




## Functional CAR models for large spatially correlated functional datasets

Lin Zhang, Veerabhadran Baladandayuthapani, Hongxiao Zhu, Keith A. Baggerly, Tadeusz Majewski, Bogdan A. Czerniak & Jeffrey S. Morris


To cite this article: Lin Zhang, Veerabhadran Baladandayuthapani, Hongxiao Zhu, Keith A. Baggerly, Tadeusz Majewski, Bogdan A. Czerniak & Jeffrey S. Morris (2015): Functional CAR models for large spatially correlated functional datasets, Journal of the American Statistical Association, DOI: [10.1080/01621459.2015.1042581](https://doi.org/10.1080/01621459.2015.1042581)


To link to this article: <http://dx.doi.org/10.1080/01621459.2015.1042581>

 View supplementary material [↗](#)

 Accepted author version posted online: 24 Jun 2015.

 Submit your article to this journal [↗](#)

 Article views: 136

 View related articles [↗](#)

 View Crossmark data [↗](#)

# Functional CAR models for large spatially correlated functional datasets

Lin Zhang<sup>1</sup>, Veerabhadran Baladandayuthapani<sup>1</sup>, Hongxiao Zhu<sup>2</sup>, Keith A. Baggerly<sup>1</sup>, Tadeusz Majewski<sup>1</sup>, Bogdan A. Czerniak<sup>1</sup>, Jeffrey S. Morris<sup>1</sup>

<sup>1</sup> *The University of Texas M.D. Anderson Cancer Center, Houston, Texas, U.S.A.*

<sup>2</sup> *Virginia Tech, Blacksburg, Virginia, U.S.A.*

## Author Footnote:

Zhang is Postdoctoral Fellow, Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston TX 77030 (E-mail: lizhang9@mdanderson.org). Baladandayuthapani is Associate Professor, Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston TX 77030 (E-mail: Veera@mdanderson.org). Zhu is Assistant Professor, Department of Statistics, Virginia Tech, Blacksburg VA 24061 (E-mail: hongxiao@vt.edu). Baggerly is Professor, Department of Bioinformatics and Computational Biology, The University of Texas M.D. Anderson Cancer Center, Houston TX 77030 (E-mail: kabagger@mdanderson.org). Majewski is Research Scientist, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston TX 77230 (E-mail: tmajewsk@mdanderson.org). Czerniak is Professor, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston TX 77230 (E-mail: bczernia@mdanderson.org). Morris is Professor, Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston TX 77030 (E-mail: jef-morris@mdanderson.org) This work has been supported by NIH grants R01-CA107304, R01-CA160736, R01-CA178744, P50-CA91846, U24-CA143883, P30-CA016672 (Biostatistics and Bioinformatics Shared Resources), and the Michael & Susan Dell Foundation (honoring Lorraine Dell). The authors thank the editor, associate editor, and reviewers for their valuable comments and

suggestions that significantly improved this manuscript. We also thank contributions from Wonyul Lee, Richard Herrick, Li Zhang, and Hui Yao.

**Abstract**

We develop a functional conditional autoregressive (CAR) model for spatially correlated data for which functions are collected on areal units of a lattice. Our model performs functional response regression while accounting for spatial correlations with potentially nonseparable and nonstationary covariance structure, in both the space and functional domains. We show theoretically that our construction leads to a CAR model at each functional location, with spatial covariance parameters varying and borrowing strength across the functional domain. Using basis transformation strategies, the nonseparable spatial-functional model is computationally scalable to enormous functional datasets, generalizable to different basis functions, and can be used on functions defined on higher dimensional domains such as images. Through simulation studies, we demonstrate that accounting for the spatial correlation in our modeling leads to improved functional regression performance. Applied to a high-throughput spatially correlated copy number dataset, the model identifies genetic markers not identified by comparable methods that ignore spatial correlations.

**KEYWORDS:** Conditional autoregressive model; Functional data analysis; Functional regression; Spatial functional data; Whole-organ histology and genetic maps

## 1 INTRODUCTION

Functional data analysis has been extensively investigated as it relates to the statistical modeling of samples of curves. The sudden development of this field has been motivated by the availability of an abundance of longitudinal or functional data collected from different scientific disciplines such as ecology, geology, social science, and genetic and medical research. See Ramsay and Silverman (2005) for a general overview of methods for functional data analysis. One important problem in functional data analysis is functional response regression, which involves regression of functional responses on a set of scalar predictors, each with its own functional regression coefficient. Morris (2015) reviewed a number of methods for functional response regression, most of which are based on an assumption that the curves are independently sampled. In this paper, we consider functional regression in settings where the functional data are spatially sampled. This work can be considered to be on the interface of functional data analysis and spatial modeling. In particular, we are interested in data for which the functions are observed on a lattice of spatial locations.

We denote a spatial functional process as

$$\{Y_i(t) : i \in D\},$$

where  $D = \{1, \dots, N\}$  is an index set of the spatial locations that can be fixed or random, and  $Y_i(t)$  are functional random variables measured on a common set of  $T$  positions. These functions can be observed over single- or multi-dimensional domains (e.g., images). Depending on the nature of the set  $D$ , we can classify spatial functional data into three types: (1) *functional point-reference data*, where  $D$  is a fixed subset of  $\mathbf{R}^r$  that contains an  $r$ -dimensional rectangle of positive volume, and  $i$  varies continuously over  $D$ ; (2) *functional areal data*, where  $D$  is again a fixed subset but partitioned into a finite number of areal units with well-defined boundaries; and (3) *functional point pattern data*, where  $D$  is itself random and its index set gives the spatial point pattern (Delicado *et al.*

*al.*, 2010). In many instances, it is of interest to model the effects of covariates (spatially varying or not) on the  $Y_i(t)$ , as well as to account for nested/ANOVA-type designs. This forms the crux of the methodological development in this article.

One common way to introduce spatial correlations among spatial functional data is to include random effect terms of non-specified functional form and utilize stationary isotropic covariance matrices that depend only on the distance between samples (Baladandayuthapani *et al.*, 2008; Staicu *et al.*, 2010), an approach extended to nonstationary spatial models in Zhou *et al.* (2010). Gromenko and Kokoszka (2013) proposed a nonparametric model for estimating the mean functions of functional point-reference data, which accounts for spatial correlations among curves using nonstationary isotropic covariance models based on functional principal component decomposition. Scheipl *et al.* (2014) presented functional additive mixed models, designed primarily for use with spline bases, that can accommodate interfunction spatial correlation through stationary separable covariance structures that are completely specified up to a scalar multiplicative constant. Analysis of functional areal data is based mainly on methods extended from the widely used conditional autoregressive (CAR) model (Besag, 1974). Knorr-Held and Besag (1998) and Schmid and Held (2004) applied the CAR model to each of the spatial and temporal main effects and their interaction term. Other methods have separately modeled the spatial and functional dependency and use parametric models to represent the functions, as seen in Bernardinelli *et al.* (1995), Martínez-Beneito *et al.* (2008), and Lawson *et al.* (2010). Other studies have modeled more complex functions nonparametrically by using a spline representation (MacNab and Dean, 2001) or Matérn covariance functions (Quick *et al.*, 2013) combined with a stationary CAR spatial process over the functional domain. As we later discuss, these separable assumptions, while convenient, are not appropriate for many complex spatially correlated functional data which exhibit considerable local heterogeneities, like in our motivating applications.

MacNab and Gustafson (2007) fit a nonseparable time-space model using a multivariate CAR,

which involves inversion of an  $NT$  by  $NT$  covariance matrix, preventing it from scaling up to large functional datasets. Torres-Avilés and Martínez-Beneito (2014) introduced a smoothed ANOVA multivariate model for spatial-temporal modeling which involves parametrically modeling the time trend for each geographic unit using preset orthogonal basis functions. Both of these papers focused on disease mapping or small area estimation. To analyze a longitudinal spatial periodontal dataset, Reich and Hodges (2008) introduced nonstationarity by fitting separate CAR models for each of the error, baseline mean, and alterations at each time point. All of these approaches focus on sparse, smooth, longitudinal data, rather than the complex, dense functional data considered in this paper and many other applications of contemporary interest.

Extensive work has been done on space-time processes, which can be viewed as a type of spatial functional data where the functions consist of time series on a lattice of spatial locations. See Cressie and Wikle (2011) for a review of spatial-temporal models. However, the goals of typical space-time analysis are different from those of functional regression, the primary focus of this paper. Time series analysis is typically interested in forecasting or extrapolating to future time points, while functional data analysis focuses on exploring and understanding relationships of variables within the fixed functional domain of the observed data. In time-space analysis, the primary interest typically lies in prediction, either for time points in the future (forecasting) for unmeasured spatial locations (kriging), or quantities for areal units (small area estimation or disease mapping). In functional regression for spatially correlated data, our goal is to relate the functional response to a set of predictors to identify regions of functions which are significantly associated with the predictors while taking the spatial correlations into account. Rather than predicting future observations, our methods are intended to discover highly-local changes within the functional domain, while borrowing strength both spatially and functionally in order to obtain more efficient estimation and more accurate inference.

In this paper, we propose a regression modeling framework for functional areal data. We introduce

functional analogs of univariate CAR models that combine basis function modeling with areal spatial process modeling. We prove that our construction induces a nonstationary CAR process in the functional domain. Compared to spatial smoothing using multivariate CAR models, we are the first to connect the spatial correlation model in the basis space to that in the data space, and show that the CAR spatial correlation structure is preserved in the data space. We show that the induced spatial smoothing parameters of the CAR process in the data space vary and borrow strength across  $t$ , which has never been examined in previous work. We posit the functional CAR model in the framework of the general functional mixed regression model, which can incorporate arbitrary covariates and multi-level designs. We show through simulation and real data analysis that the accommodation of spatial correlations in the functional regression leads to more efficient and accurate estimation and inferences than competitive methods that ignore the spatial correlations among curves.

Our model has the following features: It (i) is nonparametric, in that it allows for flexible nonparametric representations of the functions rather than pre-specified parametric functional forms; (ii) is nonstationary, as it allows both the autocovariance within curves and spatial covariance between curves to vary in strength over the domain of the functions; (iii) is nonseparable, where functional and spatial covariances display dependence; (iv) can be used with a variety of basis functions; (v) is scalable to large functional datasets sampled on a fine grid, and (vi) is defined within a general functional mixed model framework that accommodates multiple covariates with functional regression coefficients and different experimental designs through multiple levels of nonparametric random effect functions. The scalability of our approach to large functional datasets is rooted in the characteristics of the basis function transformation approach we use, including sparse representation, parallelizability of Bayesian computation in the number of basis functions, and linearity of the calculations in the number of basis functions. In addition, our fully Bayesian method can be fit in an automated fashion given just the functional responses, design matrices, and neigh-



neighborhood matrix of the samples. It yields posterior samples of all parameters that can be used to perform a wide range of potential Bayesian estimation and inference, pointwise as well as joint, while accounting for multiple testing. Our approach can be used with many different orthogonal and non-orthogonal basis functions, including wavelets, functional principal components, polynomials, splines, and Fourier basis functions. It can also be applied to functional data on higher dimensional domains such as images.

**Motivating applications:** The methods we introduce in this paper are generally applicable, but are motivated by various emerging high-throughput technologies in biomedical science that yield densely-observed functional data with spatial correlations present among the functions, including various types of neuroimaging data as well as *genomics* and *proteomics* data. An example of genomics data is whole-organ histological and genetic maps (WOHGMs). The WOHGM study provides a detailed examination of genomic changes for subregions that have different histologies within a single tissue map. It involves partitioning an intact biological tissue sample (such as a bladder tissue) into regions, for which the clinical histology is determined and genome-wide measurements of molecular quantities such as copy numbers are obtained. These data can thus be viewed as functions of genomic location for areal regions of the tissue that are expected to exhibit spatial correlation. Another example is proteomics data from imaging mass spectrometry (imaging MS) (Simpson *et. al.*, 2000; Kumar *et. al.*, 2002). The MS analysis of tissues *in situ* generates mass spectra for a grid of positions within a tissue on a glass slide, allowing for the investigation of the variability of molecular measurements across different tissue types and locations. This procedure produces spatially correlated mass spectra, and if aggregated within partitions of the tissue leads to areal functional data. These datasets are enormous, with functions sampled on a very fine grid. For example, the WOHGM dataset we analysed in this article involves over 30,000 genomic locations for which copy numbers were measured along one chromosome. Hence, existing methods for functional areal data analysis based on the construction of spatial-temporal covariance

functions cannot be scaled up to these large functional data observed on a dense grid. We apply our method to these data for functional regression while accounting for the spatial correlation, and show through simulation and real data analysis that our method has an improved performance on the identification of local functional changes compared to the methods that do not consider spatial correlations.

The outline for the rest of the paper is as follows. In Section 2, we present our hierarchical Bayesian model for spatially correlated functional data analysis. We discuss the posterior sampling algorithm and inference methods in Section 3. We report results from our simulation study in Section 4, and apply the method to the spatially correlated copy number data from the map of carcinomatous bladder tissue in Section 5. We finally conclude with a discussion in Section 6. The details of posterior sampling/inference and additional results are provided in appendices in the supplementary materials.

## 2 THE FUNCTIONAL CAR MODEL

In this section, we provide the background on the CAR model and introduce a functional CAR model based on general basis function transformations. For ease of exposition, we first consider the problem of estimating the mean function of a set of spatially correlated functions, which is the most basic form of functional response regression. In Section 2.3, we posit the functional CAR model in the general functional mixed model (FMM) framework, which can incorporate arbitrary numbers of covariates and multi-level designs. We start from a separable version of the model in the simple functional mean model, for which the covariance is the tensor product of the within-function and between-function/spatial correlations, and then present a more general, nonseparable model that allows the spatial correlation to vary across the functional domain.

## 2.1 CAR Model

The CAR model can be viewed as a discrete Markov random field, characterized by a full conditional distribution that depends on geographical adjacency. It brings in spatial smoothing in variances as part of a hierarchical model, and is well suited to model variables that exhibit subregions with strong correlations.

Let  $\{A_1, \dots, A_N\}$  represent areal units that form a lattice of  $D$  so that  $\{A_1, \dots, A_N\}$  form a partition of  $D$ , and let  $\{Y_i : i = 1, \dots, N\}$  be realizations of a Gaussian random vector where a scalar measurement  $Y_i$  is observed in the areal unit  $A_i$ . The CAR model assumes

$$Y_i | Y_{j, j \neq i} \sim \mathcal{N} \left( \sum_j b_{ij} Y_j, s_i \right).$$

Through Brook's lemma, the resulting joint distribution takes the form

$$\mathbf{Y} \sim \mathcal{MVN}(\mathbf{0}, (\mathbf{I}_N - B)^{-1} D_s),$$

where  $\mathcal{MVN}(\boldsymbol{\mu}, \Sigma)$  denotes a multi-variate normal distribution with mean  $\boldsymbol{\mu}$  and covariance  $\Sigma$ ,  $B = (b_{ij})$  with  $b_{ii} = 0, i = 1, \dots, N$ , and  $D_s = \text{diag}\{s_1, \dots, s_N\}$ .

The structure of  $B$  is specified by the shape of the lattice. To ensure symmetry and non-singularity of the covariance  $(\mathbf{I}_N - B)^{-1} D_s$ , one usually sets  $b_{ij} = \rho v_{ij} / v_{i+}$  and  $s_i = s / v_{i+}$ , where  $V = (v_{ij})$  is a symmetric neighborhood matrix with

$$v_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are neighbors} \\ 0 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases},$$

$v_{i+} = \sum_j v_{ij}$  is the number of neighbors of  $i$ , and  $\rho$  is a scalar spatial autoregressive parameter to be

estimated. This specification results in the joint distribution

$$\mathbf{Y} \sim \mathcal{MVN}(\mathbf{0}, s(D_v - \rho V)^{-1}), \quad (1)$$

denoted as  $CAR(\rho, s)$ . The parameter  $s \geq 0$  is a scalar variance parameter that controls the smoothing between neighboring sites, and the spatial correlation parameter  $\rho \in (0, 1)$  determines the extent to which neighboring units are similar in value. The value of  $\rho$  provides an indication of the strength of the marginal spatial correlation among the samples. However, they are not linearly related, as shown in Figure S1 in the supplementary materials. Given a neighborhood matrix as in our simulation study, the spatial correlation between direct neighbors is only moderate to high ( $> 0.65$ ) when  $\rho > 0.9$ . See Wall (2004) for a more detailed discussion of the implied spatial correlation of the CAR model.

## 2.2 Functional CAR Model

Given areal units  $\{A_1, \dots, A_N\}$  that form a lattice of  $D$ , suppose that  $\mathbf{Y}_{ij}(t)$  is a functional response observed in unit  $A_i$  ( $i = 1, \dots, N$ ) for subject  $j$  ( $j = 1, \dots, J$ ) on a common interval  $\mathcal{T}$ . Here, we start with a simple functional mean model

$$Y_{ij}(t) = \mu_j(t) + E_{ij}(t), \quad (2)$$

where  $\mu_j(t)$  is the mean function of the  $j$ th subject, and  $E_{ij}(t)$  is the residual error function for subject  $j$  at unit  $i$ , which is assumed to follow a Gaussian random process with mean zero. This model represents a generic nested model with two levels, in which all units from the same subject  $j$  are considered as a spatially correlated group and are assumed to have a group-specific mean function that we would like to estimate. As described in Section 2.3, other models can be used, e.g. adding a hierarchical level to estimate an overall mean function across  $j$ , when desired for a given application.

In practice,  $Y_{ij}(t)$  is observed only at a finite set of positions. Let  $\mathbf{Y}_{ij}$  be the row vector consisting of  $T$  observed values on a common grid,  $t = 1, \dots, T$ , on the interval  $\mathcal{T}$ , and let  $\boldsymbol{\mu}_j$  and  $\mathbf{E}_{ij}$  be the corresponding mean and random noise row vector. Then a discrete matrix version of this model can be written as

$$Y = X\boldsymbol{\mu} + E, \quad (3)$$

where  $Y = (\mathbf{Y}'_{11}, \dots, \mathbf{Y}'_{N1}, \mathbf{Y}'_{12}, \dots, \mathbf{Y}'_{NJ})'$  is an  $NJ$  by  $T$  matrix created by ‘stacking’ the observed discrete function responses  $\mathbf{Y}_{ij}$  as rows,  $X = (I_J \otimes \mathbf{1}_N)$  is a  $NJ$  by  $J$  design matrix for the mean functions,  $\boldsymbol{\mu}$  is a  $J$  by  $T$  matrix containing the mean functions on the grid, and  $E$  is  $NJ$  by  $T$  residual error matrices organized in the same fashion.

Suppose we could represent the functional responses sufficiently well using a truncated basis expansion of the form

$$Y_{ij}(t) = \sum_{k=1}^{T^*} \phi_k(t) y_{ijk}^* = \mathbf{y}_{ij}^* \boldsymbol{\Phi}(t), \quad (4)$$

where  $\boldsymbol{\Phi}(t) = [\phi_1(t), \dots, \phi_{T^*}(t)]'$  is the vector of basis functions, and  $\mathbf{y}_{ij}^* = (y_{ij1}^*, \dots, y_{ijT^*}^*)$  represents the row vector of the corresponding basis coefficients for the  $(i, j)$ th curve. Correspondingly, we could represent the functions on the discrete grid as

$$Y = Y^* \boldsymbol{\Phi},$$

where  $\boldsymbol{\Phi}$  is the  $T^*$  by  $T$  matrix of basis functions evaluated at the  $T$  observed points, and  $Y^*$  is an  $NJ$  by  $T^*$  matrix of the corresponding basis coefficients.

This basis representation can involve any generic basis functions  $\phi$ , such as functional principal components, wavelets, Fourier bases, or splines. For higher dimensional functions such as images, we can correspondingly use multi-dimensional basis functions, e.g., the tensor product of basis functions. The choice of basis functions can be pre-determined by the characteristics of the

application (e.g. wavelets, B-splines, Fourier basis) or determined empirically from the data (e.g. principal components). In particular, wavelets are suitable for irregular functions with spiky signals or discontinuities; Fourier bases are ideal for functions with stationary periodic features; and principal components work for sparse and smooth functional observations, which can be empirically estimated from the data using eigen-decompositions (Aston et al., 2010).

Our approach is meant for use with *lossless* or *near-lossless* basis representations. For a lossless basis representation, which has been called an isomorphic transformation (Morris et al., 2011), it holds that  $\|Y - Y^*\Phi\| = 0$ , which is true, for example, for discrete wavelet transforms (DWT) and discrete Fourier transforms (DFT). In these cases,  $T^* = T$ . In other cases, lower-rank projections of functions into the basis space with  $T^* < T$  can often be defined such that  $\|Y - Y^*\Phi\| = \epsilon \ll \|Y\|$ , in which the representation is approximately lossless, or *near-lossless*. For these reduced rank near-lossless representations, including truncated principal component analysis (PCA) and jointly thresholded wavelet bases, the number of basis functions of  $T^*$  can be selected such that a vast majority of the total energy at a preset level, e.g. 99%, in all the observed functions is preserved (Morris et al., 2011).

As opposed to approaches that incorporate basis functions into the design matrices of models fit by penalized least squares, we use what we term a *basis transformation approach*, wherein we transform each observed function from the data space into the dual space of basis coefficients, fit the basis-space version of our model, and then transform the results back for inference and interpretation. We obtain the basis coefficients in the dual space by right-multiplying both sides of model (3) by a transformation matrix  $W$ , where  $W = \Phi'(\Phi\Phi')^{-1}$  is the Moore-Penrose generalized inverse matrix of  $\Phi$ . More specifically, we can use the DWT algorithm in wavelet analysis to calculate  $Y^*$ , and transform back using the inverse-DWT, use the fast Fourier transformation (FFT) and inverse-FFT algorithm in Fourier analysis, and utilize the singular value decomposition  $Y = \mathcal{U}\mathcal{D}\mathcal{V}$  in PCA, which gives  $Y^* = \mathcal{U}\mathcal{D}^{1/2}$ ,  $\Phi = \mathcal{D}^{1/2}\mathcal{V}$ , and  $W = \mathcal{V}'\mathcal{D}^{-1/2}$ . For smoothing splines,

we can let  $\Phi$  be the Demmler-Reinsch basis functions (Brumback and Rice, 1998; Wand and Ormerod, 2008), B-spline basis functions orthogonalized to the roughness penalty, evaluated on the grid  $\mathbf{t}$  with knots at each grid point. Such a transformation yields a model in the basis space:

$$Y^* = X\mu^* + E^*, \quad (5)$$

where  $\mu^* = \mu W$  is a  $J$  by  $T^*$  matrix containing the basis coefficients for the mean functions on the grid, and  $E^* = EW$  consists of the residual errors in the basis space. Note that  $W$  is equal to  $\Phi'$  when  $\Phi$  is orthogonal.

**Separable functional CAR:** For ease of exposition, we first present a separable functional CAR model, and then generalize it to a considerably more flexible nonseparable model. The separable functional CAR model assumes that in the basis space, the  $N$ -dimensional residual error vector  $\mathbf{E}_{\cdot jk}^* = (E_{1jk}^*, \dots, E_{Njk}^*)'$  for each subject  $j$  and  $k$ th basis function follows an independent CAR prior

$$\mathbf{E}_{\cdot jk}^* | s_{jk}^*, \rho^* \sim \mathcal{MVN}(\mathbf{0}, s_{jk}^* (D_v - \rho^* V)^{-1}),$$

and we have

$$\begin{aligned} \text{vec}(E^*) &\sim \mathcal{MVN}(\mathbf{0}, \Sigma_{\text{vec}(E^*)}), \text{ with} \\ \Sigma_{\text{vec}(E^*)} &= \text{blkdiag} \left\{ s_{jk}^* (D_v - \rho^* V)^{-1} \right\}, \\ &= S^* \otimes (D_v - \rho^* V)^{-1}, \end{aligned} \quad (6)$$

where  $\text{vec}(E^*)$  is the vectorized version of the matrix  $E^*$ , created by stacking its columns,  $\text{blkdiag}$  stands for “block diagonal matrix”,  $V$  is the  $N \times N$  neighborhood matrix of the units from one subject, and  $S^* = \text{diag}(s_{11}^*, \dots, s_{J1}^*, s_{12}^*, \dots, s_{JT^*}^*)$ .

This model introduces spatial associations among the functional samples from the same subject  $j$  by assigning a CAR prior to each  $E_{\cdot jk}^*$  based on the proximity information given by the neighborhood matrix  $V$ . Here, we assume a single spatial smoothing parameter  $\rho^*$  across all subjects and

bases, and call the resulting model a separable functional CAR model since the covariance matrix  $\Sigma_{\text{vec}(E^*)}$  is represented as a tensor product of a stationary spatial/between-row covariance matrix  $R = (D_v - \rho^* V)^{-1}$  and a column covariance matrix  $S^*$ . We introduce the spatial correlations for the residual errors in the basis space. When transforming back into the data space, the spatial covariance structure is preserved; only the column covariance changes with an induced within-curve covariance  $S = \Phi' S^* \Phi$ . Thus, the residual errors  $E_{\cdot jt}$  for each subject  $j$  and each position  $t$  in the data space have the same spatial correlation structure and strength.

We assume independent Gaussian distributions for each basis coefficient with heterogeneous variances that depend on both basis index  $k$  and subject  $j$ , which induces flexible covariance structures in the data space, accounting for within-function correlation, and allowing nonstationarities with both the interfunctional variation and the intrafunctional correlations to vary across the functional index  $t$ , according to the chosen basis functions. The heterogeneous variances across basis coefficients induce intrafunctional dependencies in the data space even though the basis functions are modeled independently, since each basis function defines a linear combination of locations  $t$  within the function domain. That is, while  $S^*$  is modeled diagonally,  $S = \Phi' S^* \Phi$  is generally not diagonal. For some basis functions, this independence assumption in the basis space is strongly justified (e.g. for principal components, or for Fourier coefficients of a stationary process), or asymptotically justified (whitening property of wavelets; Craigmile and Percival, 2005), but for any  $\Phi$ , this can be used as a working approximation and the induced covariance structure investigated by  $S = \Phi' S^* \Phi$ . For smoothing spline representations, the use of a common variance component across all of the Demmler-Reinsch spline coefficients with  $S^* = s I_{JT^*}$  leads to a smoothing spline functional model (Guo, 2002) where  $s$  is related to the smoothing parameter of the curve-to-curve deviations.

**Nonseparable functional CAR:** In the separable functional CAR model, we assume the residual errors have the same spatial correlations for all basis functions, with a common spatial parameter  $\rho^*$ , which results in a uniform spatial correlation strength between any pair of neighbors at ev-



ery grid point  $t$  in the data space. However, this assumption is restrictive in many settings with spatial functional data, e.g., it is unlikely that the strength of the spatial associations between genetic functions will be the same everywhere along the genome. Hence, we seek a more flexible spatial functional model that allows the degree of spatial correlation to vary over  $t$ . For this purpose, we propose a general nonseparable functional CAR model that allows the spatial covariance parameter to vary across basis functions, i.e., the  $N$ -dimensional residual error vector  $\mathbf{E}_{\cdot jk}^* | s_{jk}^*, \rho_k^* = (E_{1jk}^*, \dots, E_{Njk}^*)'$  for subject  $j$  and the  $k$ th basis function independently follows a CAR prior  $\mathbf{E}_{\cdot jk}^* \sim \mathcal{MVN}(\mathbf{0}, s_{jk}^* (D_v - \rho_k^* V)^{-1})$ . Hence, we have

$$\begin{aligned} \text{vec}(E^*) &\sim \mathcal{MVN}(\mathbf{0}, \Sigma_{\text{vec}(E^*)}), \text{ with} \\ \Sigma_{\text{vec}(E^*)} &= \text{blkdiag} \left\{ s_{jk}^* (D_v - \rho_k^* V)^{-1} \right\} \\ &= \text{blkdiag} \left\{ S_k^* \otimes (D_v - \rho_k^* V)^{-1} \right\} \end{aligned} \quad (7)$$

where  $S_k^* = \text{diag}(s_{1k}^*, \dots, s_{Jk}^*)$  for  $k = 1, \dots, T^*$ .

The spatial smoothing parameter in the CAR prior for each basis function is assumed to be constant across  $j$  so that we are able to borrow strength across all  $J$  subjects in the spatial inference, which is especially useful when we have limited samples from each  $j$ . If desired, it is also possible to allow it to vary over subject  $j$  when required by the application and the analytic goals.

A natural question arises: *What is the induced family of the correlations when transforming back to the data space?* The theorem below and its proof assess the induced covariance structure in the data space and show that the residual errors in the original data space also follow the family of CAR spatial correlation structures for each functional location  $t$ , with the strength of the spatial associations between curves varying over the functional domain, and in this sense define a functional CAR process.

Suppose we have  $J$  subjects, and each subject  $j$  has  $N$  functional samples  $Y_{1j}, \dots, Y_{Nj}$  that are collected from a lattice with proximity information given by  $V$ , and each function  $i$  is observed on a common set of  $T$  discrete positions with a mean function  $\mu_j$  and error  $E_{ij}$ . Consider the following assumptions:

(A1) Suppose we define a basis function matrix  $\Phi$  of dimensionality  $T^* \times T$  ( $T^* \leq T$ ) that is full row rank, i.e.,  $\text{rank}(\Phi) = T^*$ , with accompanying  $T \times T^*$  transformation matrix  $W = \Phi'(\Phi\Phi')^{-1}$ , which transforms the observed functions into the basis space.

(A2) The residual errors in the basis space follow an independent CAR model with the distribution  $E_{\cdot,jk}^* \sim \mathcal{MVN}(\mathbf{0}, s_{jk}^*(D_v - \rho_k^*V)^{-1})$  for each subject  $j$  and  $k$ th basis function, as in (7) above.

**Theorem 1:** Under assumptions (A1) and (A2) and conditional on all other functional positions, the residual errors at position  $t$  in the data space,  $E_{\cdot,jt}$ , also follow a CAR spatial model with the following partial covariance matrix:

$$\Omega_{\text{vec}(E)}^{tt(j)} = \varsigma_{jt} D_v - \varrho_{jt} V,$$

where  $\varsigma_{jt} = \left( \sum_k \frac{w_{tk}^2}{s_{jk}^*} \right)$  and  $\varrho_{jt} = \left( \sum_k \frac{w_{tk}^2 \rho_k^*}{s_{jk}^*} \right)$ , with  $w_{tk}$  as the  $(t, k)$ th element of matrix  $W$ , the Moore-Penrose generalized inverse of the basis function matrix  $\Phi$ . Thus, the induced CAR spatial parameter at  $t$  in the data space for subject  $j$  is  $\rho_{jt} = \varrho_{jt} / \varsigma_{jt}$ , and the corresponding variance parameter is  $\varsigma_{jt}$ .

The proof of the above theorem is provided in the Appendix. Note that when  $\rho_1^* = \dots = \rho_{T^*}^* = \rho^*$ , then  $\Omega_{\text{vec}(E^*)}$  reduces to  $(S^*)^{-1} \otimes (D_v - \rho^*V)$ , which corresponds to the separable functional CAR model, with  $S^*$  as the within-curve covariance in the data space and  $(D_v - \rho^*V)$  as the between-curve covariance.

In addition, the partial cross-covariance matrix between the residual errors of different functional positions,  $t$  and  $t'$ , also has a CAR spatial correlation structure.

**Lemma 1:** *Under assumptions (A1) and (A2) and conditional on all other functional positions, the partial cross-covariance between the residual errors,  $E_{\cdot jt}$  and  $E_{\cdot jt'}$ , in the data space also follows a CAR covariance structure with the form*

$$\Omega_{\text{vec}(E)}^{tt'(j)} = \varsigma_{jtt'} D_v - \varrho_{jtt'} V,$$

where  $\varsigma_{jtt'} = \left( \sum_k \frac{w_{tk} w_{t'k}}{s_{jk}^*} \right)$ ,  $\varrho_{jtt'} = \left( \sum_k \frac{w_{tk} w_{t'k} \rho_k^*}{s_{jk}^*} \right)$ . Thus, the induced CAR spatial parameter for the partial cross-covariance between  $t$  and  $t'$  in the data space for subject  $j$  is  $\rho_{jtt'} = \varrho_{jtt'} / \varsigma_{jtt'}$ , with the corresponding variance parameter of  $\varsigma_{jtt'}$ . A detailed proof of the theorem and lemma is provided in the Appendix.

Remarks: This theorem and lemma show that our construction results in a CAR model for each functional position  $t$  as well as for the cross-covariances between  $t$  and  $t'$ ; hence, the functional CAR model captures correlation not only spatially across functional samples observed at  $t$ , but also across functional positions  $t$ , according to the chosen basis functions. Thus, the model borrows strength both spatially and across the functional locations in model fitting. This comprises an important general extension of CAR models to functional data, and of functional data analysis to the setting where the curves are spatially correlated with each other. Note that the only requirement of the basis function matrix  $\Phi$  in the theorem and lemma is to be of a full row rank; hence, the functional CAR model we present is quite general for different basis functions,  $\phi$ , including functional principal components, polynomials, wavelets, Fourier bases, and splines, amongst others. The choice of basis functions can be determined by the characteristics of the data and the intended analytic goals.

When  $\Phi$  is an orthogonal basis function matrix with  $T^* = T$  and  $\Phi\Phi' = I_T$ , we have  $W = \Phi'$ , and in that case  $w_{tk}$  in the induced CAR formulas would be  $\Phi_{kt}$ . Thus, the residual errors at position  $t$  in the data space,  $E_{jt}$ , has a CAR spatial correlation structure

$$\Omega_{\text{vec}(E)}^{t(j)} = \left( \sum_k \frac{\Phi_{kt}^2}{s_{jk}^*} \right) D_v - \left( \sum_k \frac{\Phi_{kt}^2 \rho_k^*}{s_{jk}^*} \right) V,$$

conditional on all other functional positions; the induced CAR spatial parameter at  $t$  in the data space for subject  $j$  is

$$\rho_{jt} = \left( \sum_k \frac{\Phi_{kt}^2 \rho_k^*}{s_{jk}^*} \right) / \left( \sum_k \frac{\Phi_{kt}^2}{s_{jk}^*} \right).$$

The results indicate that the CAR spatial correlation structures are preserved in the data space, with the spatial smoothing parameter to be a weighted average of those in the basis space. The weights are proportional to the square of the basis coefficients at position  $t$ ,  $\Phi_{kt}^2$ . Hence, the spatial correlation at position  $t$  in the data space is jointly determined by the spatial correlations of the bases that contribute to the signal at  $t$ .

In essence, the nonseparable functional CAR model leads to a nonstationary spatial correlation structure across the curves with induced CAR spatial and variance parameters  $\rho_{jt}$  and  $\varsigma_{jt}$  that vary and borrow strength across  $t$ . Furthermore, the underlying basis transformation strategy leads to calculations that are linear in the number of basis functions  $T^*$  and can be parallelized across  $T^*$  when cluster computing resources are available. These characteristics render this method scalable to enormous datasets, such as our WOHGMs. In the Bayesian framework, we assign a Beta( $a_\rho, b_\rho$ ) prior distribution on the spatial parameters  $\rho^*$ 's that concentrates mass in the (0,1) range.

The theorem and lemma extend to a more general functional response regression framework, the functional mixed model (FMM).

### 2.3 Functional CAR Mixed Models with General Basis Transformations

The functional model (2) simply consists of an overall subject-level mean plus correlated errors. Here, we generalize this functional CAR model to regression settings using an FMM framework that allows for arbitrary numbers and types of fixed effect predictors and random effect functions. The inclusion of fixed functional effects and random functional effects in the FMM enables simultaneous estimation of the population mean function and subject-specific curves, and provides flexibility in handling complex experimental designs and other induced correlation structures.

An FMM relates the functional responses to a set of scalar predictors through functional coefficients by the model

$$Y_{ij}(t) = \mathbf{X}_{ij}\mathbf{B}(t) + \mathbf{Z}_{ij}\mathbf{U}(t) + E_{ij}(t), \quad i = 1, \dots, N \quad j = 1, \dots, J \quad (8)$$

where  $\mathbf{X}_{ij}$  is a  $1 \times p$  vector of the fixed effect covariates of unit  $i$  in subject  $j$  associated with a  $p$ -dimensional vector of the fixed effect functions  $\mathbf{B}(t) = (B_1(t), \dots, B_p(t))'$ , with  $B_a(t)$  representing the partial effect of covariate  $a$  on the function at position  $t$  for  $a = 1, \dots, p$ , and where  $\mathbf{Z}_{ij}$  is a  $1 \times m$  vector of the random effect covariates of unit  $i$  in subject  $j$  associated with an  $m$ -dimensional vector of the random effect functions  $\mathbf{U}(t) = (U_1(t), \dots, U_m(t))'$ . Hence, the conditional mean function of the  $(i, j)$ th curve is  $\mu_{ij}(t) = \mathbf{X}_{ij}\mathbf{B}(t) + \mathbf{Z}_{ij}\mathbf{U}(t)$ . Note that model (2) is a special case of the FMM.

When the functional responses  $Y_{ij}(t)$  are all observed on a common grid of size  $T$ ,  $t_1, \dots, t_T$ , the discrete FMM can be represented as

$$Y = XB + ZU + E, \quad (9)$$

where  $Y$ ,  $B$ ,  $U$ , and  $E$  each has  $T$  columns that correspond to the  $T$  observed positions, and  $Y$ ,  $X$ ,  $Z$ ,  $E$  each has  $NJ$  rows created by ‘stacking’ the  $N$  discrete functions from all subjects in rows. As in Section 2.2, using a basis transformation strategy we can transform each observed function

from the data space into the basis space by right-multiplying by the transformation matrix  $W$ , and yield a basis-space version of the model:

$$Y^* = XB^* + ZU^* + E^*, \quad (10)$$

where  $Y^* = YW$  is the coefficients of the observed functions transformed into the basis space,  $X$  and  $Z$  are design matrices as in model (9),  $B^* = BW$  contains the basis coefficients for the fixed effect functions on the grid,  $U^* = UW$  contains the basis coefficients for the random effect functions on the grid, and  $E^* = EW$  consists of the residual errors in the basis space. We assume the residual errors  $E^*$  follow a nonseparable functional CAR model, leading to a functional CAR mixed model for model (10). Note that when no spatial correlation is modeled for the residual errors  $E^*$  and wavelet bases are used, model (10) parallels that in Morris and Carroll (2006).

Functional random effect: Note that random effects are not necessary to capture the within-curve correlation since our functional CAR errors already do so according to the choice of basis functions. However, random effect functions are useful to accommodate datasets with more complex designs, for example when the spatial functional data are further observed within nested, split-plot, or longitudinal designs. When random effects are present, we assume that the basis-space random effects are matrix-normally distributed with zero means:  $U^* \sim \mathcal{MN}(P, Q^*)$ , as described in Morris and Carroll (2006), where  $P$  and  $Q^*$  are the between-function and within-function covariance matrices, respectively, which implies that  $U \sim \mathcal{MN}(P, Q)$  in the data space with  $Q = \Phi' Q^* \Phi$ . Note that the between-function covariance matrix  $P$  is retained when projecting back into the data space, while the within-function covariance is changed. Again, we can assume diagonal heteroscedastic matrices in the basis space  $Q^* = \text{diag}(q_k^*)$ , which as previously discussed still accommodate within-function correlations in the data space  $Q$ , as determined by the chosen basis functions.

The between-function covariance matrix  $P$  can be chosen to accommodate different types of correlations between the random effect curves present in the data, and when the random effects are

conditionally independent,  $P = I_m$ . Although we present a single level of random effect functions here, the general model can accommodate multiple levels ( $h = 1, \dots, H$ ) of random effect functions, each of which has its own covariance matrices ( $P_h$  and  $Q_h$ ); thus, the model accommodates multi-level designs.

Functional fixed effect: Estimation of nonparametric functional coefficients of fixed effects, i.e.  $B_a(t)$ , can be improved by regularization, which we accomplish through sparsity priors on the corresponding basis coefficients. The selection priors to achieve sparsity in basis representation have been utilized in functional data analysis with different basis functions for denoising and capturing dominant signals in the data, as seen in the PCA (Aston et al., 2010; Yang et al, 2013; Zhu et al., 2014), wavelet analysis (Morris and Carroll, 2006), and Fourier analysis (Gilbert et al., 2002). The application of such priors induces a more flexible form of basis truncation, which is one of the key forms of regularization for general basis function modeling in functional data analysis (Ramsay and Silverman, 2005). Thus, we assume a spike Gaussian slab prior distribution as follows

$$\begin{aligned} B_{ak}^* &\sim \gamma_{ak} \mathcal{N}(0, \tau_{ak}) + (1 - \gamma_{ak}) \mathcal{I}(B_{ak}^* = 0) \\ \gamma_{ak} &\sim \text{Bernoulli}(\pi_{ah_k}), \quad a = 1, \dots, p \ \& \ k = 1, \dots, T^* \\ \tau_{ak} &\sim \mathcal{G}(\cdot), \end{aligned}$$

where  $\mathcal{I}(\cdot)$  is an indicator function, and  $h_k$  denotes the index of a subset of the basis coefficients including  $k$  that share a common regularization parameter, e.g. the scale of the coefficients for wavelets and Fourier series.

Here,  $\mathcal{G}(\cdot)$  is a scale mixture distribution determining the type of regularization on the basis coefficients. In this paper, we assign a vague Inverse-Gamma scale mixture for a ridge shrinkage. By using other scale mixtures, our approach can be adapted to induce other shrinkage priors for an  $L_1$ -like penalization, for example, the Bayesian lasso prior with an exponential mixture (e.g. Zhu *et. al.*, 2011), the normal gamma prior (Griffin and Brown, 2010), the normal exponential gamma

prior (Griffin and Brown, 2007), and the horseshoe prior of Carvalho et al. (2010). The selection hyperparameter,  $\pi_{ah_k}$ , is usually assigned a Beta prior. In cases where sparsity is not desired, an  $L_2$  penalization can be effectively done by specifying  $\pi_{ah_k} = 1$  for any  $a$ , and  $h_k$ , and a vague Inverse-Gamma prior for  $\tau_{ak}$ , leading to a ridge regression-type structure. If Demmler-Reinsch spline basis functions are used and  $\tau_{ak} = \tau_a$  for  $k$  corresponding to the spline terms, then this leads to smoothing spline estimates for the fixed effect functions as in Guo (2002).

The hierarchical functional CAR model can be concisely represented by the directed acyclic graph in Figure 1.

### 3 POSTERIOR SAMPLING AND INFERENCE

**Posterior sampling algorithm:** We fit our functional CAR mixed model using a Bayesian approach, with a Metropolis-Hastings-within-Gibbs sampling algorithm used to generate posterior samples of all the parameters based on the full conditional posterior distributions. We sample the fixed effects  $B^*$  and the parameters in the covariance matrices alternatively first, where their conditional distributions for Gibbs sampling are marginalized over the random effects  $U^*$ . We then sample the random effects  $U^*$  whenever they are of interest. Such a “collapsed” Gibbs sampling algorithm is advantageous as it improves the mixing properties of the MCMC chains and results in a more efficient sampler. The sampling strategies described here are based on the model in which we use spike-Gaussian-slab priors for the fixed basis coefficients  $B^*$ . For other scale mixtures, it is straightforward to substitute the MCMC update steps for respective sparse prior papers into this framework. Following we provide a brief summary of our MCMC strategy. Details of the sampling procedure are available in Appendix A in the supplementary materials.

- (a) For each basis coefficient  $k$ , sample the fixed effect  $B_{ak}^*$  from a mixture distribution of a point



mass at zero and a Gaussian distribution.

- (b) For each basis coefficient  $k$ , sample the variance components  $q_k^*$  and  $s_{jk}^*$  for the random effects and residual errors using a random-walk Metropolis-Hastings (MH) algorithm within the Gibbs sampling, where the proposal distribution for each parameter is an independent zero-truncated Gaussian distribution, for which the means are the previous values and the variances are estimated from the data based on the initial estimates of the variance of the maximum likelihood estimation (MLE).
- (c) For each basis coefficient  $k$ , sample the spatial parameter  $\rho_k^*$  from the full conditional distribution

$$f(\rho_k^*|\cdot) \propto |\Sigma_k|^{-1/2} \exp\{-\frac{1}{2}(y_k^* - XB_k^*)'\Sigma_k^{-1}(y_k^* - XB_k^*)\}(\rho_k^*)^{a_p}(1 - \rho_k^*)^{b_p},$$

where  $\Sigma_k$  is a function of  $\rho_k^*$  given by

$$\Sigma_k = q_k^*ZPZ' + S_k^* \otimes (D_v - \rho_k^*V)^{-1}.$$

Again, we use a single random-walk MH algorithm step with the proposal distribution to be a Gaussian distribution, truncated at zero and one and centered at the previous value. The proposal variance is the conditional MLE starting values estimated from the data.

- (d) Sample the parameters in the between-curve covariance matrix  $P$  for the random effect,  $\Theta_p$ , using the MH sampling algorithm that is similar to those in the previous step. Note that if the random effects are assumed to be independent across samples ( $P = I$ ), then there are no parameters to update in this step.
- (e) For each basis coefficient  $k$ , sample the random effect  $U_k^*$  from a multivariate Gaussian distribution. Note that we can omit this step when the random effects are not of interest, as the other parameters are sampled from the full conditionals marginalized over  $U$ .

- (f) For each basis coefficient  $k$ , sample the shrinkage parameter of the fixed effects,  $\tau_{ak}$ , directly from an Inverse-Gamma distribution. If  $\gamma_{ak} = 1$ , the parameters of the Inverse-Gamma full conditional are  $1/2 + a_\tau$  and  $B_{ak}^{*2}/2 + b_\tau$ , respectively; if  $\gamma_{ak} = 0$ , the parameters are the same as in the prior.
- (g) For each subset of bases functions  $h^*$  that share a common regularization parameter, sample the mixture parameter  $\pi_{ah^*}$  from a Beta distribution with the parameters  $\sum_{k:h_k=h^*} \mathcal{I}(B_{ak}^* = 1) + a_\pi$  and  $\sum_{k:h_k=h^*} \mathcal{I}(B_{ak}^* = 0) + b_\pi$ , respectively.

**Posterior inference of fixed effects** The MCMC algorithm described above yields posterior samples of all the parameters in the basis space mixed model (10). These posterior samples can then be transformed back into the data space, yielding posterior samples of the parameters in the data space mixed model (9). Specifically, we obtain posterior samples of the fixed effect function  $B_a(t)$  on the grid of  $T$  by right-multiplying the posterior samples of the corresponding vector of wavelet coefficients  $B_a^* = (B_{a1}^*, \dots, B_{aT^*}^*)$  by the  $T^*$  by  $T$  basis function matrix  $\Phi$ . From these, we can construct any Bayesian estimators and perform a wide range of Bayesian inference on the fixed effects. Here, we focus on the posterior inference of fixed effect functions since it is usually the primary interest. These methods also work for any function of the parameters in the functional CAR model and can be computed by taking combinations across coefficients, such as contrasts of fixed effect functions, as well as for any specified aggregated regions within the functions.

One of the key inferential questions of interest in FMMs is the identification of functional locations  $t$  for which the fixed effect function  $B_a(t)$  is significantly different from zero. One standard method is to use the pointwise credible bands based on MCMC samples, and flag any position  $t$  with a credible band that does not include zero. However, these pointwise credible bands do not have joint coverage probabilities, and inference based on them does not adjust for the inherent multiple testing problem. Hence, we propose two methods for significance identification with global

coverage properties: simultaneous credible band and False-discovery-rate (FDR)-based thresholding method. The former controls the global experiment-wise error rate and the latter controls the global false discovery rate.

A *simultaneous credible band* strictly controls the experiment-wise error rate at a preset level  $\alpha$ . For any fixed effect function  $\mathbf{B}_a(t)$  that are observed on  $T$  discrete positions, we can use simultaneous credible bands to identify significant regions of the curves by flagging any regions where zero is excluded. Such regions can be confidently declared significant, and the existence of any such region can be thought to reject the functional null hypothesis of  $B_a(t) = 0$  at a significance level of  $\alpha$ . Note that we can obtain the posterior samples of the difference functions between any pair of groups or any other contrasts of the fixed effect functions. Therefore, the methods for constructing simultaneous credible bands can also be applied to any contrasts of the fixed effect functions that are of interest for inference. In addition, since we have posterior samples for each position  $t$  of the entire functions, the simultaneous band inference can be carried out over any subsection or aggregation of the grid when certain subregions of the function are of particular interest.

The *FDR-based thresholding method* is used to identify regions of significant occurrence of certain events. For example, in our functional copy number application, we want to identify genomic regions with monotonic copy number gain or loss over histologic change, which may indicate an association with bladder cancer development. For this purpose, we introduce binary indicators  $C(t)$  to denote the occurrence of an event of interest, in the posterior inference, and estimate the posterior probability of the occurrence of the event at position  $t$ ,  $\hat{p}_t$ , based on the posterior samples of  $C(t)$ . Given a desired global FDR bound  $\alpha \in (0, 1)$ , the FDR-based method determines a threshold value  $\zeta_\alpha$  and flags the set of groups  $\hat{p}_t > \zeta_\alpha$  as significant regions. The threshold  $\zeta_\alpha$  is a cutpoint on the posterior probabilities that controls the expected Bayesian FDR at level  $\alpha$ . The FDR-based thresholding method can be applied to the study of any phenomenon that can be represented by binary indicators at each MCMC iteration and are of interest to researchers.

The details of constructing a simultaneous credible band and the FDR-based thresholding method are provided in Appendix B in the supplementary materials.

## 4 SIMULATION STUDIES

We conducted a simulation study based on a real MS dataset to examine the performance of the functional CAR model using wavelets as the basis functions. We considered wavelets here because the MS data are very spiky functional data in which the key features are the peaks in the mass spectra corresponding to proteins present in the cell samples. We present a brief description and the main results of the simulation studies in this section. The details of the simulation studies are provided in Appendix C in the supplementary materials. We generated functional data based on the MS data analysis of Zhu *et. al.* (2011), which were observed on an equally spaced grid of 512. The fixed effect design matrix  $X$  has five columns: columns one to four indicate four treatments, and column five indicates two levels of laser intensity. We generated 40 functions for each dataset, 10 functions for each treatment group, with estimates from Zhu *et. al.* (2011) of  $B_{ak}^*$  as the true fixed effect basis coefficients for data simulation. These 40 curves were from the same subject with  $J = 1$ , and their neighborhood matrix was generated so that the spatial correlation structure was analogous to an autoregressive process of the first order. To examine the performance of the proposed functional CAR model, we considered three different settings to simulate the residual errors  $E_{ik}^*$ : (i) Heteroscedastic and spatially correlated errors, where the variance parameters  $s_k^*$  were set the same as the estimates from Zhu, *et. al.* (2011), and the spatial parameters  $\rho_k^*$  were set such that the induced CAR parameter in the data space gradually decreased from 1 to 0, as shown in Figure S2 in the supplementary materials. (ii) Homoscedastic and spatially correlated errors, where  $s_k^*$  were set to be constant at 1, and  $\rho_k^*$  were set the same as in the first setting. (iii) Heteroscedastic and spatially independent errors, where  $s_k^*$  were set the same as the estimates from

Zhu *et. al.* (2011), and  $\rho_k^*$  were set to be 0 for all  $k$ . These three settings cover the three scenarios: (i) is where we vary both functional and spatial correlation; (ii) is where we vary only spatial correlation; and (iii) is where there is only functional correlation and no spatial correlation. We simulated a total of 30 complete datasets, with 10 for each simulation setting.

For each simulated dataset, we applied both the functional CAR model and independent wavelet-based FMM (WFMM; Morris and Carroll, 2006), using wavelets for the basis functions in both cases. We evaluated the two models in estimating the fixed effect functions  $B_a(t)$  by four measures: (i) the integrated mean squared error (IMSE) giving the accuracy of the posterior mean estimates, (ii) the integrated posterior variability (IPVar) summarizing the posterior variability about the posterior mean, (iii) the coverage probability of 95% pointwise credible bands covering the true values, and (iv) the coverage probability of 95% simultaneous credible bands covering the true values. (The definitions of IMSE and IPVar are given in the supplementary materials.) The mean and the standard deviation of the results for both methods are presented in Table 1, which summarizes each separately for the functional regions with high ( $\rho > 0.9$ ), medium ( $0.5 < \rho < 0.9$ ), and low ( $\rho < 0.5$ ) spatial correlations. The strong signals correspond to the group means  $B_1(t)$  to  $B_4(t)$  with a high signal-to-noise ratio and the weak signal corresponds to  $B_5(t)$  with a very low signal-to-noise ratio.

We see in the table that the functional CAR model performed better for strong signal estimation when medium to high spatial correlations were present in the data, with higher estimation accuracy (lower IMSE), lower posterior variability (IPVar), and higher coverage of posterior credible bands, which is more obvious when the variances of errors are relatively large, as in the homoscedastic setting. For the weak signal estimation, the functional CAR model and the independent WFMM performed similarly for the heteroscedastic setting in the presence of a medium to high spatial correlation. However, the functional CAR model did show an improved estimation of the weak signal for the homoscedastic setting, which had a signal-to-noise ratio below 1. And for both

settings (i) and (ii), the functional CAR model was slightly worse in fixed effect estimation than the independent model for regions of low spatial correlation.

To further investigate the efficiency of our proposed functional spatial model, we applied the functional CAR model to data that were simulated as spatially independent. The results showed that our model performed only slightly worse than the independent WFMM in terms of the IMSE and the coverage probabilities of the posterior credible bands. This is reasonable, as the functional model has to pay a small price to incorporate the spatial structure when there are actually no correlations between functional samples. The simulation results are further discussed in Appendix C and illustrated in Figures S2, S3, S4, and S5 of the supplementary materials.

## 5 APPLICATION: WHOLE-ORGAN HISTOLOGICAL AND GENETIC MAP (WOHGM)

### 5.1 Data Description

We applied our method to a spatially correlated copy number dataset generated through a WOHGM study of bladder cancer. Cancer develops as a series of genetic changes that eventually lead to unchecked cell division and invasion of regions normally reserved for other cells. As the aberrant cells multiply, they form spatially coherent regions of alteration, or plaques. Subsequent genetic changes and cell proliferation result in a series of nested plaques in tumor tissues. Hence, it is hypothesized that spatial changes near the tumor recapitulate the chronology of cancer development, with nesting depth providing the order of changes (Majewski *et. al.*, 2008). Tracking such spatial changes can provide valuable clues about the development of cancer. Bladder carcinoma can serve as a model for such an investigation, because it develops across the two-dimensional internal surface of a bladder, and treatment for this disease can require resection of the entire organ. By re-

secting the bladder and profiling each section, a WOHGM that integrates molecular changes with histology can be assembled. In this paper, we focus on a high-throughput copy number dataset generated by WOHGM studies.

A schematic of the WOHGM assembly is shown in Figure 2. In brief, map assembly used an entire resected bladder, which was slit from top to bottom, with the slit approximately opposite the main tumor mass. The bladder surface was then pinned to a paraffin block for subsequent gridding. The mapping grid sectioned the bladder surface into 41 rectangular sections of approximately 1cm high by 2cm wide. The maps were consistently oriented so that the slit was vertical, meaning that map sections at the left and right edges of a row were initially connected when the bladder was intact. Cells from each section were then examined and assigned a histological class ranging from N (normal tissue, shown in light gray) to D5 (invasive carcinoma, shown in dark gray). Note that the joining is approximate as the rectangular representation is obviously idealized. The copy number alterations were then measured on single nucleotide polymorphisms (SNPs) over the whole genome using Illumina microarrays for 33 out of the 41 section samples, and normalized with respect to the copy number profiles from the matched peripheral blood cells. A gain or loss of copy number when compared to the copy number in the peripheral blood cells from the same patient is depicted as green or red blocks, respectively, as shown in Figure 2. The narrow horizontal “planes” correspond to the observed copy number alterations at 50 SNP sites that were identified as having significantly monotonic copy number gains or losses with increasing severity of histological class in the following analysis. We see that the planes show a plaque-like copy number change pattern, where a gain (or loss) in copy number is found in a large contiguous area of the bladder. Such a plaque-like copy number alteration indicates the presence of strong spatial correlations among the sections in the bladder map.

Hence, the copy number dataset involves two indices of measurements: a two-dimensional spatial index that tracks the bladder map and a genetic index that tracks the chromosomal location

along the genome. Our goal here is to identify significant genetic markers that are potentially associated with bladder cancer development while accounting for the dependencies of copy number alterations along both indices. More specifically, we want to identify genes that have significant monotonic gain or loss of copy number in bladder cancer cells as their histologic classification increases in severity. When we take the copy number alterations along the genome as a functional curve, it is easy to see that the copy number data generated by the WOHGM technique fall into the functional areal data type as the mapping grid partitions the bladder surface into areal units (sections). Note that the copy number data from the WOHGM studies are very large as the copy numbers are measured at about 34,000 SNP sites along one chromosome. As a consequence, it is unrealistic to apply existing methods for functional areal data analysis that are based on constructions of spatial-temporal covariance functions to such large datasets.

## 5.2 Analysis Results

We applied our functional CAR model using wavelet basis functions and lossless transformation, and compared our algorithm with the independent WFMM in the functional estimation of the copy number dataset. Here, we focus on the copy number function along chromosome 14, where regions with plaque-like alterations were identified in initial data analysis. Copy number alterations were measured at a total of  $T = 33906$  SNP sites. The goal was to find genetic markers of monotonic copy gain or loss that correlate with increasing severity of histologic classification, which are thus potentially associated with the progression of bladder cancer. Our analysis here focuses on a single bladder map (i.e. the number of subjects  $J = 1$ ), with 33 samples collected from the map, although our model can accommodate multiple maps.

We categorized each of the 33 samples into one of three categories based on their histologic classification: Normal grade for samples with histological code N, low grade for samples with D1 or D2,



and high grade for samples with D3 to D5. We used the cell mean model for the factorial design; hence, the design matrix  $X$  was a  $33 \times 3$  matrix, with the columns indicating the histologic groups. We applied the basis transformation to each observed copy number function along chromosome 14 using the Haar wavelet, periodic boundary conditions, and decomposed to 10 wavelet levels. We used the Haar wavelet for the analysis because it has one vanishing moment and is suitable for the data with break points in copy number alterations. Our modeling effectively treated the copy number data as if they were evenly spaced along the chromosome. It has been shown that using the Haar wavelet, the approach of treating the data as evenly spaced performs comparably to approaches that model unequal spacing (Sardy *et. al.*, 1999; Hsu *et. al.*, 2005).

For the parameters  $\pi_{ah_k}$ ,  $\tau_{ak}$ ,  $q_k^*$ , and  $s_k^*$ , we specified vague proper Beta and Inverse-Gamma hyper-priors that are centered at the conditional MLEs, making this determination as described by Morris and Carroll (2006), with large variances. For the spatial smoothing parameters  $\rho_k^*$ , a skewed prior is often used in spatial CAR modeling to encourage high spatial correlations, as seen in Diva *et al.* (2007) and Bandyopadhyay *et al.* (2011). Here, we assigned a Beta(1,1) prior based on our exploratory analysis, which is uniform between 0 and 1.

We ran the MCMC chain for 5000 iterations after a burn-in of 3000. The running time was about 30 minutes for 200 iterations based on a single processor. The posterior means and standard deviations of the spatial smoothing and variance parameters are displayed in Figure S7 in the supplementary materials. From the posterior samples of the fixed effect functions, we constructed posterior samples of the respective overall mean  $C_0(t) = \frac{1}{3} \sum_{a=1}^3 B_a(t)$ ; the indicator of a same-direction copy number alteration pattern,  $C_1(t) = I[\text{sgn}\{B_1(t)\} = \text{sgn}\{B_2(t)\} = \text{sgn}\{B_3(t)\}]$ ; the indicator of monotonic copy number gain over histologic change,  $C_2(t) = I\{B_1(t) < B_2(t) < B_3(t)\}$ ; and the indicator of monotonic copy number loss over histologic change,  $C_3(t) = I\{B_1(t) > B_2(t) > B_3(t)\}$ . We then computed the posterior probabilities of the composite event,  $C_4(t) = I\{|C_0(t)| > \delta\} \cdot C_1(t) \cdot \{C_2(t) + C_3(t)\}$ , which indicates a monotonic gain or loss of copy number conditional

on the overall mean being above a threshold  $\delta$ . We chose  $\delta = 0.2$  to indicate significant copy number alterations. Based on the posterior samples of the binary indicator,  $C_4(t)$ , we then identified significant genetic markers with monotonic gain or loss of copy number over histologic change using the FDR-based thresholding method, with an FDR of  $\alpha = 0.05$ .

The results produced by the functional CAR model and the independent WFMM are presented as a Circos plot in Figure 3 (a) and (b), respectively. The outer gray layer displays the axis of the genomic locations along chromosome 14; the middle layer gives the posterior mean estimates of the overall mean  $C_0$ , with significant sites ( $> 0.2$  in magnitude) indicated in purple; the inner layer plots the posterior probabilities of monotonously amplified or decreased copy numbers over histologic change, conditional on the overall mean greater than 0.2 in magnitude, with the colours indicating the flagged significant sites based on an FDR of 0.05. In particular, SNP sites that were flagged as having significantly monotonic gain or loss of copy number are indicated separately in green or red.

From the Circos plots, we see that the functional CAR model identified 145 SNP sites that had significantly monotonic copy number gains or losses as the severity of the histological class of the cell samples increased. In contrast, the independent WFMM flagged only 9 sites. Of those 145 SNP sites flagged by the functional CAR model, 111 had monotonic loss and 34 had monotonic gain. These genetic markers were mapped into coding regions of 41 genes with known functions, as indicated in the middle of the Circos plot. We analysed the identified genes using Ingenuity Pathway Analysis (IPA) software (Ingenuity® Systems, [www.ingenuity.com](http://www.ingenuity.com)) in order to gain insight into the signaling pathways and cellular functions associated with the set of genes. The identified canonical pathways (shown in Figure S8) are mainly related to cellular proliferation, tissue development, and organ morphology. In addition, these genes are mapped to three networks/pathways, which are functionally related to cancer, cell death and survival, and cellular growth and proliferation, as shown in Figure S9. Seventeen of the genes, *ADCY4*, *BCL11B*, *DICER1*, *ELMSAN1*, *EVL*,

*FOXN3*, *MDGA2*, *PCNXL4*, *PPP2R5C*, *PRKCH*, *RALGAP1*, *RIN3*, *RNF31*, *RPGRIP1*, *TRIP11*, *UNC79*, and *VRK1*, are related to colon cancer, indicating that a partial molecular mechanism underlying the development of bladder cancer may be shared with that of colon cancer.

## 6 DISCUSSION

In this paper, we propose a Bayesian functional analytic approach that performs functional regression for spatially correlated functional data. Our approach can be viewed as merging aspects of spatial modelling and functional data analysis together in a single unified model. In the model, we introduce a spatial CAR model between curves in the basis space and prove that it induces the spatial CAR correlation structure in the data space as well. To our knowledge, we are the first to demonstrate such property of CAR models. The method incorporates the areal spatial information for functional estimation in a nonstationary and nonseparable way, and can be used for regression analysis for functions sampled on a fine grid. The method borrows strength spatially in function estimation by incorporating a spatial CAR model, given the proximity information of the sample curves. The method is flexible in allowing the spatial association between curves to vary in strength over the grid  $t$ , and is able to handle large datasets utilizing basis function transformations. In addition, the functional CAR model is generalizable for functional analysis using different basis functions, orthogonal or non-orthogonal, based on the data characteristics and analytic purpose, and can be used for multi-dimensional functional data. The computational time is linear to the selected number of basis functions and at worst to the number of subjects. With the parallelizability of the Bayesian computation in the number of basis functions, the method is scalable to large spatial functional datasets with a fine grid and moderate number of subjects.

We conducted simulation studies based on a real mass spectrometry dataset under various settings to evaluate the operating characteristics of our method. We found that by taking the spatial cor-

relation into account, the functional CAR model yielded better estimation and inference over the independent functional model when medium to high spatial correlations were present among the samples. In particular, the functional CAR model was able to smooth out spurious noise features while identifying strong signal features, as well as to distinguish weak signals when the noise level was relatively high. The improved performance can be explained by the property of the functional CAR that borrows strength spatially in local estimation, not only across observations at the same position but also from other locations within the function according to the chosen basis. This property makes the model more robust to high noise levels. The simulation results also showed that the functional CAR model experienced slight trade-offs in performance compared to the independent model when the spatial correlation was low or absent, which was expected since the functional CAR model loses some efficiency by including extra spatial parameters. In terms of identifying significant regions of change, we found that the functional CAR model consistently outperformed the independent FMM when spatial correlation was present.

The functional CAR model is a useful tool as many extremely large functional datasets with spatial correlations become available, especially in biomedical research. We presented the model for univariate functional responses and one-dimensional functions. However, it can be easily extended for higher dimensional functions with a multi-dimensional basis transformation strategy or for multivariate functional responses utilizing multivariate spatial CAR models. We leave the details of these potential extensions for future exploration.

## APPENDIX: PROOF OF THEOREM 1 AND LEMMA 1

A key technical element of this proof is to apply a linear transformation of  $\text{vec}(E^*)$  to obtain  $\text{vec}(E)$  and its covariance or inverse-covariance matrix. The linear transformation is composed of three steps:  $\text{vec}(E^*) \rightarrow \text{vec}(\{E^*\}') \rightarrow \text{vec}(E') \rightarrow \text{vec}(E)$ , where  $\text{vec}(E^*)$  is the vector obtained by staking

the columns of the matrix  $E^*$  ( $T^*$  blocks of  $NJ \times 1$  vector), and  $\text{vec}(\{E^*\}')$  is the vector obtained by stacking the rows of  $E^*$  ( $NJ$  blocks of  $T^* \times 1$  vector). Notice that the two vectors are two permutations of the same elements. This is also the case for  $\text{vec}(E)$  and  $\text{vec}(E')$ .

- Step 1: Since  $\text{vec}(E^*)$  and  $\text{vec}(\{E^*\}')$  are two permutations of the same elements, there exists some  $(NJ)T^* \times (NJ)T^*$  permutation matrix  $P^*$  such that

$$\text{vec}(\{E^*\}') = P^* \cdot \text{vec}(E^*).$$

This permutation matrix  $P^*$  has the following properties: (i)  $P^*$  is orthogonal, i.e.  $P^*(P^*)' = (P^*)'P^* = I_{NJ T^*}$ ; (ii)  $(P^*)'\text{vec}(\{E^*\}') = (P^*)'P^*\text{vec}(E^*) = \text{vec}(E^*)$ ; and (iii) If an  $NJT^* \times 1$  vector  $x$  has  $\text{cov}(x) = A \otimes B$  where  $A$  is a  $T^* \times T^*$  semi-positive definite matrix and  $B$  is a  $NJ \times NJ$  positive definite matrix, then  $\text{cov}(P^*x) = B \otimes A$ . Hence, we have  $P^* \cdot (A \otimes B) \cdot (P^*)' = B \otimes A$ .

- Step 2: Since  $\text{vec}(\{E^*\}')$  is composed of  $NJ$  blocks of  $T^* \times 1$  vectors, we can transform each block by left-multiplying  $\Phi'$  and obtain

$$\text{vec}(E') = (I_{NJ} \otimes \Phi') \cdot \text{vec}(\{E^*\}').$$

- Step 3: Again,  $\text{vec}(E)$  and  $\text{vec}(E')$  are two permutations of the same elements, and hence there exists an  $(NJ)T \times (NJ)T$  permutation matrix  $P$  such that

$$\text{vec}(E') = P \cdot \text{vec}(E), \text{ and } \text{vec}(E) = P' \cdot \text{vec}(E').$$

Combining the three steps, we have

$$\text{vec}(E) = P' \cdot (I_{NJ} \otimes \Phi') \cdot P^* \cdot \text{vec}(E^*).$$

And correspondingly, the induced covariance matrix of  $\text{vec}(E)$  is

$$\Sigma_{\text{vec}(E)} = P' \cdot (I_{NJ} \otimes \Phi') \cdot P^* \cdot \Sigma_{\text{vec}(E^*)} \cdot (P^*)' \cdot (I_{NJ} \otimes \Phi) \cdot P \quad (11)$$

Note that:  $\text{rank}(\Sigma_{\text{vec}(E)}) \leq \min\{\text{rank}(P), \text{rank}(I_{NJ} \otimes \Phi), \text{rank}(P^*), \text{rank}(\Sigma_{\text{vec}(E^*)})\} = NT^*$ . Hence, when  $T^* < T$ ,  $\Sigma_{\text{vec}(E)}$  is not of full rank and its inverse does not exist. Therefore, we work on the generalized inverse of  $\Sigma_{\text{vec}(E)}$  in a more general case.

We now derive the generalized inverse matrix of  $\Sigma_{\text{vec}(E)}$ ,  $\Omega_{\text{vec}(E)}$ . Before that, we give several properties of the tensor product and generalized inverse matrix that will be used in the derivation:

- (i) If an  $m \times n$  matrix  $A$  has full column rank and  $n \times q$  matrix  $B$  has full row rank, the  $(AB)^- = B^- A^-$ , where  $M^-$  denotes the generalized inverse matrix of the matrix  $M$ .
- (ii) Let  $W$  be the generalized inverse of matrix  $\Phi$ . We then have (a)  $(\Phi')^- = (\Phi^-)' = W'$ ; (b)  $(I \otimes \Phi)^- = I \otimes W$ ; (c)  $(I \otimes \Phi')^- = I \otimes W'$ .
- (iii) If  $A, B, C$  and  $D$  are matrices of such size that one can form the matrix products  $AC$  and  $BD$ , then  $(A \otimes B)(C \otimes D) = AC \otimes BD$ .

We now obtain  $\Omega_{\text{vec}(E)}$  by taking the generalized inverse on the right side in (11). Since the square matrix  $P$  and  $\Sigma_{\text{vec}(E^*)}$  are of full rank,  $(I_{NJ} \otimes \Phi)$  is of full row rank, and  $(I_{NJ} \otimes \Phi')$  is of full column rank, by property (i) we have

$$\Omega_{\text{vec}(E)} = P^- \cdot (I_{NJ} \otimes \Phi)^- \cdot \{(P^*)'\}^- \cdot \{\Sigma_{\text{vec}(E^*)}\}^- \cdot (P^*)^- \cdot (I_{NJ} \otimes \Phi')^- \cdot (P')^-$$

Since

$$\{\Sigma_{\text{vec}(E^*)}\}^- = \begin{bmatrix} (S_1^*)^{-1} \otimes (D_v - \rho_1^* V) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (S_{T^*}^*)^{-1} \otimes (D_v - \rho_{T^*}^* V) \end{bmatrix}$$

$$= (S^*)^{-1} \otimes D_v - \{(S^*)^{-1} (D_\rho \otimes I_J)\} \otimes V,$$

where  $S^* = \text{diag}(S_1^*, \dots, S_{T^*}^*)$ , and  $D_\rho = \text{diag}(\rho_1^*, \dots, \rho_{T^*}^*)$ . Let  $\mathcal{J}_k$  be a  $T^* \times T^*$  matrix with the  $(k, k)$ th element to be 1, and 0 elsewhere, and  $S^* = \sum_k \mathcal{J}_k \otimes S_k^*$ . Let  $W = \Phi^- = \Phi'(\Phi\Phi')^{-1}$  be the generalized inverse of the basis function matrix  $\Phi$ , we then have

$$\begin{aligned} \Omega_{\text{vec}(E)} &= P'(I_{NJ} \otimes W)P^*[(S^*)^{-1} \otimes D_v - \{(S^*)^{-1} (D_\rho \otimes I_J)\} \otimes V](P^*)'(I_{NJ} \otimes W')P \quad \text{by (ii)} \\ &= P'(I_{NJ} \otimes W)P^*\{(S^*)^{-1} \otimes D_v\}(P^*)'(I_{NJ} \otimes W')P \\ &\quad - P'(I_{NJ} \otimes W)P^*[\{(S^*)^{-1} (D_\rho \otimes I_J)\} \otimes V](P^*)'(I_{NJ} \otimes W')P \\ &= P'\{I_J \otimes (I_N \otimes W)\} \left\{ \sum_k (S_k^*)^{-1} \otimes D_v \otimes \mathcal{J}_k \right\} \{I_J \otimes (I_N \otimes W')\}P \\ &\quad - P'\{I_J \otimes (I_N \otimes W)\} \left\{ \sum_k (S_k^*)^{-1} \otimes V \otimes (\mathcal{J}_k D_\rho) \right\} \{I_J \otimes (I_N \otimes W')\}P \\ &= P' \left\{ \sum_k (S_k^*)^{-1} \otimes D_v \otimes (W \mathcal{J}_k W') \right\} P - P' \left\{ \sum_k (S_k^*)^{-1} \otimes V \otimes (W \mathcal{J}_k D_\rho W') \right\} P \quad \text{by (iii)} \\ &= \left\{ \sum_k (W \mathcal{J}_k W') \otimes (S_k^*)^{-1} \right\} \otimes D_v - \left\{ \sum_k (W \mathcal{J}_k D_\rho W') \otimes (S_k^*)^{-1} \right\} \otimes V. \end{aligned}$$

The  $N \times N$  diagonal block of  $\Omega_{\text{vec}(E)}$  corresponding to position  $t$  and subject  $j$  is

$$\Omega_{\text{vec}(E)}^{tt(j)} = \left( \sum_k \frac{w_{tk}^2}{s_{jk}^*} \right) D_v - \left( \sum_k \frac{w_{tk}^2 \rho_k^*}{s_{jk}^*} \right) V,$$

which is the partial-covariance matrix of the  $N$ -dimensional residual vector  $\mathbf{E}_{.jt} = (E_{1jt}, \dots, E_{Njt})'$  conditional on all other  $\{t' : t' \neq t\}$ . This structure corresponds to the inverse-covariance in a CAR model, with the spatial parameter  $\rho_{jt} = \left( \sum_k \frac{w_{tk}^2 \rho_k^*}{s_{jk}^*} \right) / \left( \sum_k \frac{w_{tk}^2}{s_{jk}^*} \right)$ , and the variance parameter  $s_t = \sum_k \frac{w_{tk}^2}{s_{jk}^*}$ .

Moreover, the  $N \times N$  off-diagonal block of  $\Omega_{\text{vec}(E)}$  corresponding to position  $t$  and  $t'$  of subject  $j$  is

$$\Omega_{\text{vec}(E)}^{tt'(j)} = \left( \sum_k \frac{w_{tk}w_{t'k}}{s_{jk}^*} \right) D_v - \left( \sum_k \frac{w_{tk}w_{t'k}\rho_k^*}{s_{jk}^*} \right) V,$$

which is the generalized inverse matrix of  $\text{Cov}(\mathbf{E}_{\cdot jt}, \mathbf{E}_{\cdot jt'})$  conditional on all other positions. This structure also takes the form of a CAR model, with the variance parameter  $s_{tt'} = \sum_k \frac{w_{tk}w_{t'k}}{s_{jk}^*}$  and the spatial parameter  $\rho_{jtt'} = \left( \sum_k \frac{w_{tk}w_{t'k}\rho_k^*}{s_{jk}^*} \right) / \left( \sum_k \frac{w_{tk}w_{t'k}}{s_{jk}^*} \right)$ .

Hence, in a general functional CAR model, the structure of the spatial correlations between residual errors is preserved to follow a CAR model in the data space with the spatial effect parameter changing over  $t$ , which is a function of all the spatial parameters in the basis space.



## REFERENCES

- Anderson, T. W. (1971), "The statistical analysis of time series," *Wiley Series in Probability and Mathematical Statistics*, Wiley New York.
- Aston, J. A., Chiou, J., and Evans, J. (2010), "Linguistic pitch analysis using functional principal component mixed effect models," *J. R. Statist. Soc. C*, 59, 297-317.
- Baladandayuthapani, V., Mallick, B. K., Hong, M. Y., Lupton, J. R., Turner, N. D. and Carroll, R. J. (2008), "Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis," *Biometrics*, 64, 321-322.
- Bandyopadhyay, D., Reich, B. J., and Slate, E. H. (2011), "A spatial Beta-Binomial model for clustered count data on dental caries," *Stat. Methods Med. Res.*, 20 (2), 85-102.
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M. (1995), "Bayesian analysis of space-time variation in disease risk," *Statist. Med.*, 14, 2433-2443.
- Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems," (with discussion) *J. R. Statist. Soc. B*, 36, 192-236.
- Brumback, B. A. and Rice, J. A. (1998), "Smoothing spline models for the analysis of nested and crossed samples of curves," *J. Am. Statist. Ass.*, 93 (443), 961-976.
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010), "The horseshoe estimator for sparse signals," *Biometrika*, 97(2), 465-480.
- Craigmile, P. F. and Percival, D. B. (2005), "Asymptotic decorrelation of between-scale wavelet coefficients," *IEEE Trans. on Inf. Theory*, 51, 1039-1048.
- Cressie, N., Wikle, C. (2011) *Statistics for spatio-temporal data*, John Wiley & Sons.
- Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010), "Statistics for spatial functional data: some recent contributions," *Environmetrics*, 21, 224-239.
- Diva, U., Banerjee, S., Dey, D.K. (2007), "Modelling spatially correlated survival data for individuals with multiple cancers," *Stat. Modelling*, 7 (2), 191-213.
- Gilbert, A. C., Guha, S., Indyk, P., Muthukrishnan, S., Strauss, M. (2002), "Near-optimal sparse Fourier representation via sampling," *In STOC*, pp 152-161, Montreal, Quebec, Canada.
- Griffin, J. E. and Brown, P. J. (2007), "Bayesian adaptive lassos with non-convex penalization," Technical Report, Department of Statistics, University of Warwick.
- Griffin, J. E. and Brown, P. J. (2010), "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis*, 5, 171-188.
- Gromenko, O. and Kokoszka, P. (2013), "Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination," *Computational Statistics and Data Analysis*, 59, 82-94.
- Guo W. (2002), "Functional mixed effect models," *Biometrics*, 58, 121-128.

- Hsu, L., Self, S. G., Grove, D., Randolph, T., Wang, K., Delrow, J. J., Loo, L., and Porter, P. (2005), "Denoising array-based comparative genomic hybridization data using wavelets," *Biostatistics*, 6, 211-226.
- Knorr-Held, L., Besag, J. (1998), "Modelling risk from a disease in time and space," *Statist. Med.*, 17, 2045-2060.
- Krivobokova, T., Kneib, T. and Claeskens, G. (2010), "Simultaneous confidence bands for penalized spline estimators," *J. Am. Statist. Ass.*, 105(490), 852-863.
- Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., Cheung, K. H., Miller, P., Gerstein, M., Roeder, G. S., and Snyder, M. (2002), "Subcellular localization of the yeast proteome," *Genes Dev.*, 16, 707-719.
- Lawson, A. B., Song, H. R., Cai, B., Hossain, M. M. and Huang, K. (2010), "Spacetime latent component modeling of geo-referenced health data," *Statist. Med.*, 29, 2012-2027.
- MacNab, Y. C. and Dean, C. B. (2001), "Autoregressive spatial smoothing and temporal spline smoothing for mapping rates," *Biometrics*, 57, 949-956.
- MacNab, Y. C. and Gustafson, P. (2007), "Regression B-spline smoothing in Bayesian disease mapping: With an application to patient safety surveillance," *Statist. Med.*, 26, 4455-4474.
- Majewski, T., Lee, S., Jeong, J., Yoon, D. S., Kram, A., Kim, M. S., *et. al.* (2008), "Understanding the development of human bladder cancer by using a whole-organ genomic mapping strategy," *Lab. Invest.*, 88(7), 694-721.
- Martínez-Beneito, M. A., Lopez-Quilez, A. and Botella-Rocamora, P. (2008), "An autoregressive approach to spatio-temporal disease mapping," *Statist. Med.*, 27, 2874-2889.
- Morris, J. S., Vannucci, M., Brown, P. J. and Carroll, R. J. (2003), "Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis," *J. Am. Statist. Ass.*, 98, 573-583.
- Morris, J. S. and Carroll, R. J. (2006), "Wavelet-based functional mixed models," *J. R. Statist. Soc. B*, 68(2), 179-199.
- Morris, J. S., Brown, P. J., Herrick, R. C., Baggerly, K. A., and Coombes, K. R. (2008) "Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models," *Biometrics*, 64, 479-489.
- Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P. P., Gutstein, H. (2011), "Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data," *Ann. Appl. Statist.*, 5, 894-923.
- Morris, J. S. (2015), "Functional regression", *Annual Review of Statistics and its Application*, 2, to appear.
- Müller, P., Parmigiani, G., Robert, C., and Rousseau, J. (2004), "Optimal sample size for multiple testing: the case of gene expression microarrays," *J. Am. Statist. Ass.*, 99, 990-1001.
- Quick, H., Banerjee, S. and Carlin, B. P. (2013), "Modeling temporal gradients in regionally aggregated California asthma hospitalization data," *Ann. Appl. Statist.*, 7(1), 154-176.
- Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, 2nd edition. New York: Springer-Verlag.

- Reich, B. J. and Hodges, J. S. (2008), "Modeling longitudinal spatial periodontal data: A spatially-adaptive model with tools for specifying priors and checking fit," *Biometrics*, 64, 790-799.
- Ruppert, D., Wand, M. P. and Carroll, R. J. (2003), *Semiparametric Regression*. New York: Cambridge University Press.
- Sardy, S., Percival, D. B., Bruce, A. G., Gao, H-Y. and Stuetzle, W. (1999), "Wavelet shrinkage for unequally spaced data," *Statist. Computing*, 9, 65-75.
- Scheipl, F., Staicu, A-M, Greven, S. (2014), "Functional Additive Mixed Models," *Journal of Computational and Graphical Statistics*, to appear.
- Schmid, V. and Held, L. (2004), "Bayesian extrapolation of spacetime trends in cancer registry data," *Biometrics*, 60, 1034-1042.
- Simpson, J.C., Wellenreuther, R., Poustka, A., Pepperkok, R. and Wiemann, S. (2000), "Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing," *EMBO Reports*, 1, 287-292.
- Staicu, A., Crainiceanu, C. and Carroll, R. (2010), "Fast methods for spatially correlated multilevel functional data," *Biostatistics*, 11(2), 177-194.
- Storey, J. D. (2003), "The positive false discovery rate: A Bayesian interpretation and the q-value," *Ann. Statist.*, 31, 2013-2035.
- Torres-Avilés, F., Martínez-Beneito, M. A. (2014), "STANOVA: a smoothed-ANOVA-based model for spatio-temporal disease mapping," *Stoch. Environ. Res. Risk Assess*, doi: 10.1007/s00477-014-0888-1.
- Wall, M. M. (2004), "A close look at the spatial structure implied by the CAR and SAR models," *J. Statist. Plan. Inference*, 121, 311-324.
- Wand, M. P. and Ormerod, J. T. (2008), "On semiparametric regression with O'Sullivan penalized splines," *Aust. N. Z. J. Stat.*, 50(2), 179-198.
- Yang, W., Wikle, C. K., Holan, S. H. and Wildhaber, M. L. (2013), "Ecological Prediction With Nonlinear Multivariate Time-Frequency Functional Data Models," *J. Agr. Bio. Envir. St.*, 18(3), 450-474.
- Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V. and Carroll, R. J. (2010), "Reduced rank mixed effects models for spatially correlated hierarchical functional data," *J. Am. Statist. Ass.*, 105, 390-400.
- Zhu, H., Brown, P. J. and Morris, J. S. (2011), "Robust, adaptive functional regression in functional mixed model framework," *J. Am. Statist. Ass.*, 106(495), 1167-1179.
- Zhu, H., Yao, F. and Zhang, H. H. (2014), "Structured functional additive regression in reproducing kernel Hilbert spaces," accepted by *J. R. Statist. Soc. B*. doi: 10.1111/rssb.12036.

Table 1: Simulation Results. Estimation results of the wavelet-based functional CAR and independent WFMM models in terms of integrated mean squared error (IMSE), integrated posterior variance (IPVar), probability of 95% pointwise credible bands covering the true values (Pointwise Coverage), and probability of 95% simultaneous credible bands covering the true values (Joint Coverage), summarized by mean (and standard deviations in parentheses) over 10 replications.

Signal $\rho$	IMSE			IPVar		Pointwise Coverage			Joint Coverage		
	fcar	indep		fcar	indep	fcar	indep		fcar	indep	
Strong	High	1.21 (0.41)	3.08 (0.96)	0.55 (0.11)	0.60 (0.06)	0.811 (0.073)	0.628 (0.082)	0.979 (0.022)	0.904 (0.051)		
	Medium	0.61 (0.41)	0.79 (0.38)	0.40 (0.07)	0.59 (0.10)	0.883 (0.115)	0.903 (0.077)	0.996 (0.009)	0.995 (0.007)		
	Low	1.62 (0.52)	1.52 (0.51)	1.69 (0.27)	2.07 (0.26)	0.937 (0.038)	0.970 (0.025)	0.999 (0.002)	1.000 (0.001)		
Weak	High	0.05 (0.04)	0.04 (0.02)	0.03 (0.00)	0.06 (0.01)	0.821 (0.130)	0.982 (0.016)	0.993 (0.015)	1.000 (0.000)		
	Medium	0.06 (0.03)	0.05 (0.02)	0.04 (0.01)	0.07 (0.02)	0.837 (0.112)	0.959 (0.064)	1.000 (0.000)	1.000 (0.000)		
	Low	0.25 (0.08)	0.17 (0.06)	0.14 (0.02)	0.26 (0.03)	0.838 (0.067)	0.970 (0.026)	0.995 (0.009)	1.000 (0.000)		
Spatially correlated and homoscedastic setting											
Signal $\rho$	IMSE			IPVar		Pointwise Coverage			Joint Coverage		
	fcar	indep		fcar	indep	fcar	indep		fcar	indep	
Strong	High	18.79 (8.17)	154.20 (44.42)	9.34 (2.50)	25.59 (1.42)	0.822 (0.075)	0.487 (0.074)	1.000 (0.000)	0.892 (0.043)		
	Medium	6.66 (2.56)	24.85 (5.34)	5.28 (1.07)	9.16 (0.72)	0.922 (0.049)	0.819 (0.044)	1.000 (0.000)	0.996 (0.006)		
	Low	6.23 (1.25)	6.82 (2.00)	5.37 (0.89)	7.04 (0.67)	0.912 (0.032)	0.959 (0.018)	1.000 (0.001)	0.999 (0.004)		
Weak	High	0.16 (0.03)	0.43 (0.07)	0.22 (0.04)	0.19 (0.04)	0.967 (0.026)	0.888 (0.031)	1.000 (0.000)	1.000 (0.000)		
	Medium	0.17 (0.05)	0.10 (0.01)	0.21 (0.04)	0.10 (0.01)	0.953 (0.038)	0.936 (0.025)	1.000 (0.000)	1.000 (0.000)		
	Low	0.71 (0.18)	0.47 (0.06)	0.49 (0.09)	0.33 (0.07)	0.895 (0.036)	0.904 (0.035)	1.000 (0.000)	1.000 (0.000)		
Spatially independent and heteroscedastic errors											
Signal	IMSE			IPVar		Pointwise Coverage			Joint Coverage		
	fcar	indep		fcar	indep	fcar	indep		fcar	indep	
Strong	5.63 (1.90)	4.78 (1.82)	17.27 (3.21)	20.09 (1.79)	0.907 (0.055)	0.952 (0.043)	0.998 (0.003)	0.999 (0.003)			
Weak	0.47 (0.11)	0.42 (0.09)	1.23 (0.17)	1.94 (0.21)	0.861 (0.028)	0.962 (0.027)	0.996 (0.004)	1.000 (0.001)			

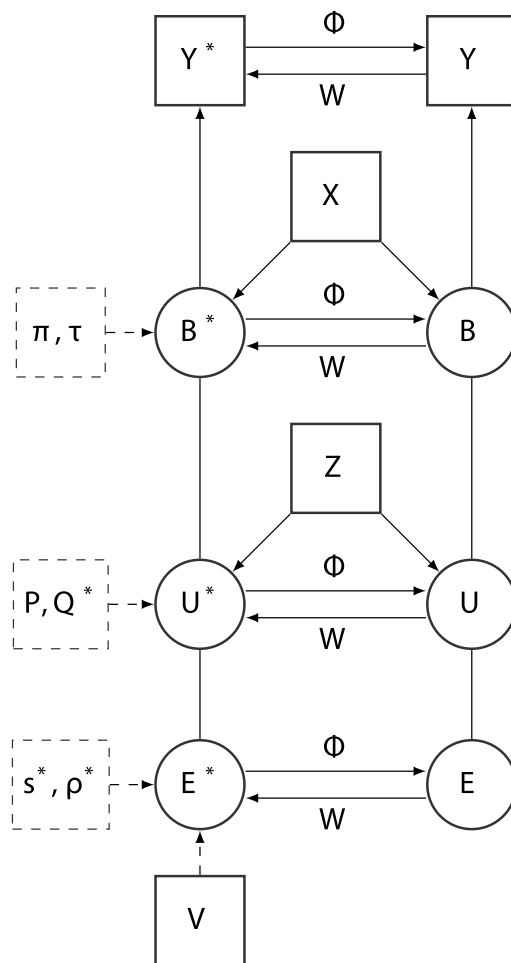


Figure 1: Directed acyclic graph for the hierarchical model of the functional CAR model. The nodes in the right column are the observed responses and functional variables in the data space; the nodes in the middle column are the transformed responses and variables projected in the basis space; the nodes in the left column are the parameters modeling the prior distributions of the variables in the basis space.  $\boldsymbol{\pi} = \{\pi_{ah_k}\}$ ,  $\boldsymbol{\tau} = \{\tau_{ak}\}$ ,  $\mathbf{s}^* = \{s_{jk}^*\}$ , and  $\boldsymbol{\rho}^* = \{\rho_{jk}^*\}$ .

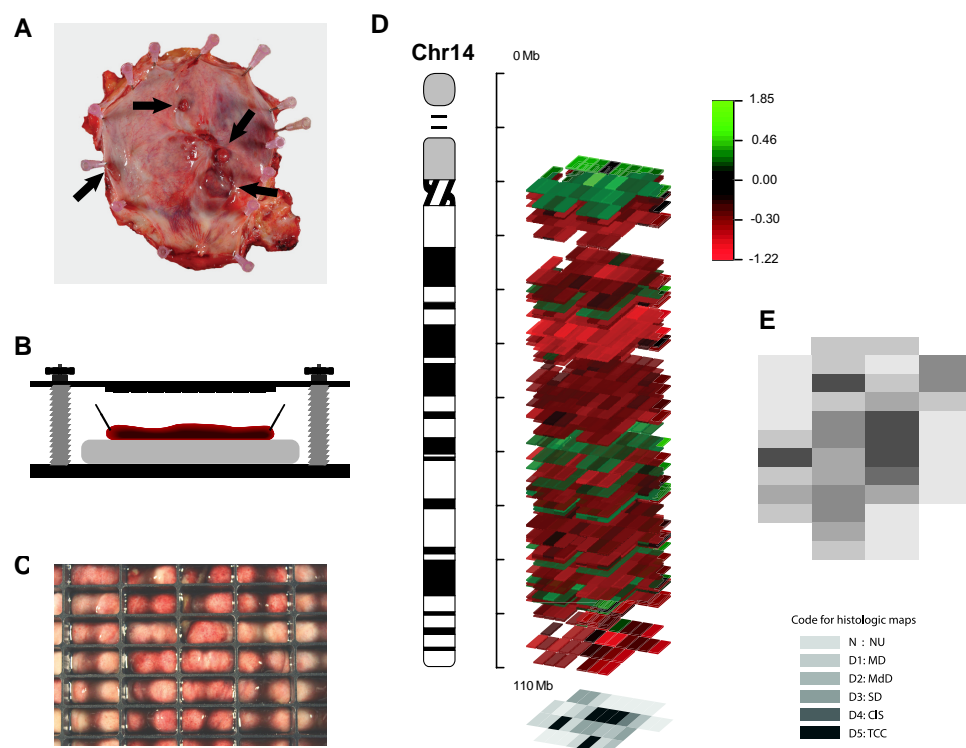


Figure 2: Schematic representation of assembly and preliminary analysis of high-resolution whole organ histologic and genetic maps (WOHGMs). A. Open cystectomy specimen with multifocal transitional cell carcinoma (indicated by black arrows). B. Open cystectomy specimen pinned down to a paraffin block for subsequent gridding. C. Top view of mapping grid superimposed over the specimen. Copy number alterations are then measured for each grid/section. D. At the bottom is a physical layout of the histologic map colour-coded from normal (light gray) to invasive carcinoma (dark gray). Plotted along the vertical chromosomal axis is the copy number variation, relative to that found in peripheral blood DNA. Gains of copy number are depicted as red blocks; losses of copy number are depicted as green blocks. E. Histologic map of cystectomy specimen. The histologic map code is as follows: NU, normal urothelium; MD, mild dysplasia; MdD, moderate dysplasia; SD, severe dysplasia; CIS, carcinoma in situ; and TCC, transitional cell carcinoma.

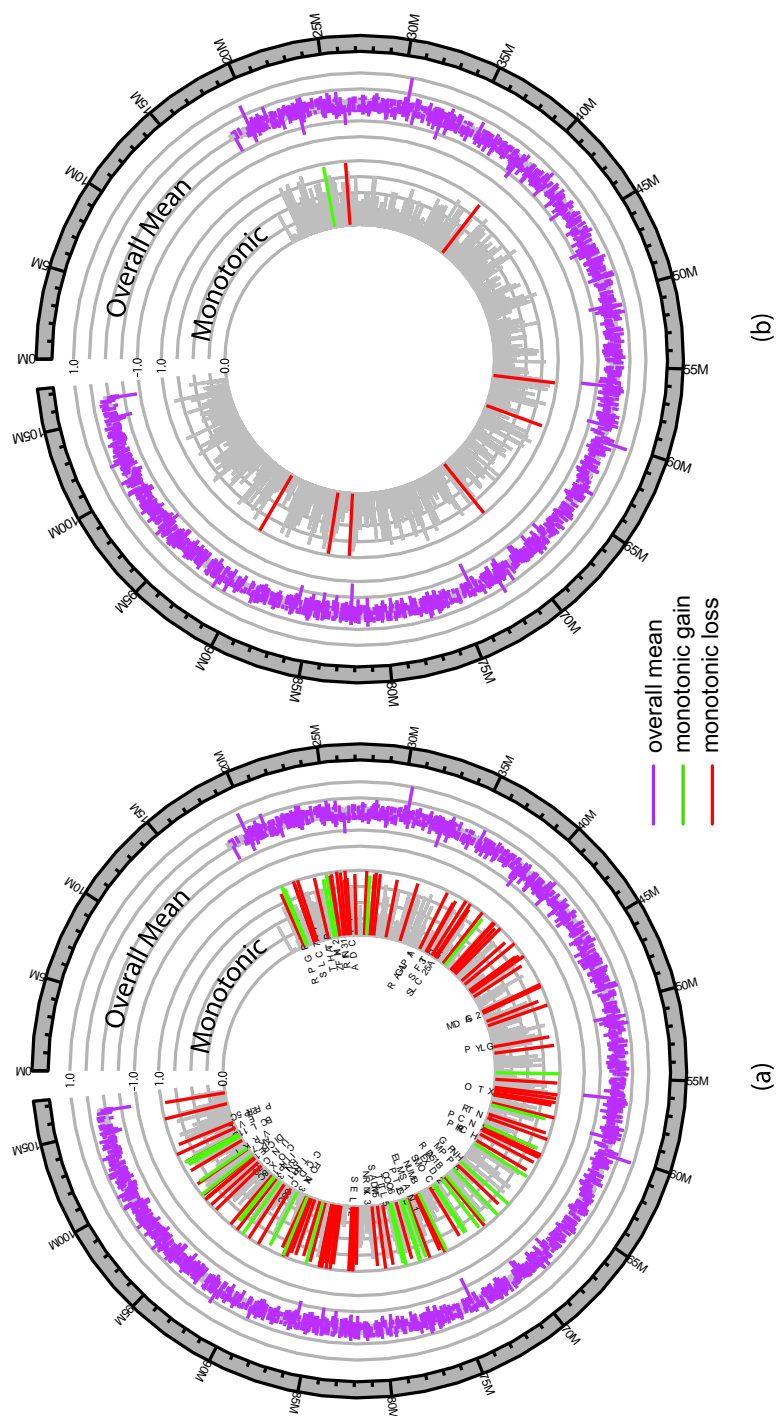


Figure 3: Circos plots of analysis results of copy number functional data on chromosome 14 obtained from (a) the functional CAR model and (b) the independent WFMM. The outermost gray circle displays the axis of the genomic locations; the middle layer gives the posterior mean estimates of the overall mean; and the inner layer presents the posterior probabilities of monotonic gain or loss of copy number. The colour lines indicate the sites flagged as significant.