Biometrics WILEY

A spatial Bayesian latent factor model for image-on-image regression

Cui Guo ⓘ | **Jian Kang** ⓘ | **Timothy D. Johnson** ⓘ

Department of Biostatistics, University of
Michigan, Ann Arbor, Michigan, USA

Correspondence Jian Kang, Department
of Biostatistics, University of Michigan,
Ann Arbor, MI 48109, USA.
Email: jiankang@umich.edu

Abstract

Image-on-image regression analysis, using images to predict images, is a challenging task, due to (1) the high dimensionality and (2) the complex spatial dependence structures in image predictors and image outcomes. In this work, we propose a novel image-on-image regression model, by extending a spatial Bayesian latent factor model to image data, where low-dimensional latent factors are adopted to make connections between high-dimensional image outcomes and image predictors. We assign Gaussian process priors to the spatially varying regression coefficients in the model, which can well capture the complex spatial dependence among image outcomes as well as that among the image predictors. We perform simulation studies to evaluate the out-of-sample prediction performance of our method compared with linear regression and voxel-wise regression methods for different scenarios. The proposed method achieves better prediction accuracy by effectively accounting for the spatial dependence and efficiently reduces image dimensions with latent factors. We apply the proposed method to analysis of multimodal image data in the Human Connectome Project where we predict task-related contrast maps using subcortical volumetric seed maps.

KEYWORDS
Bayesian predictive modeling, Gaussian processes, multimodal neuroimaging, spatial latent factor model

1 | INTRODUCTION

Image-related regression analysis has attracted increasing scientific interest in many areas, including medicine (Duff *et al.*, 2015), disease diagnosis (Suk *et al.*, 2017), and neuroscience (Zhou *et al.*, 2013; Bowman, 2014). In these applications, researchers are often interested in identifying the association between quantitative images, for example, functional magnetic resonance imaging (fMRI) (Glover, 2011), and other variables of interest, for example, patient clinical characteristics, where the quantitative images are either considered as the outcome variables or predictors.

Two types of image-related regression models have been extensively studied: scalar-on-image regression that uses images to predict scalar outcome (Reiss and Ogden, 2010; Goldsmith *et al.*, 2011; Huang *et al.*, 2013; Wang *et al.*, 2017; Kang *et al.*, 2018) and image-on-scalar regression which uses a set of scalar predictors to predict the image outcome (Gelfand *et al.*, 2003; Reiss *et al.*, 2010; Chen *et al.*, 2016; Goldsmith and Kitago, 2016; Yan and Liu, 2017). Recently there is increasing interest in developing regression models where both outcomes and predictors are multiple images, to which we refer as the image-on-image regression. It has many important applications in neuroimaging studies, for instance, one question of great interest is to understand how much the individual variations in brain activity during task performance can be explained by task-independent imaging measurements. The answer to

this question may further facilitate the researcher to assess whether the task-invoked brain activities are intrinsic features of individuals who may use different cognitive processes involving different brain circuits, and thus to a certain degree the individual variations in brain activation are independent of irrelevant task factors. For a large set of task conditions spanning several behavioral domains, Tavor *et al.* (2016) trained a simple linear regression model to predict task-based fMRI images using structural and resting-state fMRI images. Based on brain anatomical and functional parcellation, they fitted their model one parcel and one subject at a time, assuming independence and an identical linear relationship across voxels in each parcel. Then they predicted outcomes for unseen subjects by averaging fitted models across subjects. Although their linear method is simple to implement for image-on-image regression analysis, it ignores the spatial dependence structure in images by assuming voxels are mutually independent. Meanwhile, they fitted their model on the individual level but predicted new subjects using averaged estimations, which is in conflict with their assumption of individual variation in their imaging outcomes.

The spatial correlations and associations can be hard to model due to their heterogeneity, complexity and high-dimensionality with a limited sample size. For example, the spatial correlations between neural activity at different voxels may extend beyond neighboring voxels and may not decrease with increasing distance (Bowman, 2014). Also, the spatial patterns may vary across different types of brain images and even subjects. Moreover, the effects of features on outcome images may be from the whole image space but not voxel by voxel. Hence, in order to precisely describe the spatial patterns and associations of outcome and predictor images, one needs to define voxel-wise parameters in image space. These parameters are high dimensional and have complex spatial correlations, leading to a potentially overparameterized regression model along with the associated computational challenges.

In this work, we propose a spatial Bayesian latent factor model for image-on-image regression. The outcome image is modeled with a linear combination of proper basis functions, where the basis coefficients are further decomposed into a few latent factors and a loading matrix. Then latent factors are fitted with predictor images by a scalar-on-image regression model with voxel-wise spatially varying coefficients in predicting image space. This approach reduces the image dimension with proper basis functions and links the predictor and the outcome images with a few latent factors. And the voxel-by-voxel associations between predictor and outcome images are able to be evaluated. Thus, our model can make more accurate predictions compared to the linear model (Tavor *et al.*, 2016) with a small sample size.

A few image-on-image regression models have been recently proposed motivated by various applications in medical imaging. For example, Sweeney *et al.* (2013) applied voxel-wise logistic regression models incorporating multiple sclerosis imaging sequences to predict lesion incidence with T1-weighted, T2-weighted, fluid-attenuated inversion recovery (FLAIR), and proton attenuation (PD) volumes from a longitudinal study. The voxel-wise regression method evaluates the population effects and is simple to implement. However, it ignores the spatial correlations among voxels and thus may lose power to detect the association between the predictor images and the outcome images. Another voxel-wise regression model proposed by Hazra *et al.* (2019) includes prediction effects from the neighboring voxels within a given Euclidean distance. The prediction effects are assumed to be the same when voxels in the predictor image have the same distance to the voxel in the outcome image. In the outcome images, voxels are assumed to be independent over space. Their spatial association with the predictor images is restricted to small regions and only related to the spatial distance. In contrast to the above two voxel-wise regression methods, our spatial latent factor model may capture more complex spatial dependence between outcome image and the predictor images. Deep learning has been applied to medical imaging for image-on-image regression analysis, such as in image recovering and disease diagnosis (Isola *et al.*, 2017; Zhu *et al.*, 2017; Huang *et al.*, 2018). However, the performance of deep learning methods relies on very large sample sizes, which we typically do not have in medical imaging studies. Further, it is difficult to interpret the specific associations among images based on a deep neural network model.

We organize this paper as follows: We first describe our spatial Bayesian latent factor models in Section 2. In Section 3, we present the proposed Bayesian framework for estimation and prediction. In Section 4, we conduct a simulation study under different scenarios and discuss some criteria used for model evaluation and selection. Next, we illustrate the proposed method using fMRI data from the Human Connectome Project (HCP) database. We close with a discussion of future work.

2 | MODEL

In this section, we present our spatial Bayesian latent factor model for image-on-image regression analysis. We extend the classical Bayesian latent factor model for functional and longitudinal data (Montagna *et al.*, 2012) to the case where the functional predictors are images with complex spatial dependence. Our goal of statistical modeling is fundamentally different from the one by Montagna *et al.* (2018)
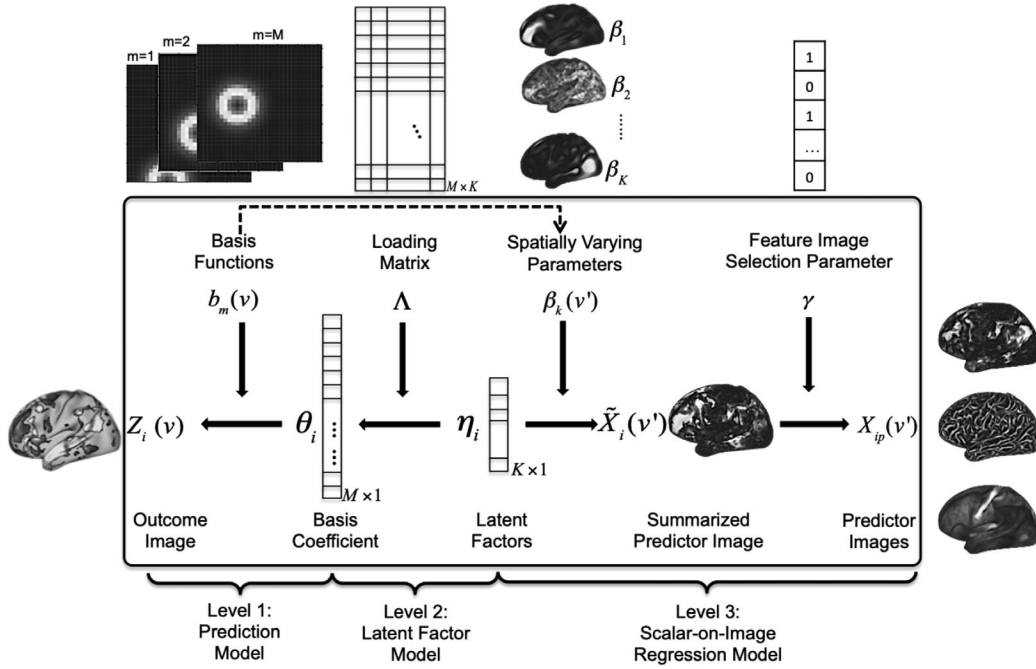
FIGURE 1 Graphical representation of the proposed spatial Bayesian latent factor model for image-on-image regression analysis

that focused on the meta-analysis of functional neuroimaging data.

Suppose the data consists of one outcome image and P predictor images from n subjects. For all subjects, we assume that both outcome and predictor images have been preprocessed and registered to the same brain region, denoted $\mathcal{R}$. Note that in practice $\mathcal{R}$ may refer to the whole brain or a specific region of interest, which we will refer to as a parcel. For each subject $i(i = 1, \ldots, n)$ at voxel $v \in \mathcal{R}$, let $Z_i(v)$ and $X_{ip}(v)(p = 1, \ldots, P)$ represent the outcome image intensity and the pth predictor image intensity, respectively. To identify the association between $Z_i(v)$ and $X_{ip}(v')$ in parcel $\mathcal{R}$ (either $v = v'$ or $v \neq v'$), we develop a spatial Bayesian latent factor model with three levels of hierarchy (see Figure 1 for illustration).

2.1 | Level 1: Approximation of outcome images

At Level 1, we approximate the outcome image using a basis expansion approach. Let $\{b_m(v)\}_{m=1}^M$ be a set of M spatial knot-dependent basis functions that can capture the variation of the outcome image in $\mathcal{R}$. For each subject i and any voxel $v \in \mathcal{R}$, we assume

$$Z_i(v) = U(v) + \sum_{m=1}^M \theta_{im} b_m(v) + e_i(v), \quad U(v) \sim \mathrm{N}(0, \sigma_u^2),$$

$$e_i(v) \sim \mathrm{N}(0, \sigma_e^2),$$

where $U(v)$ represents the population-level spatially varying intercept. As a prior specification, we assume $\{U(v)\}_{v \in \mathcal{R}}$ are independent and identically distributed as normal random variates with mean zero and variance σ_u^2. The random errors $e_i(v)$ are assumed to be independent and identically distributed and follow a normal distribution with mean zero and variance σ_e^2 over all subjects across voxels in region $\mathcal{R}$. The parameter θ_{im} is the subject-specific basis coefficient of the mth spatial basis function $b_m(v)$. The term $\sum_{m=1}^M \theta_{im} b_m(v)$ captures the spatial dependence and smoothness of the subject-specific outcome image among voxels in $\mathcal{R}$.

2.2 | Level 2: Sparse latent factor model for basis coefficients

At Level 2, we build a sparse latent factor model for the basis coefficient θ_{im}:

$$\theta_{im} = \sum_{k=1}^K \lambda_{mk} \eta_{ik} + \zeta_{im}, \qquad \zeta_{im} \sim \mathrm{N}(0, \sigma_\zeta^2),$$

where $\{\eta_{ik}\}_{k=1}^K$ represents a set of K latent factors for subject i and $\{\lambda_{mk}\}_{k=1}^K$ are the corresponding sparse loading coefficients, indicating the effects of the kth latent factor on the mth basis coefficient. For the prior specification for the sparse loading coefficients λ_{mk}, we induce the prior distribution through the parameter-expansion (PX) approach of Ghosh and Dunson (2009) leading to more

efficient posterior computation. Details are provided in Section 3. The subject-level random effects $\{\zeta_{im}\}$ explain the variation of the basis coefficient θ_{im} that cannot be explained by the latent factors η_{ik}. It also characterizes the spatial dependence among the outcome images that cannot be captured by the latent factors. Combining the Level 1 and Level 2 models, we have, for $v \neq v'$,

$$\mathrm{Cov}[Z_i(v), Z_i(v') \mid \{\eta_{ik}\}_{k=1}^K, \cdot] = \sigma_\zeta^2 \sum_{m=1}^M b_m(v) b_m(v'),$$

where "$\cdot$" represents all other parameters. This implies that the variance of random effects ζ_{im}, that is, σ_ζ^2, captures the conditional covariance between $Z_i(v)$ and $Z_i(v')$ given the latent factors and other parameters.

2.3 | Level 3: Link to predictor images

In brain imaging applications, we expect the effect size of a single predictor image from a voxel v' on the whole outcome image through latent factor η_{ik} is generally very small if not zero, but the cumulative effect size over all predictors can be large. Thus, to efficiently reduce the dimension of the parameter space, we consider a summarized predictor image $\widetilde{X}_i(v')$ as a sum of selected predictor images from $\{X_{ip}(v')\}_{p=1}^P$, that is,

$$\widetilde{X}_i(v') = \sum_{p=1}^P \gamma_p X_{ip}(v'),$$

where the latent selection indicator γ_p is assumed to follow a Bernoulli distribution with prior probability w, while π_p may include prior knowledge about important predictor images.

At Level 3, we link the latent factors with the summarized predictor image via a scalar-on-image regression model:

$$\eta_{ik} = \sum_{v' \in \mathcal{R}} \widetilde{X}_i(v') \beta_k(v') + \epsilon_{ik}, \quad \epsilon_{ik} \sim \mathrm{N}(0, \sigma_\epsilon^2),$$

where the spatially varying coefficient $\beta_k(v')$ represents the spatial effects of the summarized predictor image at voxel v' on the kth latent factor; and thus it is factor specific and shared by all subjects. To account for spatial dependence in predictors, we assign a Gaussian process (GP) prior to $\beta_k(v')$ and approximate the GP using a basis expansion approach: that is, $\beta_k(v) = \sum_{m=1}^M \alpha_{km} b_m(v)$ with $\alpha_{km} \sim \mathrm{N}(0, \sigma_\alpha^2)$. Typically, a small number of latent factors are needed to capture important feature information from the selected image predictors at the population level, con-

tributing to the outcome image prediction. The subject-level random effects ϵ_{ik} are introduced to model the dependence between the basis coefficients θ_{im}. In particular, combining Level 2 and Level 3 models, we have, for $m \neq m'$,

$$\mathrm{Cov}[\theta_{im}, \theta_{im'} \mid \{\widetilde{X}_{ip}(v'), v' \in \mathcal{R}\}, \cdot] = \sigma_\epsilon^2 \sum_{k=1}^K \lambda_{mk} \lambda_{m'k}.$$

This implies that the variance of ϵ_{ik}, that is σ_ϵ^2 controls the covariance of the basis coefficients given the summarized predictor image through the latent factor models. Thus, ϵ_{ik} contributes to the complex spatial dependence among outcome images through the correlations between basis expansion coefficients. This type of spatial dependence cannot be captured by the linear transform of the predictor images.

2.4 | Model representation

Our proposed Bayesian hierarchical model has an equivalent model representation by integrating out θ_{im} and η_{ik} (see the Web Appendix). From this representation, the conditional expectation of the outcome image $Z_i(v)$ directly links to predictor image $X_{ip}(v')$ using tensor products of spatially varying coefficients and the predictor selection indicators. Specifically, we have

$$Z_i(v) = U(v) + \sum_{v' \in \mathcal{R}} \psi(v, v') \left\{ \sum_{p=1}^P \gamma_p X_{ip}(v') \right\}$$
$$+ \tilde{\epsilon}_i(v) + \tilde{\xi}_i(v) + e_i(v),$$

$$\psi(v, v') = \sum_{k=1}^K \tilde{\lambda}_k(v) \beta_k(v'), \quad \tilde{\epsilon}_i(v) = \sum_{k=1}^K \epsilon_{ik} \tilde{\lambda}_k(v),$$

$$\tilde{\xi}_i(v) = \sum_{m=1}^M \zeta_{im} b_m(v), \tag{1}$$

where the factor-specific spatially varying coefficient $\tilde{\lambda}_k(v) = \sum_{m=1}^M \lambda_{mk} b_m(v)$ represents the spatial effects of kth latent factor on the outcome image at voxel v. The bivariate spatially varying coefficient $\psi(v, v')$ represents the total spatial prediction effects on outcome images from the selected predictor images, that is the average change in the outcome image at voxel v per unit change in the value of the summarized predictor image at voxel v'; and it is decomposed as the summation of K tensor products of spatially varying coefficients for two types of effects: the effects of summarized predictor images on the latent factors, that is, $\beta_k(v)$ and the effects of latent factors on

the outcome images $\tilde{\lambda}_k(v)$. The two spatial random effects $\tilde{\epsilon}_i(v)$ and $\tilde{\zeta}_i(v)$ constructed by ϵ_{ik} and ζ_{im} are introduced to accommodate the spatial dependence of the outcome images that cannot be explained by predictor images. The expectations of both $\tilde{\epsilon}_i(v)$ and $\tilde{\zeta}_i(v)$ are zero, and their spatial covariances are controlled by σ_ϵ^2 and σ_ζ^2, respectively. Furthermore, we can obtain the conditional covariance between the outcome images given the summarized predictor images and other parameters. For $v \neq v'$

$$\text{Cov}[Z_i(v), Z_i(v') \mid \{\widetilde{X}_i(v''), v'' \in \mathcal{R}\}, \cdot]$$

$$= \sigma_\epsilon^2 \sum_{k=1}^K \tilde{\lambda}_k(v)\tilde{\lambda}_k(v') + \sigma_\zeta^2 \sum_{m=1}^M b_m(v)b_m(v').$$

This indicates that the variance parameters of random effects σ_ϵ^2 and σ_ζ^2 along with the factor loadings λ_{mk} contribute to modeling of the complex spatial dependence of outcome images given the predictor images. This representation implies our model has the ability to retain complex spatial dependence structures in the outcome and predictors, respectively. Hence, by introducing random effects ζ_{im} and ϵ_{ik}, our model is more flexible in borrowing strength from the entire brain, or parcel, to characterize the association between the outcome images and predictor images.

2.5 | Model summary

To summarize, we propose a spatial Bayesian latent factor model for image-on-image regression. We introduce low-dimensional latent factors to link high-dimensional outcome images with predictor images. In particular, for each subject, we represent the expectation of the outcome image as a linear combination of multiple spatial basis functions, where each basis function is associated with one spatial knot. The knot-dependent coefficients are decomposed into several latent factors. Each latent factor summarizes a feature of the outcome image. We model each latent factor as a scalar outcome and assume its expectation is equal to a weighted average of a set of predictors. Each predictor is a linear transformation of one predictor image, where the corresponding spatially varying coefficients represent the spatial effects of the predictor image. We assign GP priors to the spatially varying coefficients and adopt a basis expansion approach for GP representations, which are sufficient, flexible, and convenient to preserve the complex spatial correlations. With an appropriate choice of the basis function, the model can capture well the complex spatial patterns of the outcome image using a smaller number of spatial knots. Thus, our basis expansion

approach may effectively reduce the model dimensionality. More importantly, by integrating over the whole feature image space, our model can well predict the outcome images borrowing the strength not only from the same voxels in the predictor image but also neighboring and even long-distance voxels. A small number of spatial latent factors in our model can effectively capture the association between the predictor images and the outcome image.

3 | POSTERIOR COMPUTATION

We develop a Markov chain Monte Carlo (MCMC) algorithm for posterior computation. For latent factor models, the performance of posterior computation may depend on prior specifications. In general, we can assign normal and inverse-gamma prior distributions to factor loadings and residual variances, respectively. Although those prior distributions produce conditionally conjugate posterior distributions and lead to straightforward computation using a Gibbs sampler, such routine Bayesian implementations are poorly behaved (Ghosh and Dunson, 2009). To achieve efficient posterior computation for our model, we extend the PX method proposed by Ghosh and Dunson (2009). We construct a hierarchical model for the latent factors with different covariance structures.

3.1 | Prior specifications via parameter expansion

Our model needs additional constraints to ensure that the K latent factors are identifiable. Write $\theta_i = (\theta_{im})_{M \times 1}$, $\Lambda = (\lambda_{mk})_{M \times K}$, $\eta_i = (\eta_{ik})_{K \times 1}$, $\zeta_i = (\zeta_{im})_{M \times 1}$, $\bar{X}_i = \{\tilde{X}_i(v)\}_{|\mathcal{R}| \times 1}$, $\beta_k = \{\beta_k(v)\}_{|\mathcal{R}| \times 1}$, $\beta = \{\beta_k(v)\}_{|\mathcal{R}| \times K}$, $\epsilon_i = (\epsilon_{ik})_{K \times 1}$, $\alpha_k = (\alpha_{km})_{M \times 1}$ $\alpha = (\alpha_{km})_{K \times M}$, and $b = \{b_k(v)\}_{|\mathcal{R}| \times K}$, where $|\mathcal{R}|$ represents the number of voxels in $\mathcal{R}$. The matrix representations of the original inferential models in Levels 2 and 3 are shown in Table 1.

Following the PX approach, we develop a working model and the corresponding transformations between inferential and working parameters as shown in Table 1. We introduce parameters $\Phi = \text{diag}\{\phi_1^2, \dots, \phi_K^2\}$ and the sign function $S(x) = 1$ if $x \geq 1$ and -1 otherwise. An extra working intercept term μ_i^* is included to efficiently estimate the working latent factor $\eta_i^* = (\eta_{ik}^*)$. The working factor loading matrix $\Lambda^* = (\lambda_{mk}^*)_{M \times K}$ is lower triangular with no constraints on its elements. Instead of specifying a prior distribution for Λ directly, we induce a prior distribution for Λ^* and then transform it to the prior distribution for Λ. These prior distributions placed on the working parameters are specified as follows: for

TABLE 1 Inferential and working models for PX approach with $i = 1, 2, \ldots, N$, $k = 1, \ldots, K$, $m = 1, \ldots, M$

Inferential model	Working model	Transformations
$\theta_i = \Lambda \eta_i + \zeta_i$	$\theta_i = \Lambda^* \eta_i^* + \zeta_i$	$\lambda_{mk} = S(\lambda_{kk}^*)\phi_k^{-1}\lambda_{mk}^*$
$\eta_i = \beta^T \bar{X}_i + \epsilon_i$	$\eta_i^* = \mu_i^* + [\beta^*]^T \bar{X}_i + \epsilon^*$	$\eta_{ik} = S(\lambda_{kk}^*)\phi_k(\eta_{ik}^* - \mu_{ik}^*)$
$\beta = b\alpha$	$\beta^* = b\alpha^*$	$\beta_k = S(\lambda_{kk}^*)\phi_k \beta_k^*$
$\zeta_i \sim N(0, \sigma_\zeta^2 I)$	$\zeta_i \sim N(\mathbf{0}, \sigma_\zeta^2 I)$	$\alpha_k = S(\lambda_{kk}^*)\phi_k \alpha_k^*$
$\epsilon_i \sim N(0, \sigma_\epsilon^2 I)$	$\epsilon_i^* \sim N(\mathbf{0}, \sigma_\epsilon^2 \Phi^{-1})$	$\epsilon_{ik} = S(\lambda_{kk}^*)\phi_k \epsilon_{ik}^*$

$m = 1, \ldots, M$, $\lambda_{mk}^* \sim N(0, \sigma_\lambda^2)$ if $k \leq \min(m, K)$, and $\lambda_{mk}^* \sim \delta_0$ if $\min(m, K) < k \leq K$, where δ_0 is a measure concentrated at zero. $\mu_i^* \sim N_K(\mathbf{0}, \sigma_\mu^2 I_K)$ for $i = 1, \ldots, N$; $\phi_k^2 \sim \text{Gamma}(a_\phi, b_\phi)$ and $\alpha_k^* \sim N_M(\mathbf{0}, \sigma_\alpha^2 I_M)$ for $k = 1, \ldots, K$. Hyperparameters $\sigma_\lambda^2, \sigma_\mu^2, \sigma_\alpha^2, a_\phi$, and b_ϕ can be prespecified.

According to the above working model representation and prior specifications, we develop an efficient Gibbs sampler for posterior computation (see the Appendix). Of note, the PX approach leads to an overparameterized working model and thus the posterior computation may exhibit poor mixing due to lack of identifiability (Ghosh and Dunson, 2009) for the working parameters. However, the parameters in the original inferential model are still identifiable and the Markov chain mixes much better. Estimation and prediction details are given in the Web Appendix.

3.2 | Basis functions and number of latent factors

It is challenging to choose the number of basis functions. Since the bases are locally concentrated, the more basis functions that are included in the model, the richer the spatially varying patterns of the outcome image the model can capture. On the other hand, to reduce computational costs, the number of basis functions is usually chosen to be much smaller than the number of voxels. In addition, the appropriate basis functions are unknown in advance. Conceptually, any basis functions, like B-spline bases and Gaussian kernels, can be chosen for the smooth images. Here, we choose a three-dimensional (3D) isotropic Gaussian kernel. $b_m(v) = \exp\{-b\|v - \psi_m\|^2\}$, $v \in \mathcal{R}$, $m = 1, 2, \ldots, M$, with kernel locations $\{\psi_m\}_{m=1}^M$ and parameter b controlling smoothness. Flexible approaches are available for estimating M, b, and $\{\psi_m\}_{m=1}^M$ for basis functions. One approach is to perform fully Bayesian inference with appropriate prior specifications for these parameters. However, this approach suffers very large computational burden as the basis function is reevaluated at each iteration of the MCMC algorithm. Therefore, we adopt a relatively less computationally intensive approach. We first choose a reasonable number of bases

M as well as kernel locations $\{\psi_m\}_{m=1}^M$, and then we determine the smooth parameter b by minimizing the mean squared error (MSE) and mean squared prediction error (MSPE) of outcome images via cross validation (CV). The metrics MSE and MSPE are averaged over datasets, observations, and voxels. Suppose we consider J-fold cross validation. For the jth ($j = 1, \ldots, J$) fold, let $\mathcal{I}_j^{\text{ts}}$ and $\mathcal{I}_j^{\text{tr}}$ represent the indices of the subjects in the test set and training set, respectively. Let $\hat{Z}_i(v)$ for $i \in \mathcal{I}_j^{\text{tr}}$ and $\tilde{Z}_i(v)$ for $i \in \mathcal{I}_j^{\text{ts}}$ represent the fitted and predicted outcome image at voxel v, respectively. We have $\hat{Z}_i(v) = \text{E}[Z_i(v) \mid \{\hat{\theta}_{im}\}_{m=1}^M, \hat{U}(v)] = \hat{U}(v) + \sum_{m=1}^M \hat{\theta}_{im}b_m(s)$, where $\hat{U}(v)$ and $\hat{\theta}_{im}$ are the posterior mean estimates of $U(v)$ and $\hat{\theta}_{im}$. And $\tilde{Z}_i(v) = \text{E}[Z_i(v) \mid \{X_{ip}(v'), v' \in \mathcal{R}\}_{p=1}^P, \cdot] = \hat{U}(v) + \sum_{v' \in \mathcal{R}} \sum_{p=1}^P \hat{\psi}(v, v')\hat{\gamma}_p X_{ip}(v')$, where $\hat{\psi}(v, v')$ and $\hat{\gamma}_p$ are the posterior mean estimates of $\psi(v, v')$ and γ_p. Then MSE is defined as $\sum_{j=1}^J \sum_{i \in \mathcal{I}_j^{\text{tr}}} \sum_{v \in \mathcal{R}}\{Z_i(v) - \hat{Z}_i(v)\}^2/(J|\mathcal{I}_j^{\text{tr}}||\mathcal{R}|)$, and MSPE is defined as $\sum_{j=1}^J \sum_{i \in \mathcal{I}_j^{\text{ts}}} \sum_{v \in \mathcal{R}}\{Z_i(v) - \tilde{Z}_i(v)\}^2/(J|\mathcal{I}_j^{\text{ts}}||\mathcal{R}|)$, where $|\mathcal{I}_j^{\text{tr}}|$, $|\mathcal{I}_j^{\text{ts}}|$, and $|\mathcal{R}|$ represent the training sample size, the test sample size, and the number of voxels used in the CV study, respectively.

To select the number of latent factors, a set of widely used model comparison criteria can be considered, such as deviance information criteria (DIC), Bayesian information criteria (BIC), Bayes factors (BF), and R-squared. One can fit the model multiple times with different values of K and choose the optimal value based on the above criteria. This approach is computationally intensive. We consider an alternative approach in light of our latent factor model. In our model, redundant latent factors may have very sparse, zero-, or near zero concentrated loadings. Hence, the optimal number of latent factors has a loading matrix with nonzero-concentrated and distinguishable loading vectors. Some metrics can help us to measure the variation and sparseness of each loading vector, based on which we can determine how many loading vectors are nonredundant. Here, we adopt a metric as the number of values in each loading vector outside the 90% credible interval (CI) of the whole estimated loading matrix, termed as 90% significant

loading counts (SLC). Loading vectors corresponding to important latent factors should have large 90% SLC since they account for most of the variation of the loading matrix. On the contrary, the SLC of those redundant loading vectors are close to zero and similar to each other. Some other metrics, such as mean or standard deviation of loading vectors, etc., can be used as well. We find that 90% SLC is more robust in choosing the correct number of latent factors in our simulation study. More precisely, we first fit the model with a large number of latent factors, for example, $K = 20$ in practice; then we summarize the estimated loading vectors using the 90% SLC and determine the optimal number K using a modified Elbow method (Kodinariya and Makwana, 2013). Specifically, we sort 90% SLC of all loading vectors in a descending order and create a trajectory of 90% SLC versus the number of latent factors. Our "elbow criterion" is that the optimal number K is chosen at the change point of the trajectory after which the 90% SLC almost remains unchanged and close to zero. Finally, we refit the model with the choice of K for estimation and prediction.

4 | SIMULATION STUDY

4.1 | Data generation and method

Simulation studies were conducted to compare the performance of our proposed method with two other models: linear regression (Tavor *et al.*, 2016) and voxel-wise regression. The three methods serve as the generating models in three different scenarios, respectively. In each scenario, we generate 10 datasets, each of which contains 100 observations in the training set with another 50 observations in the test set. Each simulated observation has a two-dimensional outcome image and a set of 20 two-dimensional predictor images on a 32×32 equally spaced grid. In the simulation studies, we treat the whole image space as a single parcel. Specifically, we generate predictor images from a GP with mean zero and covariance $c(v_1, v_2) = 0.01 \exp\{-15 \times d_{v_1,v_2}^2\}$, where d_{v_1,v_2} is the Euclidean distance between any two grid points v_1 and v_2.

Scenario 1. We simulate outcome images from the following linear regression model:

$$Z_i(v) = \beta_{i0} + \sum_{p=1}^{20} \beta_{ip} X_{ip}(v) + \epsilon_i(v), \quad \text{for} \quad i = 1, \dots, 150,$$

where linear coefficients β_{ip} are the same over the space but varied among observations. The error terms $\epsilon_i(v)$ are independently sampled from a normal distribution with mean zero and variance 0.1. The true value of β_{ip} is drawn from a normal distribution with mean μ_p and variance σ_p^2,

where σ_p^2 is drawn from a gamma distribution with shape 0.1 and rate 0.1. The mean μ_p is sampled from the uniform distribution on $[-3.5, -1.5] \cup [1.5, 3.5]$ if $p \leq 5$, and μ_p is generated from a uniform distribution on $[-0.5, 0.5]$, otherwise.

Scenario 2. We generate outcome images from the following voxel-wise regression model:

$$Z_i(v) = \beta_0(v) + \sum_{p=1}^{20} \beta_p(v) X_{ip}(v) + \epsilon_i(v), \text{ for } i = 1, \dots, 150,$$

where $\epsilon_i(v)$ are independently sampled from N(0, 0.1). We simulate $\beta_p(v)$ from a GP with mean zero and correlation kernel $\exp\{-15 d_{v_1,v_2}^2\}$. The marginal variance of $\beta_p(v)$ is 2.0 for $p \leq 5$ and 0.5 otherwise.

Scenario 3. We simulate data from our spatial Bayesian latent factor (SBLF) model (see Figure 1 in the Web Appendix for an illustration). We first define a set of basis functions using Gaussian kernels with equally spaced kernel where the knots are defined on grid points $\{1, \dots, 32\} \times \{1, \dots, 32\}$. We follow the PX method to generate working parameters and then take the transformations to obtain the original parameters. The true number of latent factors, K, is set to 5. Loading elements are first generated from a normal distribution. Then, for each simulated loading vector, we replace those simulated values outside its 50% CI by zero's to maintain the sparsity of the loading matrix. For the imaging predictor, indicator $\gamma_p = 1.0$ if $p \leq 5$ and $\gamma_p = 0$ otherwise.

We choose the values of parameters in these three scenarios in terms of the signal-to-noise ratio (SNR), a measure of signal strength relative to background noise, defined as $\text{Var}[\text{E}\{Z_i(v) \mid X_{i1}(v), \dots, X_{iP}(v)\}]/\text{Var}\{\epsilon_i(v)\}$. To ensure comparable results in different scenarios, we choose the parameter values such that the SNRs of simulated observations in the three scenarios have similar distributions with mean 30 and range [1, 100].

We run the MCMC algorithm for 25,000 iterations with 15,000 burn-in. We compute the posterior mean and CIs for the parameters of interest. For all the parameters with Gamma priors in Section 3.1, we set both shape and scale parameters to 1.0. We fix $\sigma_\lambda^2 = \sigma_\mu^2 = \sigma_\alpha^2 = 1.0$. For other hyperprior specifications, $\omega \sim \text{Beta}(1.0, 1.0)$. All initial values are sampled from their corresponding prior distributions, except that the initial values of γ_p are 1.0. Further, we fit the model with 1, 5, 10, and 20 latent factors, respectively.

4.2 | Results

Table 2 shows estimation and prediction accuracy for the three scenarios, including MSE, MSPE, and the proportion of observations for which our method produces smaller

TABLE 2 Simulation study results for Scenarios 1–3

Generating model	Analysis method	K	Training			Test		
			MSE	%	%*	MSPE	%	%*
	Linear	–	0.010	–	–	1.024	–	–
	Voxel-wise	–	1.016	–	–	1.050	–	–
Scenario 1	SBLF	1	0.014	0.00	100.00	1.231	23.40	27.20
(linear)		5	0.014	0.00	100.00	1.395	31.60	34.20
		10	0.014	0.00	100.00	1.474	29.40	32.60
		20	0.014	0.00	100.00	1.509	26.00	27.80
	Linear	–	0.136	–	–	0.511	–	–
	Voxel-wise	–	0.008	–	–	0.496	–	–
Scenario 2	SBLF	1	0.023	100.00	0.00	0.314	94.40	92.80
(voxel-wise)		5	0.023	100.00	0.00	0.316	94.20	92.60
		10	0.023	100.00	0.00	0.322	93.40	91.80
		20	0.023	100.00	0.00	0.341	91.80	90.80
	Linear	–	0.896	–	–	3.656	–	–
	Voxel-wise	–	2.510	–	–	3.642	–	–
Scenario 3	SBLF	1	0.149	100.00	100.00	3.436	59.80	56.40
(SBLF)		5	0.149	100.00	100.00	1.633	94.80	94.40
		10	0.149	100.00	100.00	1.923	89.40	89.00
		20	0.149	100.00	100.00	3.347	58.80	57.20

In each scenario, three models are fitted and their results compared in terms of (1) MSE, (2) MSPE, and (3) the proportions of observations with smaller MSE/MSPE using SBLF than the linear (%) or voxel-wise regressions (%*). Results of multiple values of K used in our method are included, and the true value of K used for simulations in Scenario 3 is $K = 5$.

MSE or MSPE compared to the other methods. The method used as the true generating model in each scenario has the smallest MSE for test sets. In Scenario 1, when the data are generated from the linear regression model, SBLF performs better the other methods for about 20%–30% of observations in the test sets. With a similar SNR, in Scenario 2, when the data are generated from the voxel-wise regression model, SBLF achieves a smaller MSPE and over 90% better predictions than the voxel-wise regression method. In Scenario 3, when the data are generated from SBLF, SBLF with a correct number of latent factors K leads to the best performance and an incorrect K can result in less accurate prediction.

As we discussed in Section 3.2, it is of interest to evaluate different criteria for selecting the number of latent factors K. Our simulation study in Scenario 3 indicates that some of the widely used model comparison criteria, including DIC, BIC, BF, and R-squared, could not help to identify the correct value of K. Specifically, BIC always prefers small K, while R-squared and BF favor large K. The selection of K using DIC varied widely across the 10 simulated datasets. However, the MSPE for outcomes of test sets is a robust measure for choosing K. As shown in Table 2, in Scenario 3, SBLF with a correct value of K, $K = 5$, has the smallest MSPE and the largest proportion of better performing observations than other methods in all 10 repeated studies.
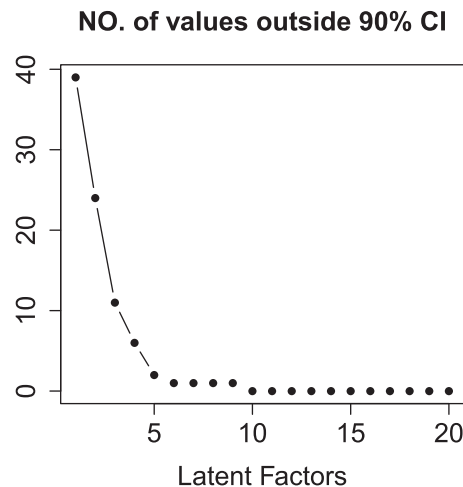
FIGURE 2 Elbow method plot with metric values of the posterior mean estimations of loading vector in simulation study Scenario 3, fitted with $K = 20$ (true $K = 5$). The metric is the number of values in each loading vectors outside the 90% CI of the whole loading matrix

Figure 2 shows designed metric for loading vectors from 1 to 20, based on which we correctly determine the value of K ($K = 5$) using our Elbow method mentioned in Section 3.2. Since posterior inference on predictor selection is biased when $K = 20$, we fit the model with $K = 5$. The

estimated posterior inclusion probability for the first five predictors was all one, and all other predictors have posterior inclusion probability equal to zero.

5 | APPLICATION

5.1 | The motivating HCP data

We apply our SBLF model to analyze a subset of neuroimaging data from the HCP. Our goal is to make predictions on the individual task-evoked images using the corresponding task-independent images. Tavor *et al.* (2016) performed a similar analysis on the same dataset using a simple linear regression approach ignoring the spatial dependence among voxels within parcels. Their analysis focused on the cortical surface imaging measurements, while our model is developed for analysis of the volumetric imaging data on 19 subcortical regions. The dataset is comprised of 98 subjects' functional and structure imaging data from the Q3 release. Details of all acquisition parameters and processing mechanisms are described in Barch *et al.* (2013).

We focus on the 19 subcortical regions consisting of 31,870 voxels. The outcome image is the faces-shapes contrast map derived from the EMOTION task fMRI data. The predictor images are 32 subcortical seed maps derived from the resting-state fMRI data. Details about the 32 subcortical seed maps can be found in Tavor *et al.* (2016). (See examples of the outcome and predictor images shown in figures in the Web Appendix.) It is well known that the amygdalae consistently associate with emotional functioning (Phan *et al.*, 2002). Hence, we report results on the left and right amygdala as examples to demonstrate our application analysis and results. There are 315 and 332 volumetric voxels within the left and right amygdala regions, respectively, and their corresponding example outcome and predictor maps are shown in the Web Appendix.

5.2 | Analysis

For SBLF, we adopt the same prior specifications as those used in the simulation study. The initial values of the predictor selection indicators are set to one. For all other parameters, the initial values are randomly sampled from their prior distributions. We specify the basis functions for the left and right amygdala containing 51 and 58 knots, respectively. To choose a good hyperparameter b in the basis functions, we use cross validation and consider three candidate values $\{1/10, 1/20, 1/30\}$. The basis functions with the three values are shown in Figures 7– 9 in the Web Appendix. To choose the number of latent factors, we start

with a large number $K = 20$ and applied the Elbow method to select the number of latent factors and refit the model with this number of latent factors. We run our proposed MCMC algorithm for 50,000 iterations with 25,000 burn-in iterations. We check the convergence of all the MCMC simulations using the Gelman–Rubin diagnostics (Gelman and Rubin, 1992). Given each selected hyperparameter b and number of latent factors K, we run five MCMC chains with different initial values. The potential scale reduction factors (PSRF) are estimated for each voxel in the outcome images. The point estimates of the PSRF range from 1.000 to 1.005 (median 1.000, mean 1.000), and the upper confidence limits have the maximum value 1.016 (median 1.000, mean 1.000), indicating convergence of the MCMC algorithm. For comparisons, we also fit the data using the simple linear regression approach (Tavor *et al.*, 2016). For each of the 19 subcortical regions, we separately perform the analysis using both methods.

5.3 | Results

Table 3 shows the 10-fold cross-validation prediction accuracy using SBLF with different values of b and K compared with the other two methods. For both the left and the right amygdala, when $b = 10$, SBLF has the smallest averaged MSPE and the largest proportions of better predicted outcomes for test sets than the other methods. Compared with the linear and voxel-wise regression methods, SBLF has the smallest MSE for fitting the outcome images in the training dataset for all the combinations of K and b. For the left amygdala region, the optimal choice for the number latent factors is 9, with MSPE = 1.168 which is smaller than that of both linear regression (1.357) and voxel-wise regression (1.540). For over 69% and 65% of outcome images in the test set, SBLF produces a smaller MPSE compared to linear regression and voxel-wise regression, respectively. These proportions for the right amygdala are even larger (77.55% and 73.47%), as shown in Table 3.

Table 2 in the Web Appendix displays results for all 19 subcortical regions compared with the linear regression method. The squared errors of outcomes from our model are about 10.2 times smaller than the linear regression method on average for all 19 regions. Our model has much larger R-squared values than the linear regression method across all regions (0.934 vs. 0.428 on average). Hence, our SBLF model performs better than the linear regression model (Tavor *et al.*, 2016) for predicting the task-evoked functional brain activity from the task-free volumetric images in the subcortical regions.

From MCMC samples of the predictor selection indicators, γ, we can estimate the posterior inclusion probability

**TABLE 3** Results for the left (1) and the right (2) amygdala

(1) Left amygdala region

Method	Bandwidth	Number of latents	Training MSE	%	%*	Test MSPE	%	%*
Linear	–	–	0.562	–	–	1.357	–	–
Voxel-wise	–	–	0.644	–	–	1.540	–	–
SBLF	$b^* = 10$	$K = 20$	0.060	100.00	100.00	1.198	66.33	63.37
		$K^* = 9$	0.063	100.00	100.00	1.168	69.39	66.30
	$b = 20$	$K = 20$	0.009	100.00	100.00	1.304	54.08	60.02
		$K^* = 8$	0.009	100.00	100.00	1.348	51.02	58.16
	$b = 30$	$K = 20$	0.005	100.00	100.00	1.830	29.59	40.82
		$K^* = 5$	0.005	100.00	100.00	1.830	20.41	40.82

(2) Right amygdala region

Method	Bandwidth	Number of latents	Training MSE	%	%*	Test MSPE	%	%*
Linear	–	–	0.651	–	–	1.539	–	–
Voxel-wise	–	–	0.735	–	–	1.866	–	–
SBLF	$b^* = 10$	$K = 20$	0.066	100.00	100.00	1.359	70.41	73.47
		$K^* = 10$	0.069	100.00	100.00	1.260	77.55	73.47
	$b = 20$	$K = 20$	0.010	100.00	100.00	1.398	68.37	74.49
		$K^* = 7$	0.010	100.00	100.00	1.941	38.78	50.00
	$b = 30$	$K = 20$	0.005	100.00	100.00	2.127	19.39	42.86
		$K^* = 5$	0.006	100.00	100.00	3.195	10.26	12.82

Performance of three different methods are compared in terms of (1) MSE, (2) MSPE, and (3) the proportions of observations with smaller MSE/MSPE using SBLF than the linear (%) or voxel-wise regressions (%*). MSE/MSPE is reported as the averaged values over all voxels, subjects, and 10-folds cross validation. Two tuning parameters, bandwidth value b for basis functions, and the number of latent factors K, are tested to determine their optimal values b^* and K^*. The SBLF model is refitted with the value of K^* determined using the Elbow method with $K = 20$.

for each predictor image, indicating the uncertainty of including the corresponding predictor images into the model. For each amygdala region, by placing a threshold value on the posterior inclusion probability, we can obtain a set of predictor images that are associated with the outcome image with certain uncertainty level. We vary the threshold from 0.0 to 0.9 and list the corresponding set of predictor images in the Web Appendix. For the right amygdala region, the posterior probability of including the 28th cortical seed map into the model is larger than 0.6. Among all the predictor images, this cortical seed map has the strongest association with the faces-shapes contrast image in the Emotion domain. Similarly, in the left amygdala region, the same predictor image also has a relative strong association (the posterior inclusion probability larger than 0.5) with the outcomes in the same task domain. However, in the left amygdala region, the 13th and 15th predictor images have more contributions to the outcome predictions given their estimations of their γs over 0.8 in the left amygdala region. These strong associations do not appear in the right amygdala regions. These two subcortical seed maps are from the cerebellum subcortical seeds, indicating the significant associations between cerebellum

structure and emotional functions in left amygdala. These results are consistent with some previous findings in neuroimaging and neurological research (Turner *et al.*, 2007; Habas *et al.*, 2009; Baumann and Mattingley, 2012; Habas, 2018). It has been shown that cerebellum is involved in neural processes underlying the regulation of emotional responses (Baumann and Mattingley, 2012). In particular, the role of the cerebellum in the modulation of neural networks that subserve processing of emotional material has been studied by analyzing the functional imaging data of lesion patients (Turner *et al.*, 2007). The results indicated that the "normal" emotional response to frightening stimuli in patients with cerebellar damage may be associated with relatively lower activity in the amygdala and other limbic and paralimbic regions. In addition, recent resting state fMRI studies (Habas *et al.*, 2009; Habas, 2018) have identified the intrinsic functional connectivity signals between the cerebellum and the amygdaloid nucleus.

Understanding how the predictor images are associated with the outcome images is also of interest. As presented (1), $\psi(v, v')$ represents the spatial prediction effect on a voxel v in the outcome image from any voxel v' in the
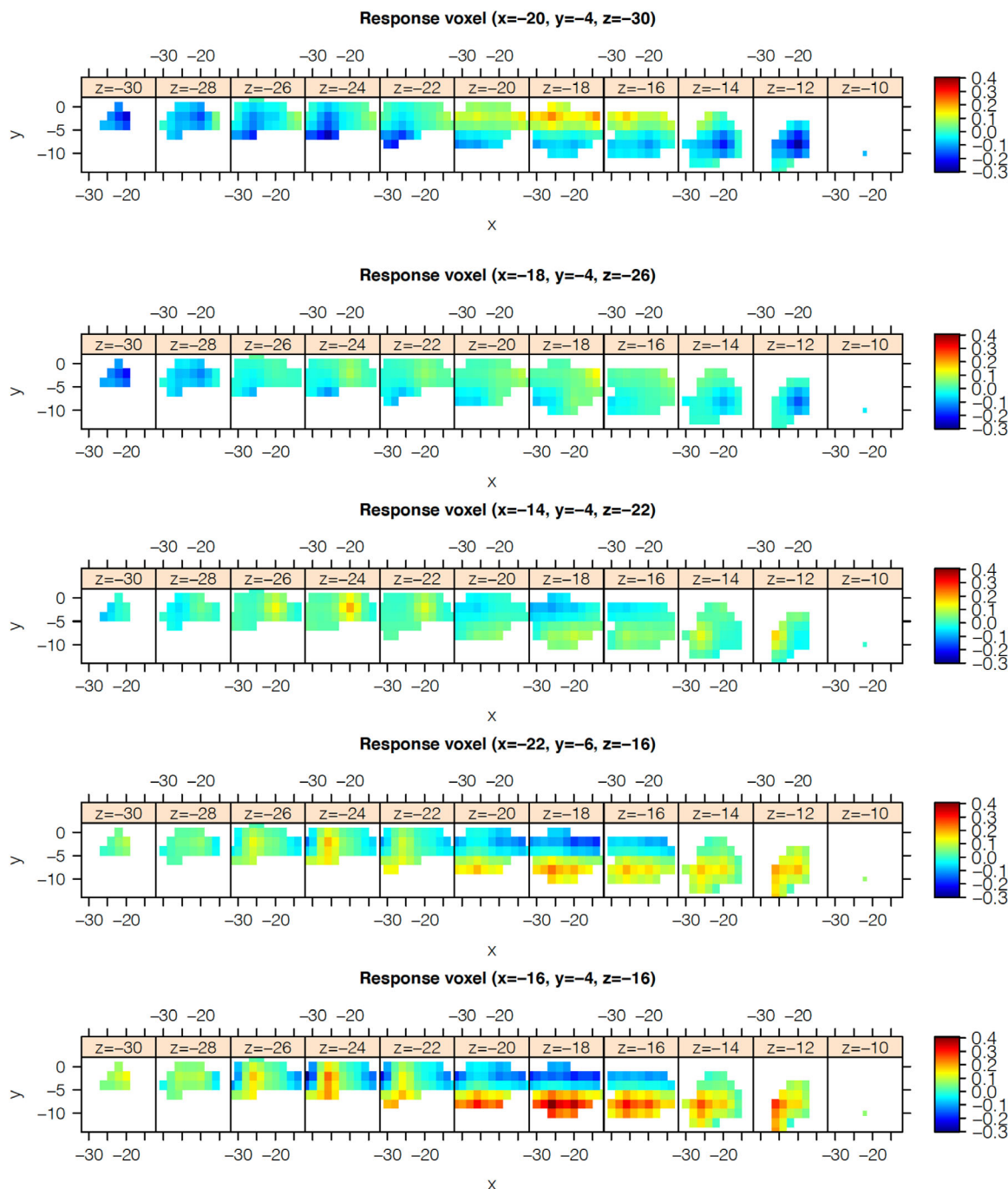
FIGURE 3 Spatially varying prediction effects $\psi(v, v')$ on five different response voxels v from all predictor voxels v'. Both v and v' are in the left amygdala maps. All maps are plotted on the same color scale varying from blue for negative values to red for positive values. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

summarized predictor image. Figure 3 shows the estimated values $\psi(v, v')$ on five outcome voxels v in the left amygdala region. The same predictor image has various effects on different outcome voxels. For example, the first outcome voxel (the first row in Figure 3) is negatively associated with nearby voxels in predictor images, while the last two voxels (the fourth and fifth rows) have more positive effects from voxels in similar locations in the

predictor images. In contrast, there is no significant effect from the predictor images on the other two voxels (the second and third rows). Meanwhile, significant associations appear among not only nearby voxels but also voxels that are farther away. For example, for the first outcome voxel ($x = -20, y = -4, z = -30$) in Figure 3, some distant voxels in the image slice ($z = -18$) are positively associated with it, while nearby voxels have significant

negative effects. These estimated associations between predictor and outcome images from our proposed model can be important in exploring brain functions and cannot be obtained by the other two methods.

6 | DISCUSSION

In this work, we propose a spatial Bayesian latent factor model for image-on-image regression. We use low-dimensional latent factors as a bridge connecting the outcome image and predictor images in the same high-dimensional imaging space. Our proposed method is flexible enough to model the spatial dependence through prespecified basis functions without imposing strong assumptions about the spatial patterns. Our SBLF model can identify the associations between the outcome image and predictor images across the whole image space, not restricted to voxels from the same locations or nearby neighbors. The low-dimensional latent factors integrate information from predictor images through a regression model with spatially varying coefficients. This regression model can include other clinical patient characteristics as well. Our method can be applied to jointly analyze multimodality imaging data, such as resting-state fMRI, task-based fMRI, and structural MRI.

We now discuss the limitations and potential future directions for our method. First, the cross-validation approach to determining the number of basis functions and the number of latent factors is very computationally intensive. An alternative is to treat these as unknown parameters and assign a multinomial prior distribution (Ghosh and Dunson, 2009), then we can make fully Bayesian inference on the model. This approach requires a trans-dimensional MCMC algorithm, which can be quite challenging in practice. Second, we make a strong assumption that the spatially varying coefficients are common for all predictor images, while the spatial predictive effects of different predictor images can be different. This assumption may reduce the power to detect the predictive effects of some predictor images and may inflate the false positive rate. We can relax this assumption by introducing predictor specific spatially varying coefficient parameters. This modification obviously increases model complexity which may require informative priors to ensure parameter identifiability and develop more efficient computational algorithms. Third, different subjects may have heterogeneous associations between task-related brain activity and resting-state activity. Our current SBLF model cannot capture this heterogeneity. We can potentially extend our model by introducing subject-specific spatially varying coefficients with clustering structures. Fourth, our current model focuses on volumetric data; however, we can modify our algorithm to brain surface data by projecting the cortex data onto a sphere and then generate smooth basis functions based on spherical harmonic kernels.

DATA AVAILABILITY STATEMENT
The HCP Open Access data is publicly available and can be downloaded from the ConnectomeDB (https://db.humanconnectome.org/) for registered users.

ORCID
Cui Guo https://orcid.org/0000-0002-3297-119X
Jian Kang https://orcid.org/0000-0002-5643-2668
Timothy D. Johnson https://orcid.org/0000-0002-7676-5347

REFERENCES
Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., et al. (2013) Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage*, 80, 169–189.

Baumann, O. and Mattingley, J.B. (2012) Functional topography of primary emotion processing in the human cerebellum. *NeuroImage*, 61, 805–811.

Bowman, F.D. (2014) Brain imaging analysis. *Annual Review of Statistics and Its Application*, 1, 61–85.

Chen, Y., Wang, X., Kong, L. and Zhu, H. (2016) Local region sparse learning for image-on-scalar regression [preprint]. *arXiv*:1605.08501.

Duff, E.P., Vennart, W., Wise, R.G., Howard, M.A., Harris, R.E., Lee, M., et al. (2015) Learning to identify CNS drug action and efficacy using multistudy fMRI data. *Science Translational Medicine*, 7, 274ra16–274ra16.

Gelfand, A.E., Kim, H.-J., Sirmans, C. and Banerjee, S. (2003) Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98, 387–396.

Gelman, A. and Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–472.

Ghosh, J. and Dunson, D.B. (2009) Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18, 306–320.

Glover, G.H. (2011) Overview of functional magnetic resonance imaging. *Neurosurgery Clinics*, 22, 133–139.

Goldsmith, J. and Kitago, T. (2016) Assessing systematic effects of stroke on motor control by using hierarchical function-on-scalar regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65, 215–236.

Goldsmith, J., Wand, M.P. and Crainiceanu, C. (2011) Functional regression via variational Bayes. *Electronic Journal of Statistics*, 5, 572.

Habas, C. (2018) Research note: A resting-state, cerebello-amygdaloid intrinsically connected network. *Cerebellum & Ataxias*, 5, 1–4.

Habas, C., Kamdar, N., Nguyen, D., Prater, K., Beckmann, C.F., Menon, V. and Greicius, M.D. (2009) Distinct cerebellar contributions to intrinsic connectivity networks. *Journal of Neuroscience*, 29, 8586–8594.

Hazra, A., Reich, B.J., Reich, D.S., Shinohara, R.T. and Staicu, A.-M. (2017) A spatio-temporal model for longitudinal image-on-image regression. *Statistics in Biosciences*, 11, 22–46.

Huang, L., Goldsmith, J., Reiss, P.T., Reich, D.S. and Crainiceanu, C.M. (2013) Bayesian scalar-on-image regression with application to association between intracranial dti and cognitive outcomes. *NeuroImage*, 83, 210–223.

Huang, H., Yu, P.S. and Wang, C. (2018) An introduction to image synthesis with generative adversarial nets [preprint]. *arXiv*:1803.04469.

Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A.A. (2017) Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134.

Kang, J., Reich, B.J. and Staicu, A.-M. (2018) Scalar-on-image regression via the soft-thresholded Gaussian process. *Biometrika*, 105, 165–184.

Kodinariya, T.M. and Makwana, P.R. (2013) Review on determining number of cluster in *k*-means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1, 90–95.

Montagna, S., Tokdar, S.T., Neelon, B. and Dunson, D.B. (2012) Bayesian latent factor regression for functional and longitudinal data. *Biometrics*, 68, 1064–1073.

Montagna, S., Wager, T., Barrett, L.F., Johnson, T.D. and Nichols, T.E. (2018) Spatial Bayesian latent factor regression modeling of coordinate-based meta-analysis data. *Biometrics*, 74, 342–353.

Phan, K.L., Wager, T., Taylor, S.F. and Liberzon, I. (2002) Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16, 331–348.

Reiss, P.T., Huang, L. and Mennes, M. (2010) Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics*, 6.

Reiss, P.T. and Ogden, R.T. (2010) Functional generalized linear models with images as predictors. *Biometrics*, 66, 61–69.

Suk, H.-I., Lee, S.-W., Shen, D. and Initiative, A.D.N. (2017) Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical Image Analysis*, 37, 101–113.

Sweeney, E., Shinohara, R., Shea, C., Reich, D. and Crainiceanu, C. (2013) Automatic lesion incidence estimation and detection in multiple sclerosis using multisequence longitudinal MRI. *American Journal of Neuroradiology*, 34, 68–73.

Tavor, I., Jones, O.P., Mars, R., Smith, S., Behrens, T. and Jbabdi, S. (2016) Task-free MRI predicts individual differences in brain activity during task performance. *Science*, 352, 216–220.

Turner, B.M., Paradiso, S., Marvel, C.L., Pierson, R., Ponto, L.L.B., Hichwa, R.D. and Robinson, R.G. (2007) The cerebellum and emotional experience. *Neuropsychologia*, 45, 1331–1341.

Wang, X., Zhu, H. and Initiative, A.D.N. (2017) Generalized scalar-on-image regression models via total variation. *Journal of the American Statistical Association*, 112, 1156–1168.

Yan, B. and Liu, Y. (2017) Smooth image-on-scalar regression for brain mapping [preprint]. *arXiv*:1703.05264.

Zhou, H., Li, L. and Zhu, H. (2013) Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108, 540–552.

Zhu, J., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O. and Shechtman, E. (2017) Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*, 30, 465–476.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 3, 4 and 5 along with the implemented C++/R codes and the demonstrating simulation data are available with this paper at the Biometrics website on Wiley Online Library. We have developed an R package SBLF for our proposed Spatial Bayesian Latent Factor Model with examples, which is publicly available on GitHub (`https://github.com/umich-biostatistics/SBLF`).