Received: 25 March 2021 | Accepted: 29 January 2022

DOI: 10.1111/biom.13631

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Bayesian spatiotemporal modeling on complex-valued fMRI signals via kernel convolutions

Cheng-Han Yu[1] | **Raquel Prado**[2] | **Hernando Ombao**[3] | **Daniel Rowe**[1]

[1]Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, Wisconsin, USA

[2]Department of Statistics, University of California, Santa Cruz, California, USA

[3]Statistics Program, King Abdullah University of Science and Technology University, Saudi Arabia

Correspondence
Cheng-Han Yu, Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI 53233, USA.
Email: cheng-han.yu@marquette.edu

Abstract

We propose a model-based approach that combines Bayesian variable selection tools, a novel spatial kernel convolution structure, and autoregressive processes for detecting a subject's brain activation at the voxel level in complex-valued functional magnetic resonance imaging (CV-fMRI) data. A computationally efficient Markov chain Monte Carlo algorithm for posterior inference is developed by taking advantage of the dimension reduction of the kernel-based structure. The proposed spatiotemporal model leads to more accurate posterior probability activation maps and less false positives than alternative spatial approaches based on Gaussian process models, and other complex-valued models that do not incorporate spatial and/or temporal structure. This is illustrated in the analysis of simulated data and human task-related CV-fMRI data. In addition, we show that complex-valued approaches dominate magnitude-only approaches and that the kernel structure in our proposed model considerably improves sensitivity rates when detecting activation at the voxel level.

KEYWORDS

autoregressive, brain activation, complex-valued time series, functional magnetic resonance imaging, Gaussian processes, kernel convolution

1 | INTRODUCTION

Functional magnetic resonance imaging (fMRI) indirectly measures neuronal activity by detecting changes in the blood oxygen level-dependent (BOLD) signal. Task-related experiments typically lead to multiple slices of images, each consisting of thousands of voxels observed every 1–3 s with hundreds of time points for each subject. In addition, the voxel time course signals are complex-valued due to phase imperfections after Fourier encoding and inverse Fourier image reconstruction, leading to complex-valued fMRI (CV-fMRI) data consisting of magnitude and phase, or real and imaginary components.

Tools for CV-fMRI analysis have been proposed in the literature. Rowe and Logan (2004) and Rowe (2005a, 2005b) model the phase directly to estimate the phase angle using a polar coordinates representation, while Lee *et al.* (2007, 2009) and Rowe (2009) use Cartesian representations for modeling CV-fMRI. Autoregressive (AR) complex-valued models have also been developed (Kociuba and Rowe, 2016; Adrian *et al.*, 2018). Recently, Yu *et al.* (2018) proposed Bayesian variable selection tools to improve detecting brain activation at the voxel level, particularly in cases of low signal-to-noise ratio (SNR), when compared to alternative complex-valued and magnitude-only models used in the majority of fMRI analyses. However, Yu *et al.* (2018) did not incorporate a spatial structure that models the fact that neighboring voxels tend to behave similarly and be activated in clusters (Lu *et al.*, 2003; Lazar, 2008).

On the other hand, a wide range of Bayesian spatiotemporal models have been developed for magnitude-only fMRI data. Smith and Fahrmeir (2007) proposed an Ising

prior to spatially smooth the variable selection indicator variables representing whether or not a given voxel is activated. Lee *et al.* (2014) extended the model by including temporal dependence. Bowman *et al.* (2008) considered a two-stage hierarchical model with an AR temporal structure and spatial dependency. Xu *et al.* (2009) proposed a hierarchical spatial model to draw inferences at the population, individual, and voxel levels, and they used Gaussian mixtures to infer activations. Zhang *et al.* (2014) used a Markov random field prior to detect activations in a nonparametric way. A number of Bayesian spatiotemporal approaches are summarized in Zhang *et al.* (2015, 2016) and Chiang *et al.* (2017). Recently, Bezener *et al.* (2018) proposed a Bayesian model for detecting activation that incorporates spatial dependence by areal parcellation of voxels. All these models are well constructed, but they only use the magnitude information and assume normally distributed errors, which may be problematic in practice. As shown in Gudbjartsson and Patz (1995) and Rowe and Logan (2004) , if both the real and imaginary parts of CV-fMRI signals have independent Gaussian errors with the same variance, the magnitude signals follow a Ricean distribution that is approximately normal only for large SNRs. Magnitude-only models have been shown to have lower power than complex-valued models when the normality is not well approximated (Rowe and Logan, 2004). In practice, SNRs may not be large enough to hold this approximate normality, particularly in cases of images with higher voxel resolutions.

We propose a flexible and computationally feasible Bayesian approach combining variable selection tools and a low-rank spatial kernel convolution (KC) for detecting activation in CV-fMRI. A complex-valued AR structure is used for modeling temporal dependence on the errors at the observational level. The KC spatial structure is added to the underlying indicator variables that determine voxel activation.

Our approach infers the parameters of spatial kernel functions, providing a flexible framework for determining the strength and range of the spatial dependency across voxels in the context of brain activation. The kernel-based approach utilizes a small set of spatial locations to achieve dimension reduction. Unlike other approaches based on areal parcellations, it does not require assigning voxels to specific parcels. Instead, our model is able to discover the underlying spatial structure of the activation process by convolving a latent process over a small set of locations with voxel and region-dependent kernel functions. Such latent structure leads to dimension reduction and efficient posterior inference. Kernel convolution provides increased spatial flexibility by capturing potential dependencies across neighboring voxels in the same region, as well as those between voxels in brain regions that are farther away.

We refer to our general model as the complex-valued kernel convolution with AR error (CV-KC-AR) model. We also develop a complex-valued model with a spatial Gaussian process (GP) structure based on areal parcellation with AR noise, referred to as CV-GP-AR. We show that both spatial models outperform the complex-valued nonspatial model of Yu *et al.* (2018) in terms of activation detection. We also show that CV-KC-AR outperforms CV-GP-AR by providing more accurate posterior probability activation maps and less false positives. Using a KC spatial structure increases the detection power when activation strength is weak. Kernel-based spatial structures eliminate or significantly reduce the number of isolated voxels incorrectly labeled as activated by encouraging clusters of brain activation while avoiding increased false positives, in contrast to other magnitude-only spatial approaches implemented in common software packages (Eklund *et al.*, 2016).

The paper is organized as follows. Section 2 describes the CV-KC-AR and CV-GP-AR models. Section 3 discusses the algorithms for posterior inference. Section 4 illustrates the performance of the proposed methods in simulation studies. Section 5 discusses the analysis of a human CV-fMRI dataset, and Section 6 presents a discussion and future extensions.

2 | BAYESIAN SPATIOTEMPORAL MODELS FOR BRAIN ACTIVATION IN CV-fMRI

We consider a spatiotemporal model for detecting brain activation at the voxel level in CV-fMRI. The temporal dependence is captured by an AR(1) noise process that has been shown to be sufficient, effective, and well balanced between detecting performance and computational efficiency (Penny *et al.*, 2003; Lindquist, 2008; Lee *et al.*, 2014). A spatial dependence structure is induced on activation probabilities by means of voxel-specific spatial random effects that are expressed as a convolution of a kernel function and Gaussian variables at a coarse set of locations.

2.1 | Kernel convolution methods

Kernel convolution has been widely used for spatial modeling in environmental sciences and geostatistics (Higdon, 1998; Fuentes, 2002; Higdon, 2002; Lee *et al.*, 2005). Consider a kernel function $k(s;\phi)$, with $s \in S$ a spatial location and parameter ϕ. Let $w(s)$ be a white noise process with mean 0, variance σ^2, and $\mathrm{Cov}(w(s), w(s')) = 0$.

Then, the (kernel-convoluted) process $x(s) = \int_S k(s-u;\phi)w(u)du$ is a stationary Gaussian process that has mean zero, $\mathrm{Var}(x(s)) = \sigma^2 \int_S k^2(s-u;\phi)du$ and $c(h) = \mathrm{Cov}(x(s), x(s')) = \sigma^2 \int_S k(t;\phi)k(t-h;\phi)dt$, $h = s - s'$. In practice, a finite sum approximation to $x(s)$ is applied, that is, for a set of sites $u_1, \ldots, u_D$ in $\mathcal{S}$, the process is defined as $x(s) = \sum_{d=1}^{D} k(s-u_d;\phi)w(u_d)$. This provides dimension reduction and allows us to simplify the computational burden for posterior inference, since a relatively small number D will effectively capture the spatial process $x(s)$.

A popular kernel is the Gaussian radial basis function due to its analytical and computational properties; however, its covariance function and corresponding random field are infinitely smooth, which is not realistic for fMRI. We use the Bezier kernels (Brenning, 2001):

$$k(z_\upsilon - u_d;\phi) = \left(1 - \frac{\|z_\upsilon - u_d\|^2}{\phi^2}\right)^{\nu}, \quad \|z_\upsilon - u_d\| < \phi, \tag{1}$$

with $\nu > 0$ a smoothing parameter and range parameter $\phi > 0$. In fMRI analysis, z_υ is the location of voxel υ, and u_d can be seen as the location of the spatial region d in the image. The representation in (1) includes commonly used kernels such as the parabolic ($\nu = 1$), quartic (biweight) ($\nu = 2$), or the triweight kernel ($\nu = 3$) up to some constant. Bezier kernels have compact support. This avoids unrealistically relating any two voxels in the image and does not require one to prespecify a neighboring structure. Instead the model learns the neighboring structure from the data by inferring the range parameter ϕ. A comparison between Bezier and Gaussian kernels is provided in Supporting Information.

2.2 | A complex-valued linear model for task-based CV-fMRI

Without loss of generality, we assume that the fMRI signals have been centered and detrended and that $x_{j,t}$ corresponds to the discretized convolution of the on-and-off signal of the stimulus sequence for task j, $j = 1, \ldots, p$, with a given hemodynamic response function (HRF) that models the hemodynamic delay in the magnetic resonance signal (Friston *et al.*, 2007). Following Yu *et al.* (2018), let $y_t^\upsilon = y_{t,Re}^\upsilon + i y_{t,Im}^\upsilon \in \mathbb{C}$ be the CV-fMRI measurement at time t and voxel υ, for $t = 1, \ldots, T$ and $\upsilon = 1, \ldots, V$. We consider a model of the form $\mathbf{y}^\upsilon = \mathbf{X}\boldsymbol{\gamma}^\upsilon + \boldsymbol{\eta}^\upsilon$, with $\mathbf{y}^\upsilon = (y_1^\upsilon, \ldots, y_T^\upsilon)'$, $\boldsymbol{\gamma}^\upsilon = \boldsymbol{\gamma}_{Re}^\upsilon + i\boldsymbol{\gamma}_{Im}^\upsilon$, where $\boldsymbol{\gamma}_{Re}^\upsilon = (\gamma_{Re,1}^\upsilon, \ldots, \gamma_{Re,p}^\upsilon)'$, and $\boldsymbol{\gamma}_{Im}^\upsilon = (\gamma_{Im,1}^\upsilon, \ldots, \gamma_{Im,p}^\upsilon)'$, and with $\mathbf{X} = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T')'$, where $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{p,t})'$. Models that infer voxel or region-specific HRFs (see, e.g., Rowe, 2001; Yu *et al.*, 2016) can be considered, but this significantly increases computational cost. Instead, we assume that the HRF is known for all t and υ and study the sensitivity of our posterior results with respect to changes in the HRF. Overall, we found no major differences across the results under different HRFs if we optimize the parameters that define each particular HRF class.

Coefficients $\gamma_j^\upsilon = \gamma_{Re,j}^\upsilon + i\gamma_{Im,j}^\upsilon$ represent activation amplitude or strength of voxel υ under task j. We view the identification of activated voxels as a variable selection problem and adopt a *spike-and-slab prior* on γ_j^υ, that is, $\gamma_j^\upsilon \sim \psi_j^\upsilon g(\gamma_j^\upsilon) + (1-\psi_j^\upsilon)I(\gamma_j^\upsilon = 0)$, where $g(\cdot)$ denotes a nonsingular complex-valued distribution. The variable ψ_j^υ is a binary variable, with $\psi_j^\upsilon = 0$ indicating that voxel υ is not activated under task j leading to $\gamma_j^\upsilon = 0$ and $\psi_j^\upsilon = 1$ indicating that the voxel is active under task j. Let $\boldsymbol{\psi}^\upsilon = (\psi_1^\upsilon, \ldots, \psi_p^\upsilon)'$. Then $\boldsymbol{\gamma}^\upsilon(\boldsymbol{\psi}^\upsilon)$ is the vector of nonzero coefficients from $\boldsymbol{\gamma}^\upsilon$ and $\mathbf{X}(\boldsymbol{\psi}^\upsilon)$ is the corresponding design matrix. Therefore, the model can be rewritten to emphasize the explicit dependence on the indicator variables as

$$\mathbf{y}^\upsilon = \mathbf{X}(\boldsymbol{\psi}^\upsilon)\boldsymbol{\gamma}^\upsilon(\boldsymbol{\psi}^\upsilon) + \boldsymbol{\eta}^\upsilon, \quad \boldsymbol{\eta}^\upsilon \overset{\text{iid}}{\sim} CN_T(0, 2\sigma_\upsilon^2\boldsymbol{\Lambda}_\upsilon, 0). \tag{2}$$

$CN_L(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$ denotes a complex-normal distribution of dimension L with complex-valued mean $\boldsymbol{\mu}$, Hermitian and nonnegative definite covariance matrix $\boldsymbol{\Gamma}$, and complex-valued symmetric relation matrix $\mathbf{C}$. Note that the dimensions of $\mathbf{X}(\boldsymbol{\psi}^\upsilon)$ and $\boldsymbol{\gamma}^\upsilon(\boldsymbol{\psi}^\upsilon)$ vary according to the nonzero elements in $\boldsymbol{\psi}^\upsilon$, that is, the columns of $\mathbf{X}$ and elements of $\boldsymbol{\gamma}^\upsilon$ are only those corresponding to nonzero ψ_j^υs, and so the notation $\mathbf{X}(\boldsymbol{\psi}^\upsilon)$ and $\boldsymbol{\gamma}^\upsilon(\boldsymbol{\psi}^\upsilon)$ above emphasizes that $\mathbf{X}$ and $\boldsymbol{\gamma}^\upsilon$ are functions of $\boldsymbol{\psi}^\upsilon$. However, for simplicity, we use $\mathbf{X}_\upsilon$ and $\boldsymbol{\gamma}^\upsilon$ instead. We assume that the noise vector $\boldsymbol{\eta}^\upsilon = (\eta_1^\upsilon, \ldots, \eta_T^\upsilon)'$ follows a complex-valued AR(1) structure with autocorrelation matrix $\boldsymbol{\Lambda}_\upsilon$ such that the real and imaginary parts have the same AR coefficient, φ_υ, and they are independent a priori through a circular normal structure. We further assume that the φ_υs are voxel specific and use empirical Bayes estimates, $\hat{\varphi}_\upsilon$s, to reduce computational burden as in Bezener *et al.* (2018), with $\hat{\varphi}_\upsilon \in (-1, 1)$ for stationarity.

Below we focus on the prior for the indicators $\boldsymbol{\psi}^\upsilon$, as spatial dependence is introduced through these binary variables. For the remaining parameters, we choose prior distributions that optimize the computational efficiency of the posterior inference. More specifically, we assume $\pi(\boldsymbol{\sigma}^2) = \pi(\sigma_1^2, \ldots, \sigma_V^2) = \prod_{\upsilon=1}^{V} \pi(\sigma_\upsilon^2)$, with $\sigma_\upsilon^2 \overset{\text{iid}}{\sim} IG(a_\sigma, b_\sigma)$. In practice, we use a popular weakly informative prior $IG(1/2, 1/2)$. We use a complex circular normal g-prior on $\boldsymbol{\gamma}^\upsilon$ and with $g = T$ (the length of the time course of the CV-fMRI signal) for both, the real and imaginary

parts, which results in

$$\boldsymbol{\gamma}^{\upsilon} \mid \boldsymbol{\psi}^{\upsilon}, \sigma_{\upsilon}^2 \overset{\text{ind}}{\sim} CN\left(\hat{\boldsymbol{\gamma}}^{\upsilon}, 2T\sigma_{\upsilon}^2\left(\mathbf{X}_{\upsilon}'\hat{\boldsymbol{\Lambda}}_{\upsilon}^{-1}\mathbf{X}_{\upsilon}\right)^{-1}, 0\right), \tag{3}$$

with $\hat{\boldsymbol{\gamma}}^{\upsilon} = (\mathbf{X}_{\upsilon}'\hat{\boldsymbol{\Lambda}}_{\upsilon}^{-1}\mathbf{X}_{\upsilon})^{-1}\mathbf{X}_{\upsilon}'\hat{\boldsymbol{\Lambda}}_{\upsilon}^{-1}\mathbf{y}^{\upsilon}$. The dimension of the prior equals the number of nonzero indicators ψ_j^{υ}s, with $\gamma_j^{\upsilon} = 0$ if $\psi_j^{\upsilon} = 0$, as explained above. This circular complex normal prior allows $\boldsymbol{\gamma}^{\upsilon}$ to be easily integrated out and hence makes the computation more efficient. It corresponds to a common unit information prior, that is, a weakly informative prior that contains the same amount of information that would be contained in a single observation.

2.3 | Hierarchical spatial priors on $\boldsymbol{\psi}^{\upsilon}$

Due to the large number of voxels in the images, it is computationally challenging to perform spatial modeling at the voxel level. To overcome this challenge, we construct a latent spatial random process based on $D \ll V$ locations that are generally but not necessarily, chosen to be equally spaced along a coarse grid over the full high-resolution image. In general, a larger D leads to better detecting performance but also imposes a heavier computation. Thus, it is important to select a prior that is relatively insensitive to the choice of D. This can be achieved by specifying the spatial structure via kernel convolution as described below.

Let $\boldsymbol{\psi}_{(j)}$ be the set of indicator variables for task j and all the voxels υ, that is, $\boldsymbol{\psi}_{(j)} = (\psi_j^1, \ldots, \psi_j^V)$. Define $\boldsymbol{S}_{(j)} = (S_j^1, \ldots, S_j^V)$ as a collection of *voxel-specific* spatial random effects for task j. The spatial prior for this model assumes that the indicator variables are conditionally independent given the spatial random effects as follows:

$$\pi\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{S}_{(j)}\right) = \prod_{\upsilon=1}^{V} \pi\left(\psi_j^{\upsilon} \mid S_j^{\upsilon}\right),$$

$$\text{where } \psi_j^{\upsilon} \mid S_j^{\upsilon} \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{1+e^{-S_j^{\upsilon}}}\right). \tag{4}$$

Let w_j^d be the latent spatial random effect for task j and site d, $z_1, \ldots, z_V$ denote the locations of the voxels, and $k(z_{\upsilon} - s_d; \boldsymbol{\phi})$ denotes a specific kernel function. Then, the voxel-specific spatial random effects in $\boldsymbol{S}_{(j)}$ are modeled as $S_j^{\upsilon} = \sum_{d=1}^{D} k(z_{\upsilon} - s_d; \boldsymbol{\phi}) w_j^d$, where s_d is the location of site d. The latent low-resolution spatial effects w_j^d are assumed to be Gaussian such that $w_j^d \mid \tau_j^2 \overset{\text{ind}}{\sim} N(0, \tau_j^2)$, and $\tau_j^2 \overset{\text{iid}}{\sim} IG(a_{\tau}, b_{\tau})$, or equivalently, $\boldsymbol{S}_{(j)} = \mathbf{K}(\boldsymbol{\phi})\mathbf{w}_{(j)}$ follows $N(0, \tau_j^2\mathbf{K}(\boldsymbol{\phi})\mathbf{K}(\boldsymbol{\phi})')$, where $\mathbf{K}(\boldsymbol{\phi}) = [k(z_{\upsilon} - s_d; \boldsymbol{\phi})]$ is the $V \times D$ matrix whose (υ, d)-th element is $k(z_{\upsilon} - s_d; \boldsymbol{\phi})$ and $\mathbf{w}_{(j)} = (w_j^1, \ldots, w_j^D)$. The parameter $\boldsymbol{\phi}$ could be a common scalar for all regions or a region-specific vector, $\boldsymbol{\phi} = (\phi^1, \ldots, \phi^D)$. It controls the shape of kernels and the smoothness of activation maps. Different priors can be assigned depending on whether $\boldsymbol{\phi}$ is a scalar or a vector. For example, $\phi^d \overset{\text{iid}}{\sim} Ga(a_{\phi}, b_{\phi})$. Region-specific a_{ϕ^d} and b_{ϕ^d} can also be considered if one has any prior information about how the spatial effect w^d contributes to the spatial structure in other regions.

Another alternative is to use a GP spatial prior structure summarized as

$$\pi\left(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{S}_{(j)}\right) = \prod_{g=1}^{G} \prod_{\upsilon \in \mathcal{R}_g} \pi\left(\psi_j^{\upsilon} \mid S_j^g\right),$$

$$\psi_j^{\upsilon} \mid S_j^g \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{1+e^{-S_j^g}}\right), \tag{5}$$

$$\left(\boldsymbol{S}_{(j)} \mid \delta_j^2, r_j\right) \overset{\text{ind}}{\sim} N\left(0, \delta_j^2\Gamma_j(\cdot, \cdot \mid r_j)\right),$$

$$\Gamma_j(i, k \mid r_j) = \exp\left(-\frac{\|s_i - s_k\|}{r_j}\right), \tag{6}$$

$$\pi\left(\delta_j^2\right) \propto \delta_j^{-2},\ r_j \sim Ga(a_j, b_j). \tag{7}$$

This model requires the specification of G spatial regions, each denoted as $\mathcal{R}_g$, and a preassignment of voxels to these regions. Furthermore, voxels in the same region share the same region-level spatial effect S_j and hence the same probability of activation. The KC model achieves spatial dependence by relating $\psi_j^1, \ldots, \psi_j^V$ via kernel convolution on $\boldsymbol{S}_{(j)}$, which leads to dimension reduction, as only a small number of latent spatial effects are needed to describe the underlying spatial process. The KC model can capture dependence between voxels that are spatially close as well as those from distant regions. Each voxel is allowed to borrow information from any other voxel through the convolution process, leading to higher resolution latent spatial effects and more precise posterior probability activation maps, as illustrated in simulations and human data analyses. In fact, posterior probability activation maps obtained from the KC model do not display the "probability inflation" commonly seen in the GP model. In the GP model when a given region contains a nonnegligible number of activated voxels, other nonactivated ones in the region will get artificially inflated probabilities of activation. On the other hand, two voxels next to each other may have significantly different

probabilities of activation due to their different region assignment. This usually occurs when the correlation between regions decays fast, and it highly depends on how large G is, and on how similar is the behavior of voxels that belong to the same region. The KC model does not have this limitation because of the "spillover effect" of the convolution structure. Thus, as shown below, while the GP method is very sensitive to the parcellation and the number of regions G, the KC approach is much less sensitive to the number of latent spatial sites D.

3 | POSTERIOR INFERENCE

In this section, we focus on posterior inference for the CV-KC-AR model via a Markov chain Monte Carlo (MCMC) scheme. The algorithm for posterior sampling in CV-GP-AR is similar to this and generalizes the algorithm of the magnitude-only GP model in Bezener *et al.* (2018).

Let $\mathbf{y} = (\mathbf{y}^1, \dots, \mathbf{y}^V)$ be the observed complex-valued image, $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\boldsymbol{\psi}) = (\boldsymbol{\gamma}^1(\boldsymbol{\psi}^1), \dots, \boldsymbol{\gamma}^V(\boldsymbol{\psi}^V))$, $\boldsymbol{\psi} = (\boldsymbol{\psi}_{(1)}, \dots, \boldsymbol{\psi}_{(p)})$, $\mathbf{w} = (\mathbf{w}_{(1)}, \dots, \mathbf{w}_{(p)})$, $\mathbf{w}_{(j)} = (w_j^1, \dots, w_j^D)$, $\boldsymbol{\tau}^2 = (\tau_1^2, \dots, \tau_p^2)$, the model parameters, and $\boldsymbol{S}_{(j)} = \mathbf{K}(\boldsymbol{\phi})\mathbf{w}_{(j)}$. Without loss of generality, we assume that the smoothness parameter ν is fixed with $\nu = 2$. This value generally provided the best performance in our simulation studies, but we also found that posterior results were not very sensitive to the choice of ν for values of $\nu \leq 5$ (see the Supporting Information). We assume that $\boldsymbol{\phi}$ is a scalar with $\boldsymbol{\phi} > 0$. Then, the posterior density of the CV-KC model is given by

$$q(\boldsymbol{\gamma}, \boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\tau}^2, \boldsymbol{\sigma}^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\gamma}, \boldsymbol{\psi}, \boldsymbol{\sigma}^2)\pi(\boldsymbol{\gamma} \mid \boldsymbol{\psi}, \boldsymbol{\sigma}^2)\pi(\boldsymbol{\sigma}^2)\pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi}) \times \pi(\mathbf{w} \mid \boldsymbol{\tau}^2)\pi(\boldsymbol{\tau}^2)\pi(\boldsymbol{\phi}). \tag{8}$$

This model is high dimensional with $V(2p+1) + Dp + p + 1$ parameters, even for single-subject analysis. To further improve computational efficiency, we take advantage of the conjugate priors and integrate out $\boldsymbol{\gamma}, \boldsymbol{\sigma}^2$, and $\boldsymbol{\tau}^2$ to obtain the marginal posterior distribution

$$q(\boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\phi} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\psi})\pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi})\pi(\mathbf{w})\pi(\boldsymbol{\phi}). \tag{9}$$

Let $\tilde{\mathbf{y}}^\upsilon = (y_{1,Re}^\upsilon, \dots, y_{T,Re}^\upsilon, y_{1,Im}^\upsilon, \dots, y_{T,Im}^\upsilon)'$, $\tilde{\boldsymbol{\gamma}}^\upsilon = (\gamma_{1,Re}^\upsilon, \dots, \gamma_{p,Re}^\upsilon, \gamma_{1,Im}^\upsilon, \dots, \gamma_{p,Im}^\upsilon)'$, $\tilde{\mathbf{X}}_\upsilon = \tilde{\mathbf{X}}(\boldsymbol{\psi}^\upsilon) = \text{blockdiag}(\mathbf{X}_\upsilon, \mathbf{X}_\upsilon)$, and $\tilde{\boldsymbol{\Lambda}}_\upsilon = \text{blockdiag}(\hat{\boldsymbol{\Lambda}}_\upsilon, \hat{\boldsymbol{\Lambda}}_\upsilon)$ whose elements are a function of $\hat{\varphi}_\upsilon$ which is fixed at its empirical Bayes estimate obtained as follows. We first fit the model $\mathbf{y}^\upsilon = \mathbf{X}_\upsilon\boldsymbol{\gamma}^\upsilon + \boldsymbol{\eta}^\upsilon$ with an IID noise assumption to obtain $\hat{\boldsymbol{\gamma}}^\upsilon$ and compute the empirical residual $\hat{\boldsymbol{\eta}}^\upsilon = \mathbf{y}^\upsilon - \mathbf{X}_\upsilon\hat{\boldsymbol{\gamma}}^\upsilon$. The empirical Bayes estimate $\hat{\varphi}_\upsilon$ is then given by $\hat{\varphi}_\upsilon = \frac{\sum_{t=2}^T Re((\eta_t^\upsilon)^*(\eta_{t-1}^\upsilon))}{\sum_{t=2}^T |\eta_{t-1}^\upsilon|^2}$ (Yu *et al.*, 2018).

In the circular normal case, after integrating $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}^2$ out, the real representation of the marginal $p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi})$ in the complex-valued model becomes

$$\begin{aligned} p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi}) &= \int p(\tilde{\mathbf{y}} \mid \tilde{\boldsymbol{\gamma}}, \boldsymbol{\psi}, \boldsymbol{\sigma}^2)\pi(\tilde{\boldsymbol{\gamma}} \mid \boldsymbol{\psi}, \boldsymbol{\sigma}^2)\pi(\boldsymbol{\sigma}^2)d\tilde{\boldsymbol{\gamma}}d\boldsymbol{\sigma}^2 \\ &= \prod_{\upsilon=1}^V (1+T)^{-q_\upsilon}|\tilde{\boldsymbol{\Lambda}}_\upsilon|^{-1/2}M(\boldsymbol{\psi}^\upsilon)^{-(T+a_\sigma)}, \end{aligned} \tag{10}$$

with $q_\upsilon = \sum_{j=1}^p \psi_j^\upsilon$ and $M(\boldsymbol{\psi}^\upsilon) = ((\tilde{\mathbf{y}}^\upsilon)'\tilde{\boldsymbol{\Lambda}}_\upsilon^{-1}\tilde{\mathbf{y}}^\upsilon - (\tilde{\mathbf{y}}^\upsilon)'\tilde{\boldsymbol{\Lambda}}_\upsilon^{-1}\tilde{\mathbf{X}}_\upsilon(\tilde{\mathbf{X}}_\upsilon'\tilde{\boldsymbol{\Lambda}}_\upsilon^{-1}\tilde{\mathbf{X}}_\upsilon)^{-1}\tilde{\mathbf{X}}_\upsilon'\tilde{\boldsymbol{\Lambda}}_\upsilon^{-1}\tilde{\mathbf{y}}^\upsilon + 2b_\sigma)$. Then, the marginal posterior is $q(\boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\tau}^2, \boldsymbol{\phi} \mid \mathbf{y}) \propto p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi})\pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi})\pi(\mathbf{w} \mid \boldsymbol{\tau}^2)\pi(\boldsymbol{\tau}^2)\pi(\boldsymbol{\phi})$. Each τ_j^2 can be integrated out to arrive at the marginal distribution of $\mathbf{w}_{(j)}$:

$$\begin{aligned} \pi(\mathbf{w}_{(j)}) &= \int \pi(\mathbf{w}_{(j)} \mid \tau_j^2)\pi(\tau_j^2 \mid a_\tau, b_\tau)d\tau_j^2 \\ &= \int \prod_{d=1}^D N(w_j^d \mid \tau_j^2)IG(\tau_j^2 \mid a_\tau, b_\tau)d\tau_j^2 \\ &= \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)} \times \frac{\Gamma(D/2 + a_\tau)}{\left(\frac{1}{2}\sum_{d=1}^D (w_j^d)^2 + b_\tau\right)^{(D/2+a_\tau)}}. \end{aligned} \tag{11}$$

As a result, we now have $q(\boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\phi} \mid \mathbf{y}) \propto p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi})\pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi})\pi(\mathbf{w})\pi(\boldsymbol{\phi})$ and can propose an MCMC algorithm to sample from this distribution. First, note that the conditional distribution for $\boldsymbol{\psi}$ is

$$q(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi}, \mathbf{y}) \propto \pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi})p(\tilde{\mathbf{y}} \mid \boldsymbol{\psi}) \propto \pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi}) \prod_{\upsilon=1}^V (1+T)^{-q_\upsilon}M(\boldsymbol{\psi}^\upsilon)^{-(T+a_\sigma)}, \tag{12}$$

with $\pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi}) = \prod_{j=1}^p \pi(\boldsymbol{\psi}_{(j)} \mid \boldsymbol{\phi}, \mathbf{w}_{(j)}) = \prod_{j=1}^p \prod_{\upsilon=1}^V \pi(\psi_j^\upsilon \mid \mathbf{K}_\upsilon'(\boldsymbol{\phi})\mathbf{w}_{(j)})$, where $\mathbf{K}_\upsilon(\boldsymbol{\phi})$ is the υth row vector of $\mathbf{K}$. Similarly, we obtain the conditional distributions for $\mathbf{w}$ and $\boldsymbol{\phi}$ from $q(\mathbf{w} \mid \boldsymbol{\psi}, \boldsymbol{\phi}, \mathbf{y}) \propto \pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi})\pi(\mathbf{w})$ and $q(\boldsymbol{\phi} \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}) \propto \pi(\boldsymbol{\psi} \mid \mathbf{w}, \boldsymbol{\phi})\pi(\boldsymbol{\phi})$.

The marginal posterior $q(\boldsymbol{\psi}, \mathbf{w}, \boldsymbol{\phi} \mid \mathbf{y})$ has dimension $p(V+D)+1$. We generate its posterior samples sequentially by sampling $\mathbf{w}$, $\boldsymbol{\psi}$, and $\boldsymbol{\phi}$. Random walk normal proposals with an adaptive variance are used for sampling w_j^d and $\log(\boldsymbol{\phi})$. The MCMC algorithm is summarized in

ALGORITHM 1 MCMC algorithm for CV-KC-AR

1. Obtain empirical Bayes estimates $\{\hat{\varphi}_v\}_{v=1}^{V}$.
2. Set initial values $\mathbf{w}^{(0)}$, $\boldsymbol{\psi}^{(0)}$, $\phi^{(0)}$
3. **for** $i = 1, 2, \ldots$ *until convergence* **do**
   - Draw $(w_j^d)^{(i)}$ in blocks using Metropolis-Hasting steps on its conditional density $q(\mathbf{w} \mid \boldsymbol{\psi}, \phi, \mathbf{y})$
   - Draw $(\psi_j^v)^{(i)}$ from its Bernoulli full conditional distribution $q(\boldsymbol{\psi} \mid \mathbf{w}, \phi, \mathbf{y})$
   - Draw $\phi^{(i)}$ in blocks using Metropolis-Hasting steps on its conditional density $q(\phi \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y})$

end

Result: Posterior samples of $\mathbf{w}$, $\boldsymbol{\psi}$, ϕ.

Algorithm 1, and the detailed sampling scheme is provided in the Supporting Information. Once posterior samples $\{(\boldsymbol{\psi}^{(m)}, \mathbf{w}^{(m)}, \boldsymbol{\phi}^{(m)})\}_{m=1}^{M}$ are obtained, we can estimate the posterior probabilities of activation, $Pr(\psi_j^v = 1 \mid \mathbf{y})$, by computing $\hat{p}_j^v = \frac{1}{M}\sum_{m=1}^{M}(\psi_j^v)^{(m)}$.

We monitor MCMC convergence via the estimated potential scale reduction factor $\widehat{R}$ (Gelman and Rubin, 1992; Brooks and Gelman, 1997). $\hat{R} < 1.1$ is generally used (Gelman *et al.*, 2004). We consider five chains with different initial values and continue simulation runs until all spatial effects $\mathbf{w}$ and ϕ have individual or multivariate $\widehat{R} < 1.1$.

4 | SIMULATION STUDIES

4.1 | Simulation study I

We simulated 50 datasets consisting of 20×20 CV-fMRI slices with a constant baseline signal and a single expected BOLD signal (i.e., $p = 1$). Three activation regions were simulated using the function `specifyregion` in the `R` package `neuRosim` (Welvaert *et al.*, 2011). Both AR and independent error structures are considered. Specifically, for $v = 1, \ldots, 20 \times 20$, and $t = 1, \ldots, 50$, the time series with AR(1) noise for each voxel v were simulated as follows:

$$\begin{aligned} y_{t,Re}^v &= \left(\beta_{v,0} + \beta_{v,1}x_t\right)\cos(\alpha_0) + \eta_{t,Re}^v, \quad \eta_{t,Re}^v \\ &= \varphi\eta_{t-1,Re}^v + \zeta_{t-1,Re}^v, \quad \zeta_{t-1,Re}^v \sim N\left(0, \sigma^2\right), \\ y_{t,Im}^v &= \left(\beta_{v,0} + \beta_{v,1}x_t\right)\sin(\alpha_0) + \eta_{t,Re}^v, \quad \eta_{t,Im}^v \\ &= \varphi\eta_{t-1,Im}^v + \zeta_{t-1,Im}^v, \quad \zeta_{t-1,Im}^v \sim N\left(0, \sigma^2\right), \end{aligned} \tag{13}$$

where x_t is the expected BOLD signal, which is obtained as a convolution of a given HRF and a stimulus indicator function that is on for $1 \le t \le 16$ and off for $16 < t \le 50$ (Lindquist, 2008). For the time series with independent noise, the simulation scheme is the same except that $\eta_{t,Re}^v$ and $\eta_{t,Im}^v$ simply follow $N(0, \sigma^2)$ independently. The expected BOLD signal, true activation map, and a parcellation of the image into $G = 25$ regions are shown in Figure 1. The simulated image and coefficient setting are similar to those in the benchmark example in Bezener *et al.* (2018). $(\beta_{v,0}, \beta_{v,1}) = (300, 2)$ and $(300, 5)$, corresponding to low and high contrast to noise ratios (CNR). We set $\alpha_0 = \pi/4$, $\sigma^2 = 3$ and $\varphi = 0.5$.

4.1.1 | Fitted complex-valued models

We fitted three models to the simulated AR data: CV-KC-AR, CV-GP-AR, and C-EMVS-AR. The C-EMVS-AR model is model (vi) in Section 4 of Yu *et al.* (2018) that assumes voxel-specific variances and AR coefficients. We also fitted CV-KC, CV-GP, and C-EMVS models to data simulated with independent noise ($\varphi_v = 0$) to study performance improvements purely due to the spatial structure of the KC and GP models.

We use the centroids of parcellated regions shown in Figure 1 as our spatial locations for KC models in order to provide a fair comparison to the GP models since this implies $D = G$, but note that KC models do not require us to use locations on a regular grid. Specifically, we considered a CV-KC-AR model with the likelihood (2) and the KC spatial prior described above, with $\tau^2 \sim IG(1/2, 1/2)$,$\sigma_v^2 \overset{\text{iid}}{\sim} IG(1/2, 1/2)$, $\nu = 2$ and $\phi \sim IG(1/2, 1/2)$. The CV-GP-AR model has the same likelihood and the GP prior structure described above, with $r_j \sim \chi_8^2$. CV-KC and CV-GP have the same hyperparameter settings as CV-KC-AR and CV-GP-AR, but assume $\varphi_v = 0$ for all v. Finally, C-EMVS-AR is given by the same likelihood and priors $\gamma^v \mid \psi^v \sim (1 - \psi^v)CN_1(0, 2v_0\sigma_v^2, 0) + \psi^v CN_1(0, 2\sigma_v^2, 0)$,$\sigma_v^2 \sim IG(1/2, 1/2)$,$\psi^v \mid \theta \sim Bernoulli(\theta)$, $\theta \sim Beta(1, 1)$, and $\varphi_v \sim U(-1, 1)$. C-EMVS has the same specification as C-EMVS-AR, but sets $\varphi_v = 0$.

For both KC and GP models, after warming up the first 1000 samples and thinning the sequences by

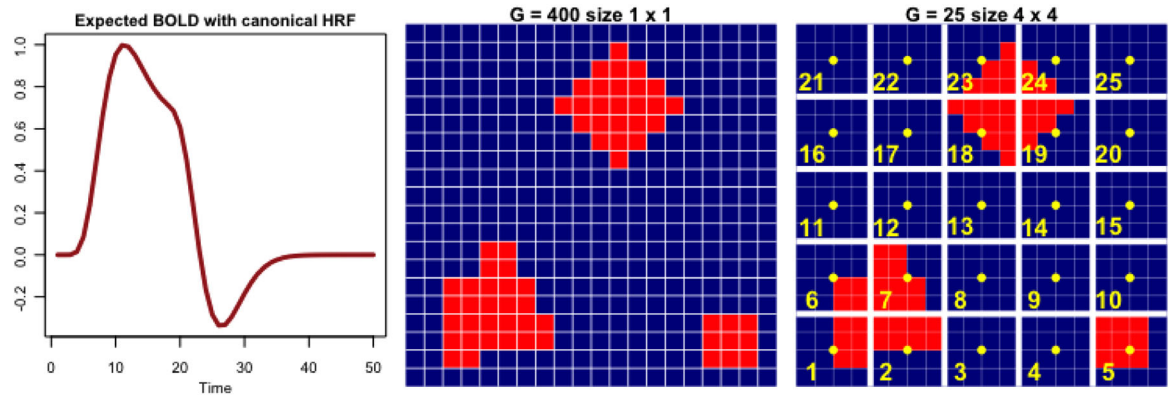


FIGURE 1 Left: Expected BOLD signal with the canonical HRF used in the simulation. Middle: Activation map of a 20×20 image. Right: Activation map with white spatial region grid lines. The number of regions is $G = 25$. Each region is of size 4×4 containing 16 voxels. Yellow dots represent the centroids of regions. Yellow numbers are region numbers. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

keeping every 25th draw, 2000 draws were used for analysis. Such posterior samples have criteria $\widehat{R} < 1.1$ for all spatial effects $\mathbf{w}$ and kernel parameter ϕ, which approximately indicates convergence. The effective sample size of the kernel parameter ϕ is about 400. The EM algorithm in the C-EMVS approach converges when $\|\theta^{(l)} - \theta^{(l-1)}\| < 10^{-3}$, where $\theta^{(l)}$ denotes the full parameter vector at the lth iteration.

We detect activation and construct activation maps from posterior probabilities with voxel v classified as active if $Pr(\psi^v = 1 \mid \mathbf{y}) > t^*$. The threshold $t^* = 0.5$ corresponds to the so-called "median probability model," which has been shown to be the optimal predictive model under certain assumptions for the squared error loss (Barbieri and Berger, 2004). $t^* = 0.8722$ has been used in some Bayesian spatial approaches for magnitude-only data (Smith and Fahrmeir, 2007; Lee *et al.*, 2014). This threshold corresponds to a p-value 0.05 for the likelihood ratio statistic $-2\log((1 - Pr(\psi^v = 1 \mid \mathbf{y}))/Pr(\psi^v = 1 \mid \mathbf{y}))$ (Raftery, 1996). Alternative t^* values based on a Bayesian decision theoretic approach can be used. Müller *et al.* (2004, 2006) show that when a loss function of the form $c\mathbf{E}(FD \mid \mathbf{y}) + \mathbf{E}(FN \mid \mathbf{y})$ is considered, where FD is false discoveries and FN is false negatives, the optimal t^* that minimizes this loss function is $\frac{c}{1+c}$.

4.1.2 | Results of the analysis of the complex-valued simulated data

We consider $t^* = 0.5$ for detecting activation. The activation performance with different t^* values can be examined by looking at the resulting receiver operating characteristic (ROC) curves (see the Supporting Information).

In many applied settings, fMRI signals have been shown to be temporally correlated. Failing to model such temporal structure may result in false activation detection as shown in Figure 2. Without modeling such temporal structure of the simulated data with AR noise, the CV-KC model leads to more false positives and reduced power in activation detection. In fact, although not shown here, its false-positive rate goes up with the strength of temporal correlation. Among the three AR models, CV-KC-AR balances true positives and negatives well and is overall the best. Although CV-KC-AR has lower sensitivity (91.8%) than C-EMVS-AR (95.6%), it greatly reduces false positives and hence has a better overall measure accuracy rate of 98.6%, compared to 98.2% for C-EMVS-AR and 96.9% for CV-GP-AR. In fact, CV-GP-AR has significantly lower sensitivity (82.1%).

To purely focus on the benefits of adding a spatial modeling structure, we fitted nontemporal models to non-AR simulated data. The models CV-KC, CV-GP, and C-EMVS all led to nearly perfect detecting performance when $\beta_{v,1} = 5$ so we focused on the low CNR case with $\beta_{v,1} = 2$. We found that the proposed CV-KC has sensitivity 78.7%, accuracy 86.2%, F1 score 87.3%, compared to 57.7%, 93.3%, and 72.5% for CV-GP and 66.1%, 94.0%, and 77.0% for C-EMVS. We also found that KC models with site-specific ϕ_ds had similar performance measures to KC models with a common ϕ.

The intuition behind the improved performance of KC models is that they lead to a higher resolution latent spatial process S^v, as shown in Figure 3. The posterior effect $\boldsymbol{S}$ is computed via $\mathbf{K}(\tilde{\phi})\tilde{\mathbf{w}}$, where $\tilde{\phi}$ and $\tilde{\mathbf{w}}$ are the posterior means of ϕ and $\mathbf{w}$. The spatial process map correctly identifies three activation areas. The probabilities of activation are obtained from the posterior means of $\boldsymbol{S}$ via logistic transformation. These estimated probabilities of

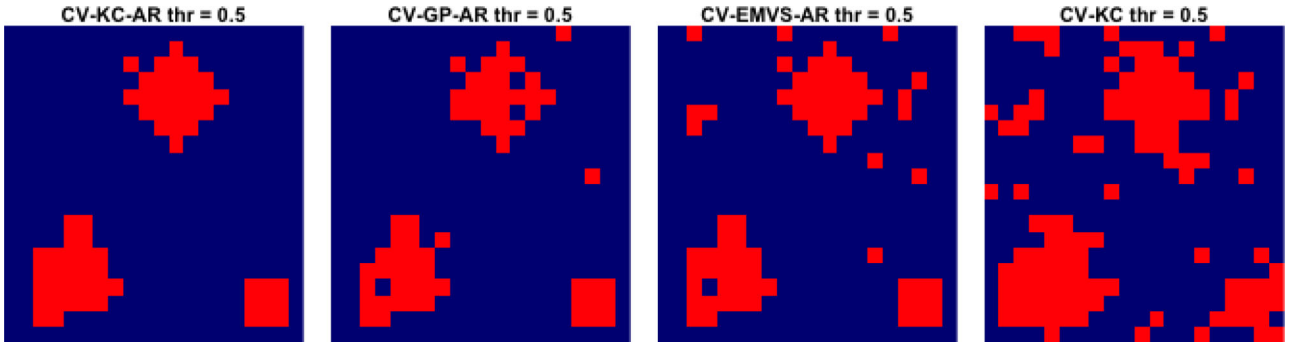
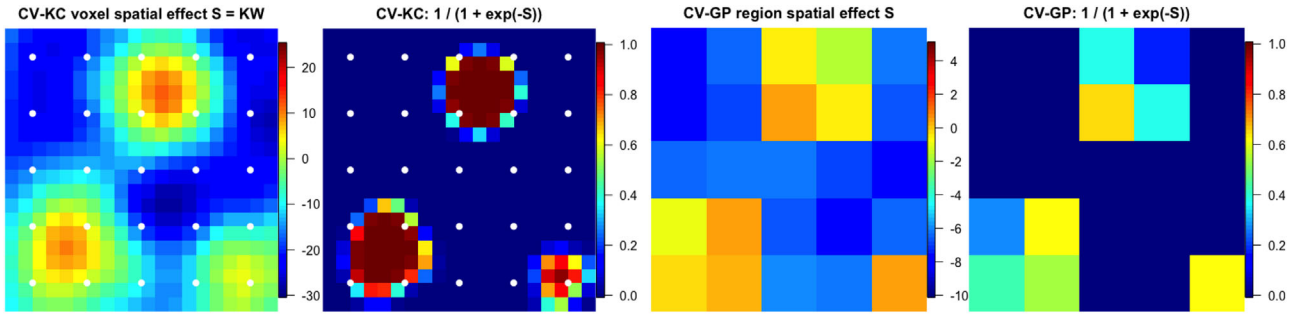


FIGURE 2 Activation plots for the simulated dataset with AR(1) noise. A threshold value of 0.5 was used. Left to right: CV-KC-AR, CV-GP-AR, CV-EMVS-AR, CV-KC. This figure appears in color in the electronic version of this article, and any mention of color refers to that version



(a) CV-KC voxel-level spatial effect (b) CV-GP region-level spatial effect

FIGURE 3 Posterior mean of spatial random effects and corresponding logistic transformation for the CV-KC and CV-GP with $\beta_{v,1} = 5$ (high CNR). White dots represent the location of latent spatial sites used for KC. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

activation are close to one for voxels that are truly activated. On the other hand, as shown in Figure 3B, GP models require voxels in the same region to share the same spatial effect. In fact, these region-specific probabilities can be seen as the overall mean probability of activation in a region. For example, the probability of activation for region 18 is estimated at the posterior mean of S^{18}, $\tilde{S}^{18}$, as $1/(1 + e^{-\tilde{S}^{18}}) \approx 0.626$ which approximates the proportion of activated voxels in the region, $10/16 = 0.625$. Hence, based on the region-specific spatial effects, the GP model says that there are about 10 voxels out of 16 being activated in region 18, but it cannot further specify which voxels are more likely to be activated as there is no additional voxel-level information. Furthermore, the detection performance of GP models is sensitive to the number of regions G and to how the image is parcellated. In contrast, CV-KC models are much less affected by D. As shown in the Supporting Information, a CV-KC with $D = 25$ can lead to the same or even better detection performance than a CV-GP with $G = 100$. Moreover, the computing time of CV-GP grows faster than that of CV-KC as G gets larger. When $G = 100$, CV-GP takes about 2.3 times longer to obtain the same type of activation results obtained by CV-KC with $D = 100$. Another important advantage of the KC approach is that it easily allows for multiresolution analysis (see the Supporting Information).

4.2 | Realistic simulation study: Multiple-slice analysis

We analyze realistic simulated data that imitates the human echo planar volume imaging for whole brain studies. In the majority of fMRI experiments, volume images are measured one slice at a time in an interleaved fashion and so the volume data consist of a collection of two-dimensional (2D) slabs.

The data were generated using a discrete version of the magnetic resonance (MR) signal equation after steady-state magnetization (Karaman *et al.*, 2015). Brains were scanned seven times sequentially along the axial direction to generate seven image slices. Each slice has dimension 96×96 over $T = 490$ time points. There are two 5×5 squares as true activation regions that have positive magnitude contrast (β_1) and phase contrast (α_1). Therefore, two $5 \times 5 \times 5$ cubes are the activation regions of the

whole brain if slices are combined. Slices 1 and 7 have no activated voxels, and their magnitude and phase contrasts are zero. Slice 4 has the largest maximum magnitude and phase CNR with $\text{CNR}_\text{M} = \beta_1/\sigma = 0.5/1$ and $\text{CNR}_\text{P} = \alpha_1/\text{SNR}_\text{M} = (\pi/120)/25$, respectively. Slices 3 and 5 have weaker activations than slice 4, and slices 2 and 6 have the weakest CNR_M and CNR_P. Technical details of the data-generating process are discussed in the Supporting Information.

Before fitting our models, we shrink the image size to 64×64 by removing voxels outside the brain. We set $D = G = 64$ and use the centroids of the GP regions as the KC spatial locations. We fitted CV-KC, MO-KC, CV-GP, and MO-GP to each slice of the image. All model and MCMC settings are the same as those used in the previous simulation.

Since all models have a 100% true negative rate for slices 1 and 7, we focus on detecting power for slices 2–6. The posterior probability maps in Figure 4 show that the CV models outperform MO models and KC models outperform GP models for both complex-valued and magnitude-only data. When activation strengths are weak (slices 2 and 6), activated voxels are barely detected for all models, although KC shows slightly more detection power. However, the KC models detect more true positives than the GP models for slices 3, 4, and 5. If the number of true positives are the same for both KC and GP models, those activated voxels will generally have higher posterior probabilities when the KC model is used. True positive rates averaged across slices with $t^* = (0.5, 0.8722)$ are (0.39, 0.34), (0.38, 0.30), (0.22, 0.14), and (0.16, 0.10) for CV-KC, CV-GP, MO-KC, and MO-GP, respectively (low true positive rates are due to low detecting power of slices 2 and 6), while F1 scores are (0.49, 0.45), (0.49, 0.41), (0.31, 0.21), (0.25, 0.17), respectively. The KC methods are robust to threshold values, and the CV methods dominate the MO approaches.

5 | ANALYSIS OF HUMAN CV-fMRI DATA

We analyze human data recorded during an fMRI experiment performed on a 3.0-T General Electric Signa LX MRI scanner with a gradient-echo echo-planar pulse sequence. Slices within each volume are measured sequentially in an interleaved fashion. The experiment consisted of a unilateral right-handed finger-tapping task performed with a visual cue indicating whether to tap or rest. Therefore, scientifically one should expect to see brain activation in the motor area on the left hemisphere and the supplementary motor area of the blob in the center.

A block-designed experiment with an initial 20 s of rest followed by 16 epochs of 15 s on and off was used. The full dataset is composed of seven 2.5 mm thick axial slices of dimension 96×96 and 510 time points with a volume time to repetition (TR) of 1 s. We demonstrate that our CV-KC method is able to produce activation results that are consistent with previous analyses with the additional benefit of reducing spurious activations, such as detecting activations outside the brain or in regions that are not implicated in the finger-tapping task. We analyze multiple 2D slice images using our methods as this describes the echo-planar image measurement technique. With the TR of 1 s for these data and interleaved slices, adjacent slices are measured half a second apart. Here we present the results of the analysis for a single slice, the same one analyzed in Karaman *et al.* (2014).

We begin by parcellating the image into 36 equal-sized squared spatial regions containing 36 voxels each, and then fit the CV-KC and CV-GP models to the CV-fMRI data from one subject. Figure 5 shows the activation maps derived from the CV-KC, CV-GP, and C-EMVS models under $t^* = 0.5$ for detecting activation. Clearly, spatial models eliminate isolated voxels around the brain margin incorrectly labeled as active by the C-EMVS approach. In addition, spatial modeling encourages activation in clusters of voxels. Both the CV-KC and CV-GP models detect more activated voxels in the expected motor-related brain regions located on the left of the brain. Moreover, with the fact that activation is expected to be in gray matter (GM), and the estimated classification of whether a voxel is GM in Karaman *et al.* (2014), most of our activated voxels are shown to be in GM. These results are biologically plausible because execution of complex cognitive tasks usually involves populations of neurons spanning across many voxels rather than a single voxel. Although not shown here, CV-KC-AR and CV-GP-AR produce almost identical activation maps to CV-KC and CV-GP with CV-KC-AR outperforming CV-KC-AR and C-EMVS-AR.

To focus on a smaller portion of the image that reduces areas outside the brain, an image of size 56×56 was further examined. This image was represented by 64 latent spatial sites for KC analysis, or parcellated into 64 regions each containing 49 voxels for GP analysis. Figure 6 shows the voxel-level spatial effects inferred by the CV-KC model and the region-specific spatial effects inferred by the CV-GP model for this reduced image. The CV-KC model generates smooth and localized maps of the estimated spatial effects that show the strength of the spatial effects adequately concentrated around the active regions. On the other hand, the maps of estimated spatial effects from CV-GP are neither smooth nor localized. We note that for this particular dataset, the activation results are not so different for the CV-KC and CV-GP approaches when t^* values of 0.5 or 0.8722 are used for detecting activation. However, once again, we see that the

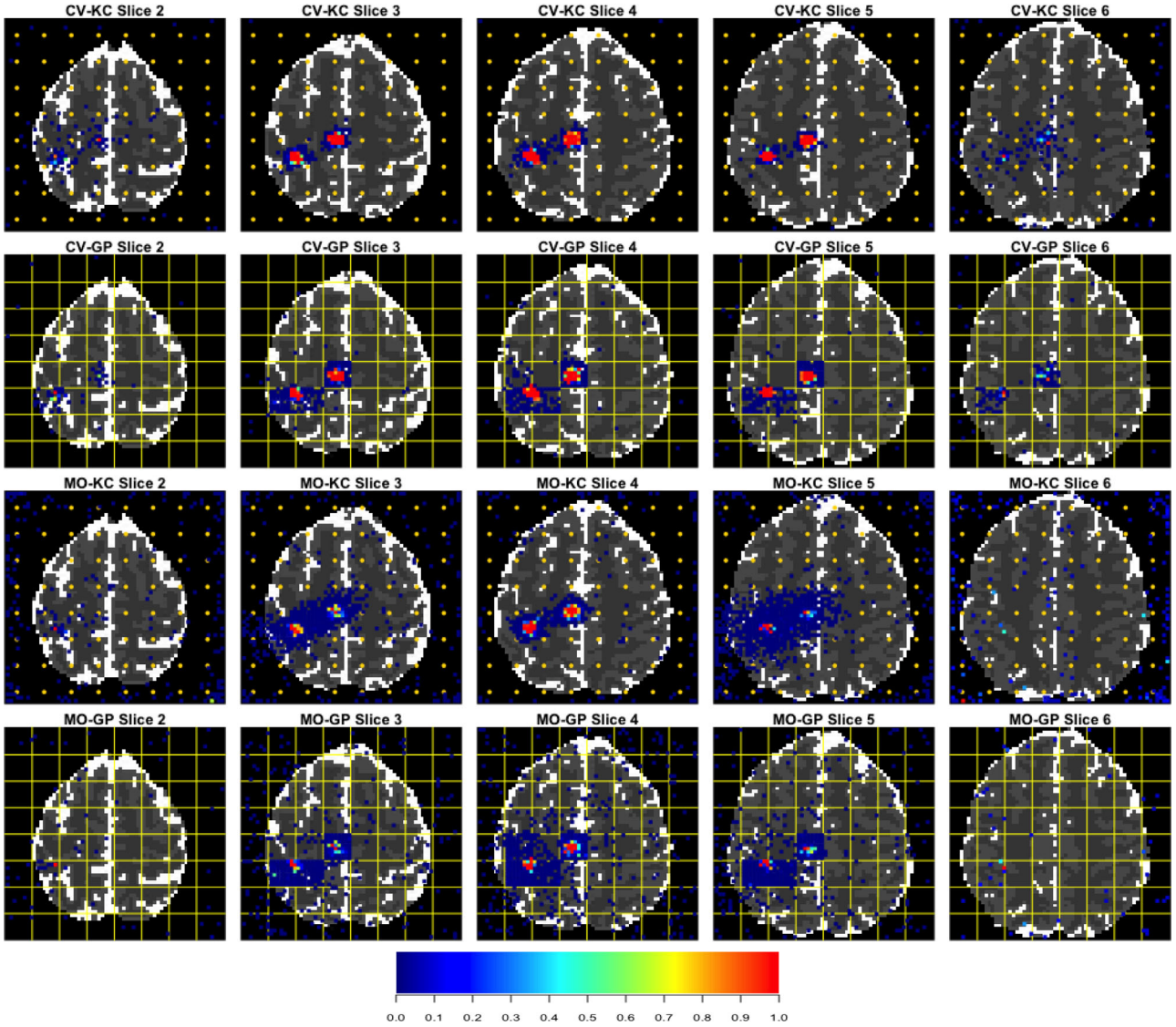


FIGURE 4 Posterior probability maps. From left to right: Slice 2 to slice 6. From top to bottom: CV-KC, CV-GP, MO-KC, MO-GP. The posterior probability colorbar is shown at the bottom. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

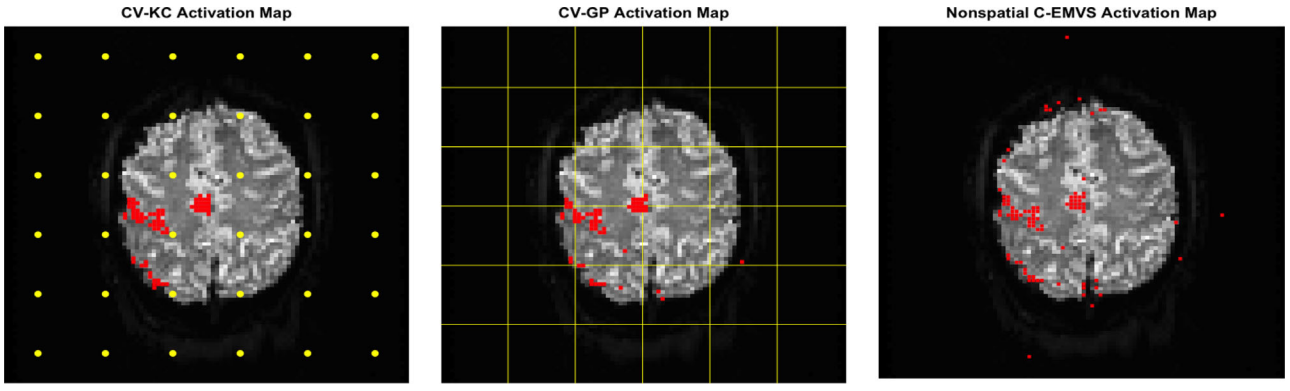
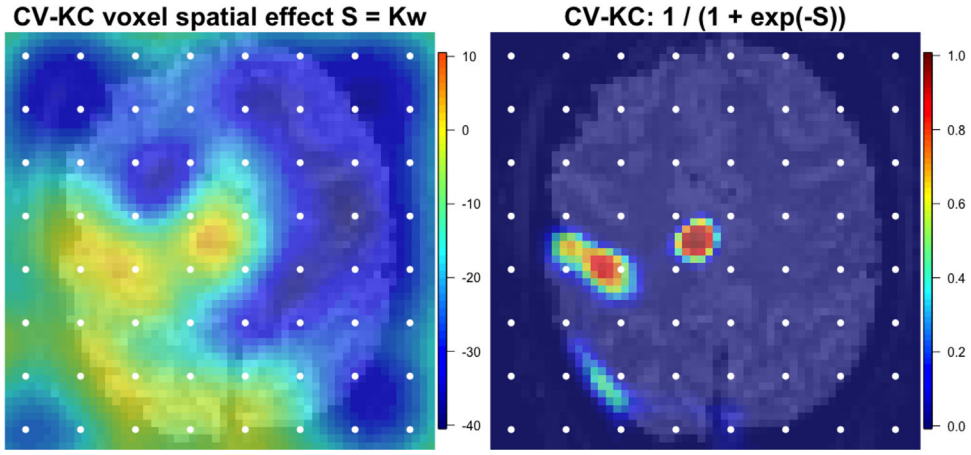
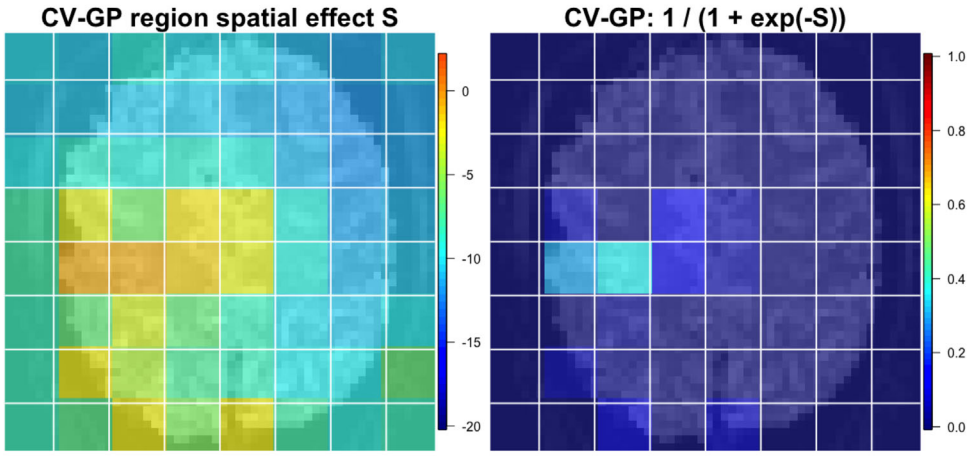


FIGURE 5 Activation of human subject CV-fMRI. Left: Activation map from CV-KC with threshold 0.5. Middle: Activation map from CV-GP with threshold 0.5. Right: Activation map from the nonspatial C-EMVS approach with threshold 0.5. Yellow dots represent the location of latent spatial sites used for KC. This figure appears in color in the electronic version of this article, and any mention of color refers to that version



(a) CV-KC voxel-level spatial effect



(b) CV-GP region-level spatial effect

FIGURE 6 Posterior mean of the spatial random effects and their corresponding logistic transformations for the complex-valued human data: (A) results from the CV-KC model; (B) results from the CV-GP model. White dots represent the location of latent spatial sites used for KC. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

CV-KC model leads to probability maps (in the Supporting Information) that are less uncertain than the maps obtained from the CV-GP model. This is a clear advantage of the CV-KC approach, as more precise posterior probability maps for activation lead to final activation results that are less sensitive to the choice of t^*.

6 | DISCUSSION

We propose a new Bayesian spatiotemporal model via kernel convolution for detecting brain activation from CV-fMRI signals at the voxel-specific level. Our simulation studies show that the Bayesian complex-valued spatial models lead to greater power and less false positives than models that do not include a spatial structure, especially for small CNR.

Assuming the same number of latent spatial effects ($G = D$), our KC model outperforms the GP model by cleverly borrowing spatial information from neighboring voxels. Therefore, the CV-KC approach produces more precise posterior probability maps. Moreover, the CV-KC model is not very sensitive to the number of spatial sites, leading to dimension reduction and decreasing the computational time for posterior inference. The proposed KC models also avoid the need for predetermining the shape of regions and centroid locations for these regions which is potentially a serious problem for the GP models.

Future extensions will consider modeling functional connectivity through additional spatial effects that can include anatomical boundaries and long-distance spatial dependencies. While our proposed method is mainly for single-subject studies, its general spatiotemporal structure lays a foundation of complete CV-fMRI multi-subject

modeling. In a multisubject study, two types of parameter, S_j^v and $a_{(s)}^d$ for subject s, can be considered, with S_j^v a common spatial/group-level effect that borrows information about local dependence from all subjects, and $a_{(s)}^d$ is the subject-level effect that accounts for individual differences.

We strongly recommend the use of CV-fMRI data and models for detecting brain activation. Adding a spatial structure improves detection, but is computationally more costly than using a complex-valued model with no spatial structure. Having this in mind, we advocate for the use of complex-valued models with a spatial kernel convolution. As shown here, these models lead to flexible spatial modeling and more accurate estimated probability activation maps and are less sensitive to the choice of the threshold value that determines activation. Moreover, the CV-KC models lead to dimension reduction and allow us to easily incorporate multiresolution components to obtain more precise results without significantly increasing the computational burden.

DATA AVAILABILITY STATEMENT

The raw data that support the findings in this paper are available from the corresponding author upon reasonable request. The processed data are openly available at https://github.com/chenghanyustats/cv-kc-ar/tree/main/data

ORCID

Cheng-Han Yu https://orcid.org/0000-0002-6977-6400

REFERENCES

Adrian, D.W., Maitra, R., and Rowe, D.B. (2018) Complex-valued time-series modeling for improved accuracy of activation detection in fMRI studies. *Annals of Applied Statistics*, 12, 1451–1478.

Barbieri, M.M. and Berger, J.O. (2004) Optimal predictive model selection. *Annals of Statistics*, 32, 870–897.

Bezener, M., Hughes, J., and Jones, G. (2018) Bayesian spatiotemporal modeling using hierarchical spatial priors, with applications to functional magnetic resonance imaging. *Bayesian Analysis*, 13, 1261–1313.

Bowman, D.F., Caffo, B., Bassett, S.S., and Kilts, C. (2008) A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage*, 39, 146–156.

Brenning, A. (2001) Geostatistics without stationarity assumptions within geographical information systems. *Freiberg Online Geoscience*, 6, 1–108.

Brooks, S.P. and Gelman, A. (1997) General methods for monitoring convergence of iterative simulation. *Journal of Computational and Graphical Statistics*, 7, 434–455.

Chiang, S., Guindani, M., Yeh, H.J., Haneef, Z., Stern, J.M., and Vannucci, M. (2017) A Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Human Brain Mapping*, 38, 1311–1332.

Eklund, A., Nichols, T., and Knutsson, H. (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *PNAS*, 113, 7900–7905.

Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., and Penny, W.D. (2007) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. London: Academic Press.

Fuentes, M. (2002) Spectral methods for nonstationary spatial processes. *Biometrika*, 89, 197–210.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004) *Bayesian Data Analysis*. Boca Raton, FL: CRC press.

Gelman, A. and Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.

Gudbjartsson, H. and Patz, S. (1995) The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34, 910–914.

Higdon, D. (1998) A process-convolution approach to modeling temperatures in the North Atlantic ocean. *Environmental and Ecological Statistics*, 5, 173–190.

Higdon, D. (2002) Space and space-time modeling using process convolutions. In: By A.H. El-Shaarawi, V. Barnett, C.W.A. and P.C. Chatwin (Eds.) *Quantitative Methods for Current Environmental Issues*. Oxford: Springer London, pp. 37–56.

Karaman, M., Bruce, I.P., and Rowe, D.B. (2015) Incorporating relaxivities to more accurately reconstruct MR images. *Magnetic Resonance Imaging*, 33, 374–384.

Karaman, M.M., Bruce, I.P., and Rowe, D.B. (2014) A statistical fMRI model for differential T2* contrast incorporating T1 and T2 of gray matter. *Magnetic Resonance Imaging*, 32, 9–27.

Kociuba, M.C. and Rowe, D.B. (2016) Complex-valued time-series correlation increases sensitivity in fMRI analysis. *Magnetic Resonance Imaging*, 34, 765–770.

Lazar, N.A. (2008) *The Statistical Analysis of Functional MRI Data*. New York: Springer-Verlag.

Lee, H., Higdon, D., Calder, K., and Holloman, C. (2005) Efficient models for correlated data via convolutions of intrinsic processes. *Statistical Modelling*, 5, 53–74.

Lee, K.J., Jones, G.L., Caffo, B.S., and Bassett, S.S. (2014) Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis*, 9, 699–732.

Lee, J., Shahram, M., and Pauly, J.M. (2009) Combining complex signal change in functional MRI. *Magnetic Resonance in Medicine*, 62, 1358–1360.

Lee, J., Shahram, M., Schwartzman, A., and Pauly, J.M. (2007) Complex data analysis in high-resolution SSFP fMRI. *Magnetic Resonance in Medicine*, 57, 905:917.

Lindquist, M.A. (2008) The statistical analysis of fMRI data. *Statistical Science*, 23, 439–464.

Lu, Y., Jiang, T., and Zang, Y. (2003) Region growing method for the analysis of functional MRI data. *NeuroImage*, 20, 455–465.

Müller, P., Parmigiani, G., and Rice, K. (2006) FDR and Bayesian multiple comparisons rule. In: Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M. and West, M. (Eds.) *Bayesian Statistics 8*. Oxford: Oxford University Press, pp. 349–370.

Müller, P., Parmigiani, G., Robert, C., and Rouseau, J. (2004) Optimal sample size for multiple testing: the case of gene expression microarrays. *Journal of American Statistical Association*, 99, 990–1001.

Penny, W., Kiebel, S., and Friston, K. (2003) Variational Bayesian inference for fMRI time series. *NeuroImage*, 19, 727–741.

Raftery, A.E. (1996) Hypothesis testing and model selection. In: Gilks, W.R., Richardson, S. and Spiegelhalter, D. (Eds.) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, pp. 163–188.

Rowe, D.B. (2001) A Bayesian source separation for reference function determination in fMRI. *Magnetic Resonance in Medicine*, 46, 374–378.

Rowe, D.B. (2005a) Modeling both the magnitude and phase of complex-valued fMRI data. *NeuroImage*, 25, 1310–1324.

Rowe, D.B. (2005b) Parameter estimation in the magnitude-only and complex-valued fMRI data models. *NeuroImage*, 25, 1124–1132.

Rowe, D.B. (2009) Magnitude and phase signal detection in complex-valued fMRI data. *Magnetic Resonance in Medicine*, 62, 1356–1357.

Rowe, D.B. and Logan, B.R. (2004) A complex way to compute fMRI activation. *NeuroImage*, 23, 1078–1092.

Smith, M. and Fahrmeir, L. (2007) Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, 102, 417–431.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., and Rosseel, Y. (2011) neurosim: An R package for generating fMRI data. *Journal of Statistical Software*, 44, 1–18.

Xu, L., Johnson, T.D., Nichols, T.E., and Nee, D.E. (2009) Modeling inter-subject variability in activation locations of fMRI data: a Bayesian hierarchical spatial modeling approach. *Biometrics*, 650, 1041–1051.

Yu, Z., Prado, R., Quinlan, E.B., Cramer, S.C., and Ombao, H. (2016) Understanding the impact of stroke on brain motor function: A hierarchical Bayesian approach. *Journal of the American Statistical Association*, 111, 549–563.

Yu, C., Prado, R., Ombao, H., and Rowe, D.B. (2018) A Bayesian variable selection approach yields improved detection of brain activation from complex-valued fMRI. *Journal of American Statistical Association*, 113, 1395–1410.

Zhang, L., Guindani, M., and Vannucci, M. (2015) Bayesian models for fMRI data analysis. *WIREs Computational Statistics*, 7, 21–41.

Zhang, L., Guindani, M., Versace, F., Engelmann, J.M., and Vannucci, M. (2016) A spatio-temporal non-parametric Bayesian model of multi-subject fMRI data. *Annals of Applied Statistics*, 10(2), 638–666.

Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014) A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage*, 95, 162–175.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2.1, 3, 4.1.2 and 4.2 are available with this paper at the Biometrics website on Wiley Online Library. The R code of the proposed method is available online with this paper and also on GitHub at https://github.com/chenghanyustats/cv-kc-ar.

How to cite this article: Yu, C.-H., Prado, R., Ombao, H., Rowe, D. (2023) Bayesian spatiotemporal modeling on complex-valued fMRI signals via kernel convolutions. *Biometrics*, 79, 616–628. https://doi.org/10.1111/biom.13631

Supporting Information for Bayesian Spatiotemporal Modeling on Complex-Valued fMRI Signals via Kernel Convolutions by Yu, C., Prado, R., Ombao, H. and Rowe, D.

Contents

1 MCMC posterior sampling for the CV-KC-AR

The following sampling scheme is for the MCMC algorithm of CV-KC-AR. In practice, we improve the mixing considerably by sampling the parameters in blocks and also use the empirical covariance matrix from posterior samples in previous iterations to generate proposal candidates.

- Sampling $\mathbf{w}$: this involves pD steps. For $j = 1, \ldots, p$ and for each d with $d = 1, \ldots, D$, we have that the conditional density of w_j^d is given by

$$\begin{aligned} q\left(w_j^d \,\middle|\, \mathbf{w}_{(-j)}^{(-d)}, \boldsymbol{\psi}, \phi, \mathbf{y}\right) &\propto \pi\left(\boldsymbol{\psi} \mid \mathbf{w}, \phi\right) \pi\left(\mathbf{w}_{(j)}\right) \\ &\propto \pi\left(w_j^d\right) \prod_{v=1}^{V} \pi\left(\psi_j^v \,\middle|\, S_j^v\left(w_j^g; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-d)}\right)\right), \end{aligned}$$

  where $\mathbf{w}_{(-j)}^{(-d)}$ are all latent spatial random effects w_l^h, $l \neq j$, $h \neq d$. A random walk proposal $\left(w_j^d\right)^* \sim N\left(\left(w_j^d\right)^{(m)}, \sigma^2_{w_j^d}\right)$ is used, where $\left(w_j^d\right)^{(m)}$ is the current state and $\sigma^2_{w_j^d}$ is a tuning parameter. Hence the Metropolis-Hastings ratio is

$$\frac{q\left(\left(w_j^d\right)^* \;\middle|\; w_{-j}^{-d}, \boldsymbol{\psi}, \phi, \mathbf{y}\right)}{q\left(\left(w_j^d\right)^{(m)} \;\middle|\; w_{-j}^{-d}, \boldsymbol{\psi}, \phi, \mathbf{y}\right)} = \frac{\prod_{v=1}^{V} \pi\left(\psi_j^v \;\middle|\; S_j^v\left(\left(w_j^d\right)^*; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-d)}\right)\right)}{\prod_{v=1}^{V} \pi\left(\psi_j^v \;\middle|\; S_j^v\left(\left(w_j^d\right)^{(m)}; \mathbf{K}_v(\phi), \mathbf{w}_{(-j)}^{(-d)}\right)\right)} \times \frac{\pi\left(\left(w_j^d\right)^*\right)}{\pi\left(\left(w_j^d\right)^{(m)}\right)}.$$

- Sampling $\boldsymbol{\psi}$: this involves pV updates. For $j = 1, \ldots, p$ and $v = 1, \ldots, V$ we have that $Pr\left(\psi_j^v = 1 \;\middle|\; \boldsymbol{\psi}_{(-j)}^{(-v)}, \mathbf{w}_{(j)}, \phi, \mathbf{y}\right) \propto \theta\left(\psi_j^v = 1\right)$, with $\theta\left(\psi_j^v = 1\right) = \pi\left(\psi_j^v = 1 \;\middle|\; S_j^v = \mathbf{K}_v'(\phi)\mathbf{w}_{(j)}\right)(1+T)^{-\psi_j^v} M\left(\boldsymbol{\psi}^v\right)^{-(T+a_\sigma)}$, and $\boldsymbol{\psi}_{(-j)}^{(-v)}$ are all binary indicator variables ψ_l^u, $l \neq j$, $u \neq v$. Therefore,

$$\left(\psi_j^v \;\middle|\; \boldsymbol{\psi}_{(-j)}^{(-v)}, \mathbf{w}_{(j)}, \phi, \mathbf{y}\right) \sim Bernoulli\left(\frac{\theta\left(\psi_j^v = 1\right)}{\theta\left(\psi_j^v = 1\right) + \theta\left(\psi_j^v = 0\right)}\right).$$

- Sampling ϕ: Let $\xi = \log(\phi)$, then, the full conditional of ξ is

$$\begin{aligned} q\left(\xi \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right) &= q\left(\phi(\xi) \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right) e^{\xi} \\ &= Ga\left(e^{\xi} \mid a_\phi, b_\phi\right) \cdot e^{\xi} \cdot \prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_j^v \;\middle|\; S_j^v\left(\mathbf{K}_v(e^{\xi}); \mathbf{w}_{(j)}\right)\right). \end{aligned}$$

  We use a normal random walk proposal on ξ, i.e., $\xi^* \sim N\left(\xi^{(m)}, \sigma_\xi^2\right)$, where $\xi^{(m)}$ is current state value and σ_ξ^2 is a tuning parameter. Then, the Metropolis-Hastings ratio is given by

$$\frac{q\left(\xi^* \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right)}{q\left(\xi^{(m)} \mid \boldsymbol{\psi}, \mathbf{w}, \mathbf{y}\right)} = \frac{\prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_j^v \mid S_j^v\left(\mathbf{K}_v(e^{\xi^*}); \mathbf{w}_{(j)}\right)\right)}{\prod_{j=1}^{p} \prod_{v=1}^{V} \pi\left(\psi_j^v \mid S_j^v\left(\mathbf{K}_v(e^{\xi^{(m)}}); \mathbf{w}_{(j)}\right)\right)} \times \frac{Ga\left(e^{\xi^*} \mid a_\phi, b_\phi\right) \cdot e^{\xi^*}}{Ga\left(e^{\xi^{(m)}} \mid a_\phi, b_\phi\right) \cdot e^{\xi^{(m)}}}.$$

  For the general case that $\phi = \left(\phi^1, \ldots, \phi^D\right)$, each ϕ^d can be updated using a scheme similar to the one described above.

2 Additional aspects related to Simulation Study 1

2.1 MCMC convergence

Figure 1 shows the trace plots of the latent spatial effects w_{24} and w_{25} as well as the kernel parameter ϕ for the Simulation Study 1 in Section 4. After warming

up the first 1000 samples and thinning the sequences by keeping every 25th draw, 2000 draws were used for analysis.

The trace, sample autocorrelation function, and the ergonic mean plots show that the collected samples are representative of the parameters' posterior distribution. In the ergonic mean plot, the black curve shows the posterior mean, and the two red-dashed curves indicate the 95% credible interval. After 1500 posterior samples are collected, the posterior mean and the 95% credible interval are stablized.

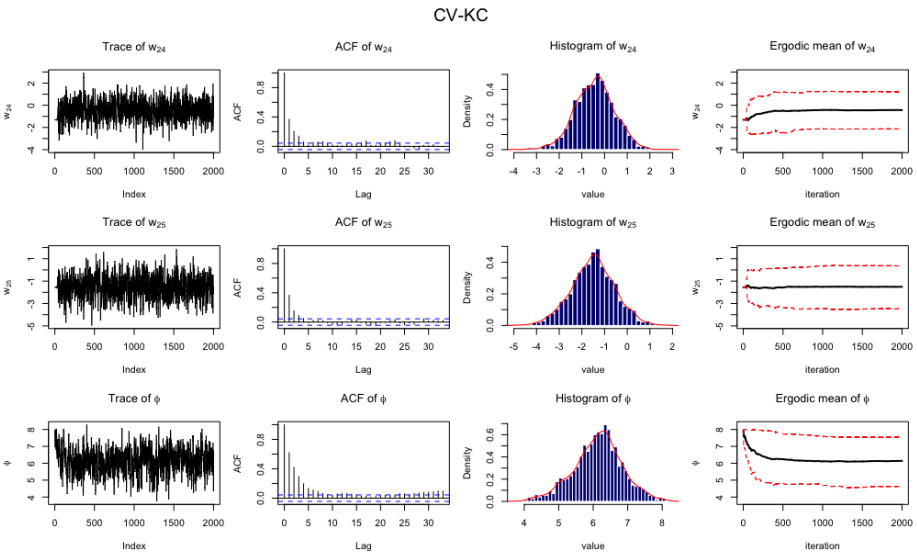


Figure 1: Convergence Diagnostics Plotting.

2.2 Model performance: Effect of t^* and ROC curves

Here we focus on comparing the performance of the different models under different values of the threshold t^* that is used to classify voxels as active. We show that incorporating a spatial structure via KC leads to improved performance in detecting activation at the voxel level not only for complex-valued fMRI but also for magnitude-only fMRI, and this is the case for different values of t^*. More specifically, in order to compare model performance under different values of t^* we examine receiver operating characteristic (ROC) curves for the different models using the data from Simulation study I.

Figure 2 shows the ROC curves obtained under different values of t^* for 3 complex-valued models, CV-KC, CV-GP and C-EMVS (left plot) as well as results for 3 magnitude-only models (right plot). We see that the spatial models clearly dominate the models with no spatial structure in both complex-valued and magnitude-only cases, and that the models that incorporate the spatial

structure via kernel convolution dominate all the other models regardless of the value of t^*.

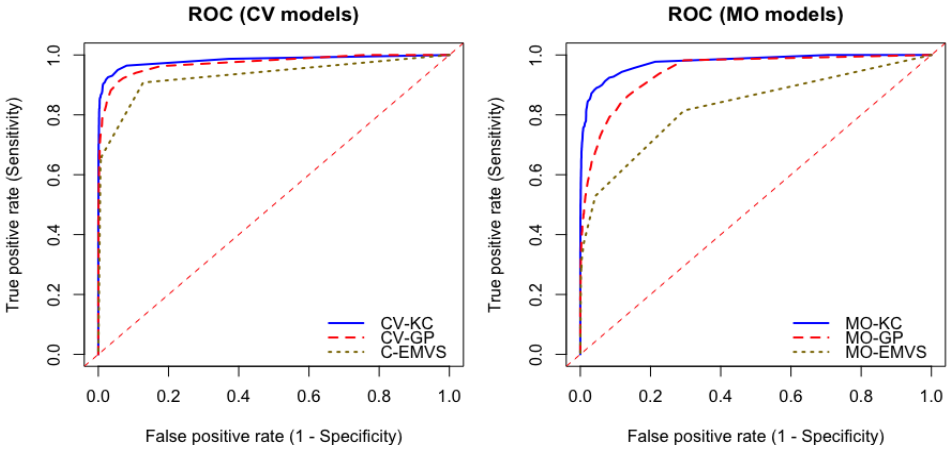


Figure 2: ROC curve for the CV (left) and MO (right) models: KC (blue-solid), GP (red-dash) and EMVS (brown-dot).

2.3 Multi-resolution model

Another advantage of the kernel-based method is that it allows for multi-resolution approaches. For example, we can use more centroid locations in areas that are more likely to be activated. Also, kernels used in areas with more or less centroid locations could be different to capture different activation patterns. The mathematical formulation is straightforward. Suppose that D sites are chosen by using a coarse grid and H regions are created using a fine grid. Denote the "coarse" kernel functions by $k_c\left(z_v - s_d, \phi_c\right)$ and the "fine" kernel functions by $k_f\left(z_v - u_h, \phi_f\right)$, respectively, where s_d denotes the coordinate of site d in the the coarse grid, and u_h denotes the location of site h in the fine grid. Then, the voxel-level overall spatial effect is

$$S_j^v = \sum_{d=1}^{D} k_c\left(z_v - s_d; \phi_c\right) w_j^d + \sum_{h=1}^{H} k_f\left(z_v - u_h; \phi_f\right) b_j^h = S_{c,j}^v + S_{f,j}^v,$$

where b_j^h is another Gaussian variable independent of w_j^g. The posterior sampling scheme for this multi-resolution model is essentially the same as the scheme used for the model with only one resolution described above. Figure 3 shows how multi-resolution works. In this example, we put more centroid locations on the regions that include the true activated voxels. This is not uncommon, as in many practical scenarios neuroscientists have some knowledge about the areas in the brain where they expect to see active voxels in a particular experimental

setting. The middle picture shows the posterior mean of the overall voxel-level spatial effects. The spatial effects from the coarse grid become a very smooth and flat layer because all possible ups and downs are captured by the spatial effects from the fine grid. The estimated probability activation map obtained by computing the logistic transformation of the overall spatial effects shows values near one for activated voxels and essentially zero for non-activated voxels. This shows that properly using a multi-resolution scheme can further improve activation detection without dramatically increasing the computation time.

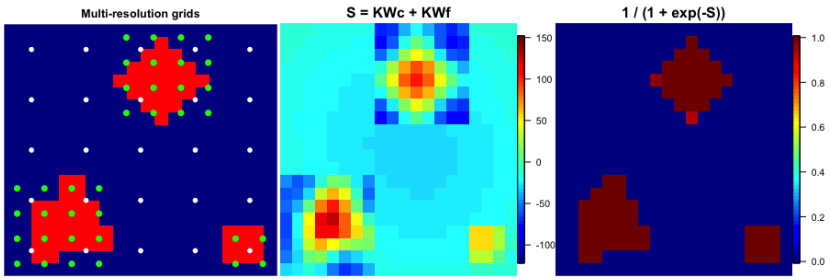
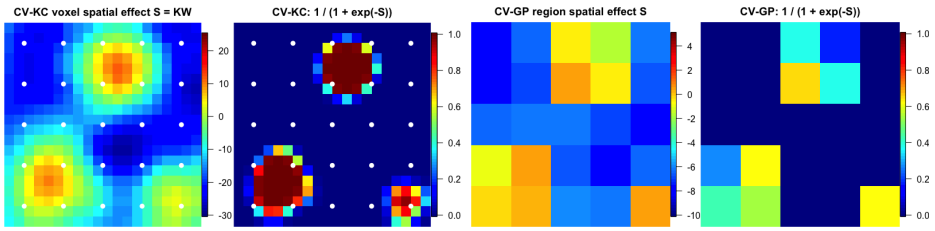


Figure 3: CV-KC multi-resolution spatial effects. From left to right: (1) Two resolution sites. Sites on the coarse grids are colored white and sites on the fine grids are colored green. (2) Posterior mean of overall voxel-level spatial effects, i.e., sum of the spatial effects from coarse and fine grids. (3) Logistic transformation of the overall spatial effects.



(a) CV-KC voxel-level spatial effect (b) CV-GP region-level spatial effect

Figure 4: Posterior mean of spatial random effects and corresponding logistic transformation for the CV-KC and CV-GP with $\beta_{v,1} = 5$ (high CNR). White dots represent the location of latent spatial sites used for KC.

2.4 Kernel effects

Figure 5 shows the estimated weights of the site-specific spatial effects $w^1, \ldots, w^{25}$ on each voxel in the CV-KC model. Specifically, in this example with 25 sites,

the voxel-level spatial effect is given by

$$S^v = \sum_{d=1}^{25} k\left(z_v - s_d\right) w^d = k\left(z_v - s_1\right) w^1 + \cdots + k\left(z_v - s_{25}\right) w^{25},$$

and hence the weights of the effects $w^1, \ldots, w^{25}$ are column vectors $\{\mathbf{K}_{v,1}\}_{v=1}^{400}$, $\ldots$, $\{\mathbf{K}_{v,25}\}_{v=1}^{400}$ of the 400×25 kernel matrix $\mathbf{K}$. Figure 5 shows each of these columns in an image. Since a common ϕ is used for all spatial regions, the radii of the kernel effects are the same. The effects displayed in the plot are obtained by evaluating the Bezier kernel function $k\left(z_v - s_d; \tilde{\phi}\right)$ at the posterior mean of ϕ, $\tilde{\phi}$. Then, this figure shows how a given voxel is related to other voxels in the image. In general, voxels in a given region g are mostly influenced by the spatial effect of the same region, as indicated in the picture, but they are also affected by the spatial effects of other regions with the kernel values shown in the picture acting as weights for the different effects. For example, voxels in region 2 (see Figure the main paper for region numbers) are heavily affected by w^2 with the largest weights given by $k\left(z_v - s_2; \tilde{\phi}\right), v \in \mathcal{R}_2$. This is shown in the image corresponding to Column 2 of K in Figure 5. However, voxels in region 2 are also impacted by the spatial effects from regions 1, 3, 6, 7, and 8 with smaller weights $k\left(z_v - s_g; \tilde{\phi}\right), v \in \mathcal{R}_2$, for $g = 1, 3, 6, 7, 8$.

2.5 Effect of smooth parameter ν and comparison to a convolution model with Gaussian kernels

We examine the effect of the smooth parameter ν in the Bezier kernel function by detecting activation performance and recovering the latent spatial process when $\nu = 1, 2, 3$ and 5. Figure 6 shows the Bezier and Gaussian kernels with different values of parameters, where the Gaussian kernel is defined as $k\left(s_v - u_d; \phi\right) \propto \exp\{-\|s_v - u_d\|^2 / 2\phi^2\}$. For the Bezier kernels, with a fixed value of parameter ϕ, a smaller value of ν leads to a larger degree of smoothness. Also, when ν is small, the Bezier correlation decays faster as ϕ decreases. In the case of the Gaussian kernels, the correlation decays exponentially with distance, but any two voxels in the image are correlated. Instead, for Bezier kernels, two voxels in the image are uncorrelated if their distance is longer than ϕ. Note also that Gaussian kernels generally produce smoother spatial latent processes.

Table 1 shows that in our simulation study the kernel method is not so sensitive to what value of ν is used in the model. This is because the KC model can learn the value ϕ, balancing the effect of ν and ϕ, and hence produce a kernel effect that is best for detecting activation. When the Gaussian kernel is used, the latent spatial process will be smoother than the spatial process generated by the Bezier kernel. The performance measures for the Gaussian kernel are similar to those for the Bezier kernels, but for all the measures one of the Bezier kernel options always outperforms the Gaussian kernel, as shown in Table 1.

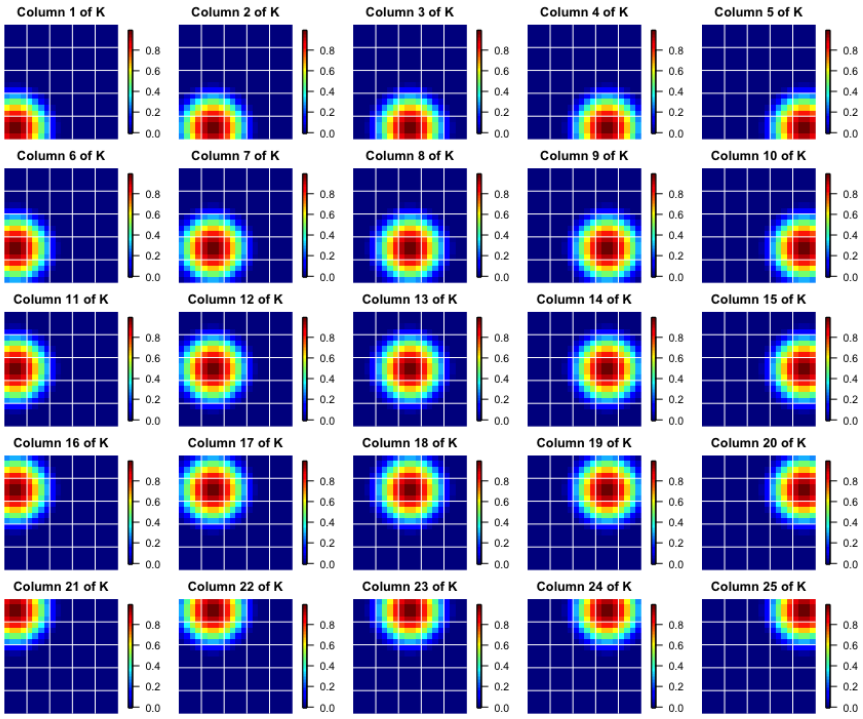


Figure 5: Kernel effect with the kernel function $k\left(z_v - s_g; \tilde{\phi}\right)$ evaluated at the posterior mean of ϕ, $\tilde{\phi}$. Each image shows the effect of spatial region g.

2.6 Selection of the number of spatial sites for KC

The spatial process S^v constructed from KC generally works better by using smoothing kernels that are not very peaked, for example, Gaussian, Bezier, and tricube kernels. If a very peaked kernel is used, such as $k(s) = |s|^{-0.5} \exp\{-|s|\}$, a very dense set of locations $\{s^d\}_{d=1}^D$ would be needed, making the convolution method impractical. For a given a normal kernel width, it is recommended that the spacing of the s^ds locations is no larger than one standard deviation of the kernel (Finkenstadt et al., 2007). Instead of pre-specifying a kernel width, our model flexibly learns the neighboring structure from the data by inferring the range parameter ϕ. The smaller number of locations we use, the larger ϕ would need to be in order to allow voxels to borrow information from neighboring voxels and produce a smooth underlying spatial effect map. However, by doing so over smoothing would be unavoidable, and it could potentially mask possible activation patterns, resulting in low detecting power, as shown in Figure 7 when $D = 16$. From simulation and human data analysis, we suggest that the spacing between centroid locations is no larger than the size of a potential activation region.

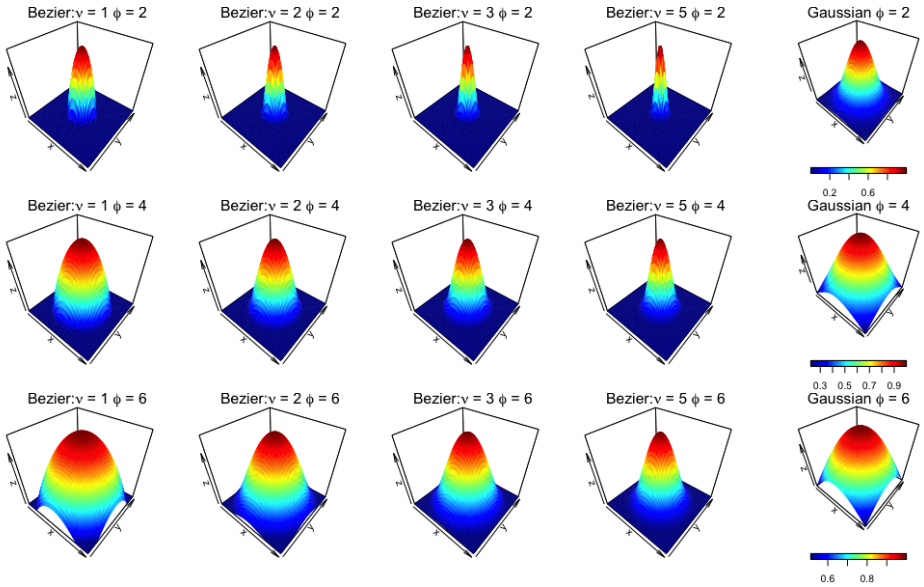


Figure 6: Bezier and Gaussian kernel functions.

Table 1: Performance measures of CV-KC under the Bezier kernel with $\nu = 1, 2, 3$ and 5 and the Gaussian kernel. The threshold value is 0.5.

Kernel type	Sensitivity	Specificity	Precision	Accuracy
Bezier $\nu = 1$	0.7387	0.9985	0.9893	0.9582
Bezier $\nu = 2$	0.7629	0.9982	0.9870	0.9618
Bezier $\nu = 3$	0.7790	0.9979	0.9858	0.9640
Bezier $\nu = 5$	0.7677	0.9982	0.9877	0.9625
Gaussian	0.7516	0.9982	0.9875	0.9600

2.7 Effect of the number of spatial knots and computation time

Given the fact that the GP model forces voxels in the same region to share the same spatial effect, its detection performance is sensitive to the number of regions and to how the image is parcellated. In general, the larger G is, the better detection the model produces. Figure 7 shows how activation probability maps change with the number of spatial regions G and the number of spatial sites D. Notice that CV-KC is much less affected by D. In fact, CV-KC with $D = 25$ can lead to the same or even better activation detection than CV-GP with $G = 100$. To examine how G and D affect the computation efficiency of CV-GP and CV-KC, we ran the algorithm 10 times for each model, and obtained their average computing time on simulated data with $V = 400$, $T = 50$ and $p = 1$ under low CNR case. The results are shown in Table 2. The algorithms were

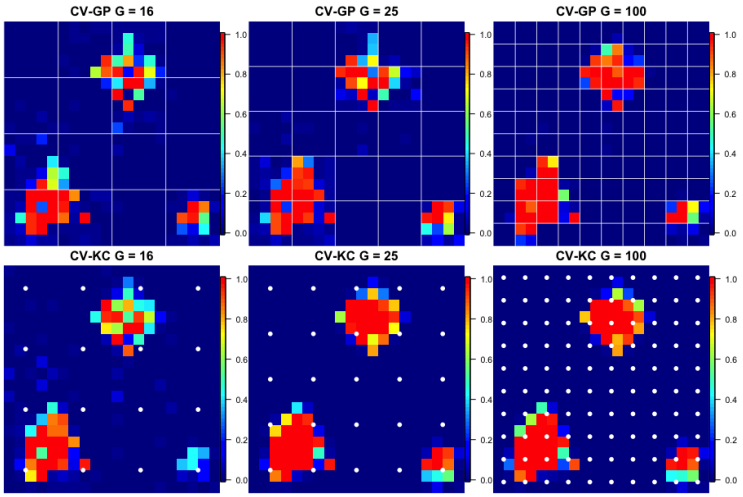


Figure 7: Effect of G and D under the low CNR case. Top: posterior probability maps of CV-GP. Bottom: posterior probability maps of CV-KC. From left to right are maps with $G = D = 16, 25, 100$, respectively. White dots represent the location of latent spatial sites used for KC.

Table 2: Average computing time for the CV-GP and CV-KC MCMC algorithms based on 10 runs. The time unit is seconds per 1000 MCMC iterations.

G/D	4	16	25	100
CV-GP	0.234	0.297	0.409	3.363
CV-KC	0.367	0.508	0.598	1.479

run in `R` version 4.1.1 with `C++` code embedded using the `Rcpp` package without any parallel computing scheme. The hardware used for running the algorithm is MacBook Pro 13 inch 2020 version with the operating system macOS Big Sur 11.5.2. Note that the computing time of CV-GP grows faster than that of CV-KC as G gets larger. For these data, when $G = 100$, CV-GP takes about 2.3 times longer to obtain the same type of activation results obtained by CV-KC.

3 Data generating process of the realistic simulated data

The data were generated using a discrete version of the magnetic resonance (MR) signal equation after steady state magnetization (Karaman et al., 2015):

$$s(k_x, k_y|t) = \int\int \rho(x,y) e^{\frac{-t}{T_2^*(x,y)}} \left(1 - e^{\frac{-TR}{T_1(x,y)}}\right) e^{i\Gamma_H \Delta B(x,y) t} e^{-i2\pi(k_x x + k_y y)} \mathrm{d}x\mathrm{d}y, \quad (1)$$

where $s(k_x, k_y|t)$ is the k-space location at intra slice time t, $\rho(x, y)$ is the proton spin density (PSD), $T_2^*(x, y)$ is the transverse relaxation rate (TRR), $T_1(x, y)$ is the longitudinal relaxation rate (LRR), $\Delta B(x, y)$ is the magnetic field inhomogeneity (MFI), and Γ_H is the proton gyromagnetic ratio (Haacke et al., 1999). The k-space points in (1) are defined by the temporal integral of the magnetic field gradients $G_x(\cdot)$ and $G_y(\cdot)$:

$$k_x = \frac{\Gamma_H}{2\pi} \int_0^t G_x(t')\mathrm{d}t', \qquad \text{and} \quad k_y = \frac{\Gamma_H}{2\pi} \int_0^t G_y(t')\mathrm{d}t'.$$

Brains are scanned 7 times sequentially along the axial direction to generate 7 slices of image. The input used in data generating process, 3.0 T tissue specific physical parameters, (Peters et al., 2006) was given in Table 3 and displayed in Figure 9. To simplify the process, the data set was generated with no MFI $\Delta B(x, y)$. The k-space array after being reconstructed then yielded a real-valued image with a maximum around one.

The defined true magnitude in the simulation were the real-value images scaled by multiplying a magnitude signal-to-noise ratio $\mathrm{SNR_M} \times \sigma/\overline{GM}$, where $\sigma = 1$ is the standard deviation of noise added to the simulation, $\mathrm{SNR_M} = 25$ is the signal-to-noise ratio, and $\overline{GM} = 0.3545$ is the average grey matter value before scaling. Baseline phase shown in Figure 8 was then added to each tissue type according to $\alpha_{0,OB} = 0$ for outside brain, $\alpha_{0,WM} = \pi/12$ for white matter, $\alpha_{0,GM} = \pi/6$ for grey matter, and $\alpha_{0,CSF} = \pi/4$ for cerebrospinal fluid. The remaining imaging parameters were selected to mirror those of an experimental dataset with field of view $(FOV) = 240$ mm, time to echo $(TE) = 50$ ms, flip angle $(FA) = 90°$, effective echo spacing $(EESP) = 720$ μs, and bandwidth $(BW) = 125$ kHz.

	Grey Matter	White Matter	Cerebrospinal Fluid	Outside Brain
PSD	0.83	0.71	1.00	10^{-9}
T_2^*	59.7	54.6	2200	10^{10}
T_1	1331	832	4000	10^{-6}

Table 3: Tissue physical parameter values.

Each slice in the simulated data has dimension 96×96 over $T = 490$ time points. The two 5×5 squares shown in the magnitude contrast and phase contrast of Figure 8 are true activation regions. Therefore, two $5 \times 5 \times 5$ cubes are the activation regions of the whole brain if slices are combined. Each active voxel has different intensity and the voxels near the center of the region have stronger intensities than the ones around the edges of the region. At each time point, the magnitude contrast (β_1) in Figure 8 was multiplied by a task response waveform and then added to the magnitude baseline (β_0) in Figure 8 to form the image magnitude. At each time point, the phase contrast (α_1) in Figure 8 was multiplied by a task response waveform then added to the phase baseline (α_0) in Figure 8 to form the image phase. Notice that slice 1 and slice 7 have no activated voxels. Both magnitude and phase contrasts are zero in these two

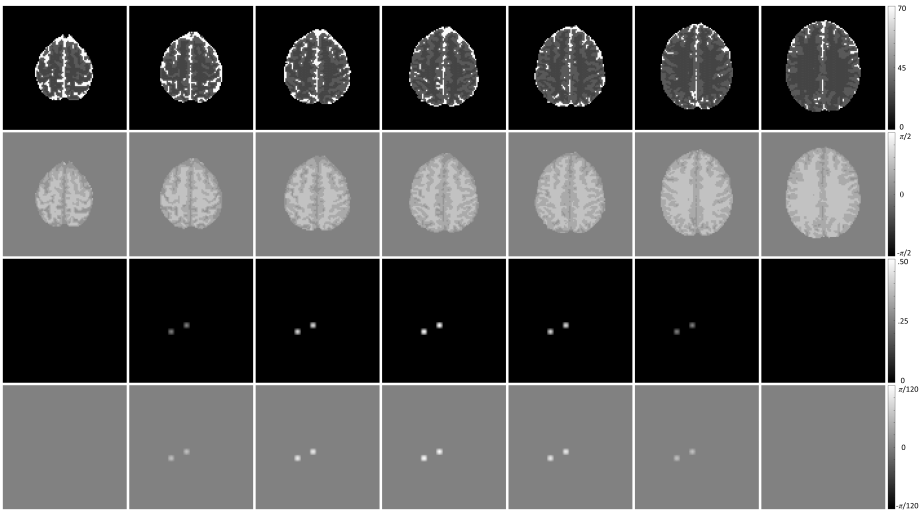


Figure 8: True simulated image values. From left to right: slice 1 to slice 7. From top to bottom: (1) Magnitude baseline; (2) Phase baseline; (3) Magnitude contrast; (4) Phase contrast.

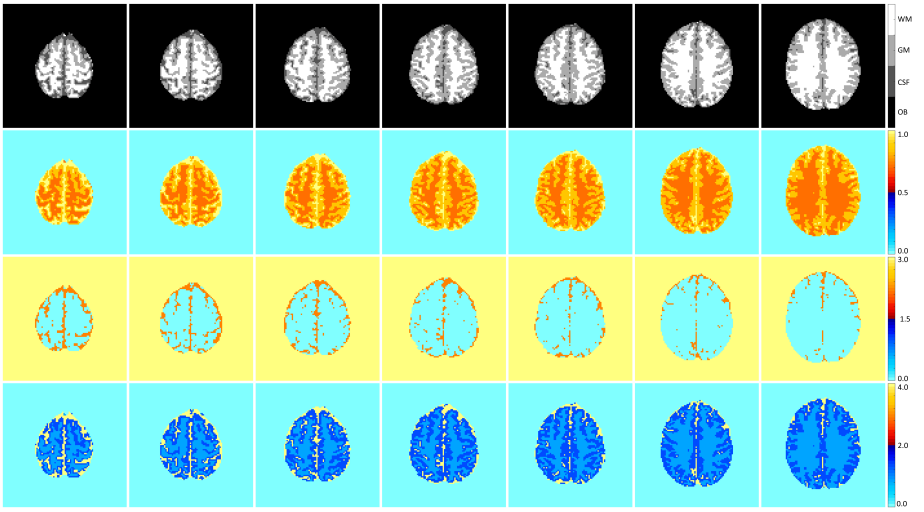


Figure 9: Slices representing physical parameters. From left to right: slice 1 to slice 7. From top to bottom: (1) Grey Matter (GM), White Matter (WM), Cerebrospinal Fluid (CSF) and Outside Brain (OB); (2) PSD $\rho(x, y)$; (3) TRR, $T_2^*(x, y)$; (4) LRR, $T_1(x, y)$. The units of measurement for T_2^* and T_1 are milliseconds.

slices of image. Independent zero mean and unit variance normal noise was also

added to the real and imaginary parts at each time point. In this simulation, the maximum magnitude CNR and phase CNR occur in the center of each ROI. The contrast values (β_1 and α_1) in each ROI were then multiplied by an unnormalized Gaussian kernel with full-width-at-half-max ($FWHM$) = 4 voxels. Slice 4 has the largest maximum magnitude and phase CNR with value $\text{CNR}_\text{M} = \beta_1/\sigma = 0.5/1$ and $\text{CNR}_\text{P} = \alpha_1/\text{SNR}_\text{M} = (\pi/120)/25$, respectively. Slice 3 and 5 have weaker activations than slice 4. Their maximum magnitude and phase CNR are 0.42045 and $(0.841\pi/120)/25$. Slice 2 and 6 have weakest CNR with the maximum value 0.25 for magnitude and $(0.5\pi/120)/25$ for phase. Slice 1 and 7 have no contrasts.

4 Real Human Data

Figure 10 shows the images of the magnitude, and phase angle, as well as images of the real and imaginary components of CV-fMRI data recorded in a subject who participated in a finger-tapping block design experiment. The images were obtained from a 96×96 slice taken at time point 400 (of a time course of length 510) from the full CV-fMRI data, and is part of the human CV-fMRI dataset analyzed in the human data analysis section in the main paper.

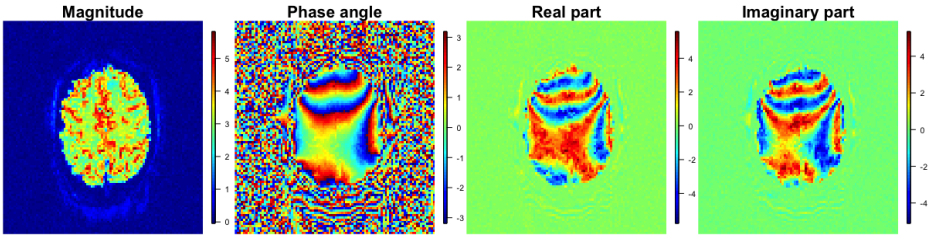


Figure 10: Human CV-fMRI images. From left to right: magnitude image, phase angle image, and real and imaginary images. The CV-fMRI data was recorded during a finger-tapping experiment. The 96×96 slices were taken at time point 400.

References

Finkenstadt, B., Held, L., and Isham, V. (2007). *Statistical Methods for Spatio-Temporal Systems*. Chapman and Hall/CRC.

Haacke, E., Brown, R., Thompson, M., and Venkatesan, R. (1999). *Magnetic Resonance Imaging: Principles and Sequence Design*. John Wiley and Sons, New York.

Karaman, M., Bruce, I. P., and Rowe, D. B. (2015). Incorporating relaxivities to more accurately reconstruct MR images. *Magnetic Resonance Imaging* **33,** 374–384.

Peters, A. M., Broooks, M. J., Hoogenrad, F. G., Gowland, P. A., Francis, S. T., Morris, P. G., and Bowtell, R. W. (2006). Comparison of T2* measurements in human brain at 1.5, 3 and 7 T. In *Proc. Intl. Soc. Magn. Reson. Med.*, volume 14, page 926, Seattle, WA, USA. John Wiley & Sons.