Received: 4 December 2019 | Revised: 15 January 2021 | Accepted: 26 January 2021

DOI: 10.1111/biom.13445

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Geostatistical modeling of positive-definite matrices: An application to diffusion tensor imaging

Zhou Lan[1] | **Brian J. Reich**[2] | **Joseph Guinness**[3] | **Dipankar Bandyopadhyay**[4] | **Liangsuo Ma**[4] | **F. Gerard Moeller**[4]

[1] Yale School of Medicine, New Haven, Connecticut

[2] North Carolina State University, Raleigh, North Carolina

[3] Cornell University, Ithaca, New York

[4] Virginia Commonwealth University, Richmond, Virginia

Correspondence
Brian J. Reich, North Carolina State University, Raleigh, NC 27695 U.S.A.
Email: brian_reich@ncsu.edu

Funding information
Foundation for the National Institutes of Health, Grant/Award Numbers: P30CA016059, R01DE024984, U54DA038999

Abstract

Geostatistical modeling for continuous point-referenced data has extensively been applied to neuroimaging because it produces efficient and valid statistical inference. However, diffusion tensor imaging (DTI), a neuroimaging technique characterizing the brain's anatomical structure, produces a positive-definite (p.d.) matrix for each voxel. Currently, only a few geostatistical models for p.d. matrices have been proposed because introducing spatial dependence among p.d. matrices properly is challenging. In this paper, we use the spatial Wishart process, a spatial stochastic process (random field), where each p.d. matrix-variate random variable marginally follows a Wishart distribution, and spatial dependence between random matrices is induced by latent Gaussian processes. This process is valid on an uncountable collection of spatial locations and is almost-surely continuous, leading to a reasonable way of modeling spatial dependence. Motivated by a DTI data set of cocaine users, we propose a spatial matrix-variate regression model based on the spatial Wishart process. A problematic issue is that the spatial Wishart process has no closed-form density function. Hence, we propose an approximation method to obtain a feasible Cholesky decomposition model, which we show to be asymptotically equivalent to the spatial Wishart process model. A local likelihood approximation method is also applied to achieve fast computation. The simulation studies and real data application demonstrate that the Cholesky decomposition process model produces reliable inference and improved performance, compared to other methods.

KEYWORDS

Cholesky decomposition, diffusion tensor imaging, geostatistical modeling, positive-definite matrix, spatial random fields, spatial Wishart process

1 | INTRODUCTION

Diffusion tensor imaging (DTI), a magnetic resonance imaging (MRI) technique, is used to measure the diffusion process of water molecules in the brain (Soares *et al.*, 2013). Estimated 3×3 positive-definite (p.d.) matrices summarize the water diffusion at each location in the brain. The p.d. matrix is also called a diffusion tensor (DT), representing the covariance of the local 3D Brownian motion (Schwartzman, 2006; Dryden *et al.*, 2009). Since DTI has been used extensively to map white matter tractography in the brain, it has an advantage over other

MRI-based techniques in revealing abnormal topological organization in the brain (Lo *et al.*, 2010). A primary clinical objective is to understand how covariates (eg, age, gender, drug use) affect DTs, reflecting its effects on brain structure.

Incorporating spatial dependence is important for achieving efficient and valid inference in imaging data analysis (Spence *et al.*, 2007; Wu *et al.*, 2013; Xue *et al.*, 2018). Recently, Lan *et al.* (2021) also revealed that incorporating spatial dependence leads to improved performance in the DTI region of difference selection, validated by an application to a cocaine user data set (Ma *et al.*, 2017). In this paper, we focus on geostatistical methods, ie, methods for random fields commonly used in the geostatistical literature (Gelfand *et al.*, 2010). Current geostatistical modeling only focuses on random variables following univariate or multivariate distributions (eg, univariate/multivariate Gaussian). However, the voxel-level variable in DTI is a p.d. matrix. In the literature, only a few relevant works have been proposed for spatially-varying p.d. matrices (Gelfand *et al.*, 2004). This triggers our study of geostatistical modeling of p.d. matrices.

Previous attempts to analyze DTI data can be broadly classified into univariate modeling and matrix-variate modeling. To avoid the complexity caused by matrix-variate data, univariate modeling projects a DT onto a descriptive scalar quantity such as the magnitude of isotropy, magnitude/fractional of anisotropy, or mode of anisotropy (Ennis and Kindlmann, 2006). Among these scalar quantities, fractional of anisotropy is the most popular (see Lane *et al.*, 2010; Ma *et al.*, 2017). Several statistical methods are proposed using fractional of anisotropy of diffusion tensors as responses (eg, Zhu *et al.*, 2013; Liu *et al.*, 2016). However, since these projections are surjective (eg, different DTs may project onto the same scalar quantity), the loss of information caused by univariate modeling is irreversible. To this end, matrix-variate modeling has been proposed via parameterizing the DTs as matrix-variate random distributions such as the lognormal distributions (Schwartzman, 2006, 2016) or Wishart distribution (eg, Lee and Schwartzman, 2017). Martín-Fernández *et al.* (2004) proposed a Gaussian Markov random field for DTI, but the model does not ensure the diffusion tensor in the space of p.d. matrices. Therefore, in general, most matrix-variate models have not yet been extended to spatial modeling because incorporating spatial dependence for p.d. matrices is nontrivial.

To mitigate these issues, we propose a spatial matrix-variate regression model. The covariates are incorporated through the Cholesky decomposition (Zhu *et al.*, 2009), and the coefficients are spatially-varying to capture local covariate effects. The Cholesky decomposition is commonly used to model p.d. matrices (eg, Pourahmadi, 2007; Dryden *et al.*, 2009). Here, we propose a new Cholesky decomposition process model for spatial data that approximates the original spatial model via Cholesky decomposition of p.d. matrices as the responses, study some theoretical properties, and develop algorithms for scalable computing.

The spatial dependence among p.d. matrices is achieved by the spatial Wishart process, a spatial random field (stochastic process) supporting spatially dependent Wishart matrices. The first use of the spatial Wishart process was a prior for a spatially-varying covariance matrix (Gelfand *et al.*, 2004). In this paper, the spatial Wishart process is used as a model for p.d. matrix observations. Considering that the literature comprehensively describing the statistical properties of the spatial Wishart process is sparse, we further prove that the spatial Wishart process as a random field on uncountable locations is valid and almost-surely continuous given some conditions.

Although the model based on the spatial Wishart process is elegant with several nice properties, a bottleneck of the spatial Wishart process is that its probability density function is intractable (Viraswami, 1991). Therefore, instead of directly modeling the DTs, we propose a new Cholesky decomposition process model that approximates the original model by taking the Cholesky decomposition of p.d. matrices as the responses. The Cholesky decomposition process model is composed of six univariate spatial Gaussian processes, where the parameters retain the interpretations of the original model. Via theoretical results and simulation studies, we show that the Cholesky decomposition process model is an asymptotic approximation and a useful working model. The theoretical results make an important contribution of both direct and potential value in applications of DTI and other fields, simplifying matrix-variate models relying on dependent Wishart matrices (eg, Karagiannidis *et al.*, 2003; Kuo *et al.*, 2007; Smith and Garth, 2007) to multivariate models relying on Gaussian distributions.

We also deal with massive spatial data by Vecchia's method (Vecchia, 1988; Datta *et al.*, 2016), a local likelihood approximation that approximates the joint density of spatial variables as a product of conditional densities. To demonstrate our proposal, we further investigate its performance using simulation studies, and provide data analysis on cocaine user data (Ma *et al.*, 2017), in comparison to the univariate spatially-varying coefficient process model (Gelfand *et al.*, 2003). The novelty of our work is exploring spatial associations in modeling p.d. matrix-variate data under the framework of geostatistical modeling, with some key theoretical contributions of multivariate analysis and applications to DTI.

2 | SPATIAL WISHART PROCESS MODEL

A typical DTI data set (Ma *et al.*, 2017) usually includes DTs from each subject $i \in \{1, \dots, N\}$ at each voxel $\boldsymbol{s} \in \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_n\}$, and subject-level covariates (eg, age, education level, medical treatments). The primary clinical objective is to detect local covariate effects on DTs. Let $\boldsymbol{A}_i(\boldsymbol{s})$ be the 3×3 p.d. DT matrix of subject i measured at voxel $\boldsymbol{s}$ and $\boldsymbol{X}_i$ be a design matrix containing an intercept and d subject-level covariates. The p.d. DT matrices are modeled as parameterized Wishart matrices (Dryden *et al.*, 2009) to have mean matrix $\boldsymbol{\Sigma}_i(\boldsymbol{s})$ and degrees of freedom m, denoted as $\boldsymbol{A}_i(\boldsymbol{s}) \sim \mathcal{W}(\boldsymbol{\Sigma}_i(\boldsymbol{s}), m)$. To model spatial dependence and ensure that $\boldsymbol{A}_i(\boldsymbol{s})$ is a p.d. matrix, we decompose $\boldsymbol{A}_i(\boldsymbol{s})$ as

$$\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{L}_i(\boldsymbol{s})\boldsymbol{U}_i(\boldsymbol{s})\boldsymbol{L}_i(\boldsymbol{s})^T. \tag{1}$$

In this decomposition, $\boldsymbol{U}_i(\boldsymbol{s})$ has mean $\boldsymbol{I}_3$ and is the spatially dependent *residual term* modeling variation that cannot be explained by the covariates and the *regression term* $\boldsymbol{L}_i(\boldsymbol{s})$ is the lower-triangle Cholesky matrix of the mean matrix $\boldsymbol{\Sigma}_i(\boldsymbol{s})$, ie, $\mathbb{E}\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{\Sigma}_i(\boldsymbol{s}) = \boldsymbol{L}_i(\boldsymbol{s})\boldsymbol{L}_i(\boldsymbol{s})^T$. The spatial Wishart process for $\boldsymbol{U}_i(\boldsymbol{s})$ is described in Section 2.1, and the regression construction for $\boldsymbol{L}_i(\boldsymbol{s})$ as a function of $\boldsymbol{X}_i$ is described in Section 2.2. We refer to this model as the spatial Wishart process model in the rest of the paper.

2.1 | *Residual term*: Spatial Wishart process

In this subsection, we introduce the spatial Wishart process as a means of modeling spatial dependence. Gelfand *et al.* (2004) provide the construction of the spatial Wishart process, which is stated as follows. For $j \in \{1, 2, \dots, m\}$, let $\{\boldsymbol{Z}_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ be a mean-zero p-dimensional multivariate Gaussian process with a $p \times p$ cross-covariance matrix $\boldsymbol{\Sigma}$ and a (univariate) spatial dependence function $\mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})$ determined by parameters $\boldsymbol{\Phi}$, ie, $\text{cov}(\boldsymbol{Z}_j(\boldsymbol{s}), \boldsymbol{Z}_j(\boldsymbol{s}')) = \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi}) \times \boldsymbol{\Sigma}$, denoted as $\boldsymbol{Z}_j \sim \mathcal{GP}(\boldsymbol{0}, \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi}), \boldsymbol{\Sigma})$. If for each $\boldsymbol{s} \in \mathcal{D}$, we collect $\boldsymbol{U}(\boldsymbol{s}) = \sum_{j=1}^{m} \boldsymbol{Z}_j(\boldsymbol{s})\boldsymbol{Z}_j^T(\boldsymbol{s})/m \sim \mathcal{W}(\boldsymbol{\Sigma}, m)$, then the collection $\{\boldsymbol{U}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ is a spatial Wishart process, a random field supporting spatially dependent Wishart matrices. The spatial Wishart process can be understood as a two-level hierarchical model where the spatial dependence of the Wishart matrix $\boldsymbol{U}(\boldsymbol{s})$ is induced by the latent spatial Gaussian processes $\{\boldsymbol{Z}_j\}$. Also, in light of the application to DTI, we assume $p = 3$ by default.

In practice, the number of locations in $\mathcal{D}$ is usually finite. However, Gelfand *et al.* (2010) emphasize the importance of ensuring a valid mathematical specification of a spatial stochastic process when the potential number of locations is uncountable, ie, that the process satisfies the Kolmogorov existence theorem (Oksendal, 2003). Theorem 1 confirms that the spatial Wishart process satisfies the conditions of the Kolmogorov existence theorem and is a valid stochastic process.

Theorem 1 (Spatial Wishart Process). *The spatial Wishart process $\{\boldsymbol{U}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ is a valid stochastic process (random field), which gives proper finite-dimensional distributions for any collections of locations in the spatial domain.*

In addition to showing that the random field is valid, we show that it is almost-surely continuous (Property 1). The proof of almost-sure continuity is based on Kent (1989) (see Web Appendix A).

Property 1 (Almost-Sure Continuity). Let $\{\boldsymbol{U}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ be a spatial Wishart process. If the correlation function $\mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})$ has a second-order Taylor series expansion with remainder that goes to 0 at a rate of $2 + \delta$ for some $\delta > 0$, $\boldsymbol{U}(\boldsymbol{s})$ converges weakly to $\boldsymbol{U}(\boldsymbol{s}_0)$ with probability one as $||\boldsymbol{s} - \boldsymbol{s}_0|| \to 0$.

Since neuroimaging data are usually collected at a high resolution and the effect of the disease at proximally located/neighboring voxels can be similar (see Wu *et al.*, 2013; Xue *et al.*, 2018), the residuals $\boldsymbol{U}_i(\boldsymbol{s})$ should be continuous and spatially dependent. Therefore, we model the residuals $\{\boldsymbol{U}_i(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ for $i \in \{1, 2, \dots, N\}$ as realizations of a spatial Wishart process with degrees of freedom m, cross-covariance matrix $\boldsymbol{I}$, and correlation function $\mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})$, denoted as

$$\boldsymbol{U}_i \sim \mathcal{SWP}(m, \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi}), \boldsymbol{I}). \tag{2}$$

Setting the cross-covariance matrix to $\boldsymbol{I}$ preserves the marginal distribution $\boldsymbol{A}_i(\boldsymbol{s}) \sim \mathcal{W}(\boldsymbol{\Sigma}_i(\boldsymbol{s}), m)$.

In spatial statistics and neuroimaging, understanding the spatial dependence is essential. We quantify spatial dependence of the spatial Wishart process using the expected squared Frobenius norm $\mathcal{V}(\boldsymbol{s}, \boldsymbol{s}') = \mathbb{E}||\boldsymbol{U}(\boldsymbol{s}) - \boldsymbol{U}(\boldsymbol{s}')||_F^2$, where $||.||_F$ is the Frobenius norm. The expected squared Frobenius norm can also be understood as a generalized variogram (Cressie, 1992) for matrix-variate data (Lan *et al.*, 2021), where an increasing spatial dependence of $\boldsymbol{U}(\boldsymbol{s})$ and $\boldsymbol{U}(\boldsymbol{s}')$ leads to a smaller $\mathcal{V}(\boldsymbol{s}, \boldsymbol{s}')$. Through the variogram, we find that the spatial Wishart process $\mathcal{SWP}(m, \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi}), \boldsymbol{\Sigma})$ is separable (Cressie, 1992) since

$$\mathcal{V}(\boldsymbol{s}, \boldsymbol{s}') = \gamma(m, \boldsymbol{\Sigma})[1 - \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})^2], \tag{3}$$

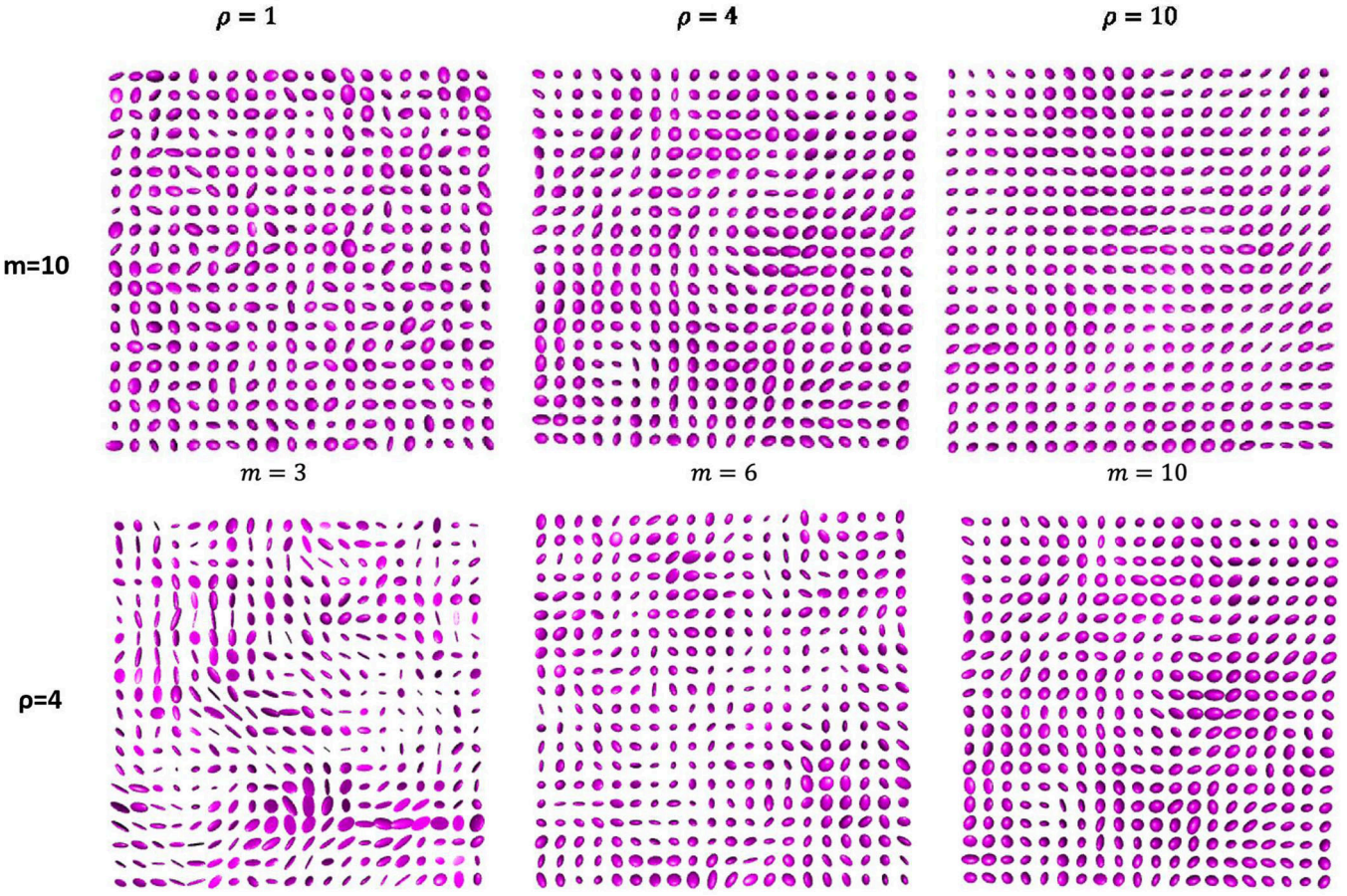

FIGURE 1 The diffusion tensors are generated by spatial Wishart process with mean matrix $\boldsymbol{I}$. The images of the first row are generated with $m = 10$ and $\rho \in \{1, 4, 10\}$ (left, middle, right). The images of the second row are generated with $\rho = 4$ and $m = \{3, 6, 10\}$ (left, middle, right). This figure appears in color in the electronic version of this article, and any mention of color refers to that version

where the term $1 - \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})^2$ is the spatial term and $\gamma(m, \boldsymbol{\Sigma}) = \frac{2}{m} Tr(\boldsymbol{\Sigma}\boldsymbol{\Sigma}) + \frac{2}{m} Tr^2(\boldsymbol{\Sigma})$ is the nonspatial term. The property of spatial separability makes the residual variation more transparent: The cross-dependence of p.d. matrices depends on $\gamma(m, \boldsymbol{\Sigma})$ such that an increasing m, primarily controlling the variance of a Wishart matrix, leads to smaller cross-dependence; a larger spatial correlation of the underlying Gaussian processes leads to larger spatial dependence.

The spatial dependence can be visualized via realizations of the standard spatial Wishart processes ($\boldsymbol{\Sigma} = \boldsymbol{I}$) on a 20×20 grid with spacing of 1 between adjacent grid points. Given that the spatial correlation function is exponential, ie, $\mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\rho) = \exp(-\frac{||\boldsymbol{s}-\boldsymbol{s}'||}{\rho})$, we visualize the p.d. matrices in two dimensions as ellipsoids in Figure 1. In Figure 1, the p.d. matrices are simulated with $m = 3$ and $\rho = 1, 4, 10$, where a larger ρ leads to stronger spatial dependence. In Figure 1, the p.d. matrices are simulated with $\rho = 4$ and $m = 3, 6, 10$, where a larger m leads smaller cross-dependence. Since the three cases in Figure 1 maintain the same level of spatial dependence, we may also identify that the spatial and nonspatial variations are separable.

2.2 | *Regression term*: Cholesky decomposition

Expressing the mean matrix $\boldsymbol{\Sigma}_i(\boldsymbol{s})$ in terms of $\boldsymbol{X}_i$ is not straightforward (Zhu *et al.*, 2009; Yuan *et al.*, 2012) because the responses $\boldsymbol{A}_i(\boldsymbol{s})$ are in the space of p.d. matrices but the covariates $\boldsymbol{X}_i$ are in the Euclidean space. Following Zhu *et al.* (2009), we regress the (k, l)th element of $\boldsymbol{L}_i(\boldsymbol{s})$, denoted as $l_{ikl}(\boldsymbol{s})$ on $\boldsymbol{X}_i$ as

$$\log l_{ikk}(\boldsymbol{s}) = \boldsymbol{X}_i \boldsymbol{\beta}_{kk}(\boldsymbol{s}), \quad l_{ikl}(\boldsymbol{s}) = \boldsymbol{X}_i \boldsymbol{\beta}_{kl}(\boldsymbol{s}) \quad \text{for } k > l, \tag{4}$$

where $\boldsymbol{\beta}_{kl}(\boldsymbol{s}) = [\beta_{0kl}(\boldsymbol{s}), \beta_{1kl}(\boldsymbol{s}), \ldots, \beta_{dkl}(\boldsymbol{s})]^T$ is the spatially-varying coefficient vector and $\beta_{jkl}(\boldsymbol{s})$ is the

coefficient associated with the jth covariate. The roles of the coefficients $\boldsymbol{\beta}_{kl}(\boldsymbol{s})$ can be explained as linear effect on $\log l_{ikk}(\boldsymbol{s})$ or $l_{ikl}(\boldsymbol{s})$. To model the spatial dependence of the mean effects in (4), we assign a mean-zero spatial Gaussian process prior on

$$\boldsymbol{\beta}(\boldsymbol{s}) = [\boldsymbol{\beta}_{11}(\boldsymbol{s})^T, \boldsymbol{\beta}_{22}(\boldsymbol{s})^T, \boldsymbol{\beta}_{33}(\boldsymbol{s})^T, \boldsymbol{\beta}_{21}(\boldsymbol{s})^T, \boldsymbol{\beta}_{31}(\boldsymbol{s})^T, \boldsymbol{\beta}_{32}(\boldsymbol{s})^T]^T,$$

denoted as

$$\boldsymbol{\beta} \sim \mathcal{GP}\big(\mathbf{0}, \mathcal{K}\big(\boldsymbol{s}, \boldsymbol{s}' | \boldsymbol{\Phi}_\beta\big), \sigma_\beta^2 \boldsymbol{I}\big),$$

where $\boldsymbol{\Phi}_\beta$ is a set of spatial parameters controlling the spatial dependence of mean process, and σ_β^2 is the variance of the Gaussian process.

3 | CHOLESKY DECOMPOSITION PROCESS MODEL

In Section 2, we show that the spatial Wishart process model provides a means of modeling spatially dependent p.d. matrices, and the roles of all the parameters are easily interpretable. Viraswami (1991) and others (eg, Blumenson and Miller, 1963; Smith and Garth, 2007) show that a closed-form probability density function of the spatial Wishart process model is not available in general, except for some special cases. Therefore, to approximate the spatial Wishart process model, we further propose the Cholesky decomposition process model.

The Cholesky decomposition process model is specified on the Cholesky decomposition elements of $\boldsymbol{A}_i(\boldsymbol{s})$, denoted as $\{t_{ikl}(\boldsymbol{s}) : k \geq l, \boldsymbol{s} \in \mathcal{D}\}$, where $t_{ikl}(\boldsymbol{s})$ is the (k,l)th element of the lower Choelesky factor of $\boldsymbol{A}_i(\boldsymbol{s})$. Next, we will show that the Cholesky decomposition process is an asymptotic approximation to the spatial Wishart process. This means the interpretations of parameters are shared with those in Section 2. The Cholesky decomposition process model is below.

$$\begin{aligned}
&\textbf{Diagonal}: \sqrt{2}\log t_{ikk} \sim \mathcal{GP} \\
&\Big(\sqrt{2}\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}), C(\boldsymbol{s}, \boldsymbol{s}' | \boldsymbol{\Phi}_u), \sigma_m^2\Big) \quad \text{for } k = 1, 2, 3, \\
&\textbf{Off-Diagonal:} t_{ikl} | \bar{t}_{ikk} \sim \mathcal{GP}\big(\boldsymbol{X}_i\boldsymbol{\beta}_{kl}(\boldsymbol{s}), \\
&C(\boldsymbol{s}, \boldsymbol{s}' | \boldsymbol{\Phi}_u)\bar{t}_{ikk}(\boldsymbol{s})\bar{t}_{ikk}(\boldsymbol{s}'), \sigma_m^2\big) \quad \text{for } k > l. \qquad (5)
\end{aligned}$$

In this expression, $\sqrt{2}\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s})$ and $\boldsymbol{X}_i\boldsymbol{\beta}_{kl}(\boldsymbol{s})$ are marginal means of the diagonal and off-diagonal Gaussian processes at location $\boldsymbol{s}$, respectively. Also, $\bar{t}_{ikk}(\boldsymbol{s}) = \exp(\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{kk}(\boldsymbol{s}))$ and $\widehat{\boldsymbol{\beta}}_{kk}(\boldsymbol{s})$ is the ordinary least squares estimates computed using only data at voxel $\boldsymbol{s}$ from regressing $\log t_{ikk}(\boldsymbol{s})$ on $\boldsymbol{X}_i$. The prior on $\boldsymbol{\beta}(\boldsymbol{s}) = [\boldsymbol{\beta}_{11}(\boldsymbol{s})^T, \boldsymbol{\beta}_{22}(\boldsymbol{s})^T, \boldsymbol{\beta}_{33}(\boldsymbol{s})^T, \boldsymbol{\beta}_{21}(\boldsymbol{s})^T, \boldsymbol{\beta}_{31}(\boldsymbol{s})^T, \boldsymbol{\beta}_{32}(\boldsymbol{s})^T]^T$ is the same as what we defined in Section 2.

To provide a rigorous mathematical validation, we also prove that $\{\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{T}_i(\boldsymbol{s})\boldsymbol{T}_i(\boldsymbol{s})^T : \boldsymbol{s} \in \mathcal{D}\}$ is a valid stochastic process and almost-surely continuous (Theorem 2), where $\boldsymbol{T}_i(\boldsymbol{s})$ is the lower-triangle Cholesky matrix.

Theorem 2 (Cholesky Decomposition Process). *$\{\boldsymbol{A}_i(\boldsymbol{s}) = \boldsymbol{T}_i(\boldsymbol{s})\boldsymbol{T}_i(\boldsymbol{s})^T : \boldsymbol{s} \in \mathcal{D}\}$ is a valid stochastic process that gives proper finite-dimensional distributions for any collections of locations in the spatial domain. Also, if the correlation function $C(\boldsymbol{s}, \boldsymbol{s}' | \boldsymbol{\Phi})$ has a second-order Taylor series, expansion with remainder that goes to 0 at a rate of $2 + \delta$ for some $\delta > 0$, $\boldsymbol{A}(\boldsymbol{s})$ converges weakly to $\boldsymbol{A}(\boldsymbol{s}_0)$ with probability one as $||\boldsymbol{s} - \boldsymbol{s}_0|| \to 0$.*

To link to the spatial Wishart process model, we set $C(\boldsymbol{s}, \boldsymbol{s}' | \boldsymbol{\Phi}_u) = \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}' | \boldsymbol{\Phi}_u)^2$ and $\sigma_m^2 = \frac{1}{m}$. Given the asymptotic properties in Theorem 3, we may conclude that asymptotically ($m \to \infty$), the two models are equivalent and the parameters in the two models have the same interpretations: $\boldsymbol{\beta}_{kk}(\boldsymbol{s})$ controls the mean of $\log t_{ikk}(\boldsymbol{s})$ and partially describes local variation of $t_{ikl}(\boldsymbol{s})$; $\boldsymbol{\beta}_{kl}(\boldsymbol{s})$ controls the mean of $t_{ikl}(\boldsymbol{s})$; $\boldsymbol{\Phi}_u$ controls the spatial residual dependence. Furthermore, if we modify $\bar{t}_{ikk}(\boldsymbol{s})$ in (5) such that $\bar{t}_{ikk}(\boldsymbol{s}) = \exp(\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}))$, the condition that $N \to \infty$ can be omitted to show that $\{\sqrt{m}[e_{ikl}(\boldsymbol{s}_1), \ldots, e_{ikl}(\boldsymbol{s}_n)]^T | \boldsymbol{S}_{e_{ikl}}\}$ is equal in distribution to $\{\sqrt{m}[t_{ikl}(\boldsymbol{s}_1), \ldots, t_{ikl}(\boldsymbol{s}_n)]^T | \bar{t}_{ikk}\}$. However, we show that the specification in (5) leads to computationally efficient Gibbs sampling for coefficients (Section 3.1) and a reasonable trade-off according to the simulation results showing the closeness of parameter estimation (Section 4).

Theorem 3 (Asymptotic Properties). *For $i \in \{1, 2, \ldots, N\}$, let $\{t_{ikl}(\boldsymbol{s}) : k \geq l, \boldsymbol{s} \in \mathcal{D}\}$ and $\{e_{ikl}(\boldsymbol{s}) : k \geq l, \boldsymbol{s} \in \mathcal{D}\}$ be Cholesky decomposition elements of $\{\boldsymbol{A}_i(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ following the Cholesky decomposition process model and the spatial Wishart process model, respectively. For $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$, we have the following asymptotic results:*

- ***Diagonal:*** *As $m \to \infty$, $\sqrt{m}[\log e_{ikk}(\boldsymbol{s}_1) - \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}_1), \ldots, \log e_{ikk}(\boldsymbol{s}_n) - \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}_n)]^T$ is equal in distribution to $\sqrt{m}[\log t_{ikk}(\boldsymbol{s}_1) - \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}_1), \ldots, \log t_{ikk}(\boldsymbol{s}_n) - \boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}_n)]^T$ for $k = 1, 2, 3$;*
- ***Off-Diagonal:***
  - *Let $\boldsymbol{\mu}_{m,kl}$ (dependent of m) and $\boldsymbol{\mu}_{kl}$ (independent of m) be the means of $\{[e_{ikl}(\boldsymbol{s}_1), \ldots, e_{ikl}(\boldsymbol{s}_n)]^T | \boldsymbol{S}_{e_{ikl}}\}$ and $\{[t_{ikl}(\boldsymbol{s}_1), \ldots, t_{ikl}(\boldsymbol{s}_n)]^T | \bar{t}_{ikk}\}$, respectively. As $m \to \infty$, $\boldsymbol{\mu}_{m,kl}$ converges in probability to $\boldsymbol{\mu}_{kl}$, and*

$cor(e_{ikl}(\boldsymbol{s}), e_{ikl}(\boldsymbol{s}')|\boldsymbol{S}_{e_{ikl}})$ *converges in probability to* $cor(t_{ikl}(\boldsymbol{s}), t_{ikl}(\boldsymbol{s}')|\bar{t}_{ikk}) = \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})^2 = C(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi})$;

- *If* $\boldsymbol{\beta}_{kl}(\boldsymbol{s}) = \boldsymbol{0}$ *for all* $\boldsymbol{s} \in \mathcal{D}$ *and* $k > l$, *then as* $m \to \infty$ *and* $N \to \infty$, $\{\sqrt{m}[e_{ikl}(\boldsymbol{s}_1), \dots, e_{ikl}(\boldsymbol{s}_n)]^T|\boldsymbol{S}_{e_{ikl}}\}$ *is equal in distribution to* $\{\sqrt{m}[t_{ikl}(\boldsymbol{s}_1), \dots, t_{ikl}(\boldsymbol{s}_n)]^T|\bar{t}_{ikk}\}$, *for* $k > l$;
- *If* $\boldsymbol{\beta}_{kl}(\boldsymbol{s}) = \boldsymbol{0}$ *for all* $\boldsymbol{s} \in \mathcal{D}$ *and* $k > l$, *and that* $\bar{t}_{ikk}(\boldsymbol{s}) = \exp(\boldsymbol{X}_i\boldsymbol{\beta}_{kk}(\boldsymbol{s}))$, *then as* $m \to \infty$, $\{\sqrt{m}[e_{ikl}(\boldsymbol{s}_1), \dots, e_{ikl}(\boldsymbol{s}_n)]^T|\boldsymbol{S}_{e_{ikl}}\}$ *is equal in distribution to* $\{\sqrt{m}[t_{ikl}(\boldsymbol{s}_1), \dots, t_{ikl}(\boldsymbol{s}_n)]^T|\bar{t}_{ikk}\}$, *for* $k > l$.

In the above expression, $\boldsymbol{S}_{e_{ikl}}$ *denotes the selected components of the latent Gaussian processes of the Wishart process described in Section 2.* $\boldsymbol{S}_{e_{ikl}}$ *are expressed as follows:* $\boldsymbol{S}_{e_{i31}} = \boldsymbol{S}_{e_{i32}} = \{[Z_{ij1}(\boldsymbol{s}), Z_{ij2}(\boldsymbol{s})]^T : \boldsymbol{s} \in \mathcal{D}, j \in \{1, 2, \dots, m\}\}$ *and* $\boldsymbol{S}_{e_{i21}} = \{Z_{ij1}(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}, j \in \{1, 2, \dots, m\}\}$, *where* $\boldsymbol{Z}_{ij}(\boldsymbol{s}) = [Z_{ij1}(\boldsymbol{s}), Z_{ij2}(\boldsymbol{s}), Z_{ij3}(\boldsymbol{s})]^T$ *and* $\boldsymbol{Z}_{ij} \sim \mathcal{GP}(\boldsymbol{0}, \mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi}_u), \boldsymbol{I})$ *(the term is independent distributed over j, and i is a given and fixed subject subscript).*

The asymptotic results apply to large degrees of freedom, m, when the off-diagonal coefficients are zeros. In both models, large m corresponds to small residual variability, ie, images with small noise. This is a reasonable condition in our motivating data (see Section 5) where the estimated residual variance is small. In comparison to the spatial Wishart process model, the Cholesky decomposition process model is more computationally convenient because it transforms matrix responses to scalars who follow Gaussian processes. Moreover, since the underlying mechanism of the DT's spatial dependence is unknown, both models can be treated as proposed geostatistical models for DTI. All the proofs for the results in this section are summarized in Web Appendix B.

3.1 | COMPUTATIONAL DETAILS

In this subsection, we give the computational details of this model. We fit the model using Markov chain Monte Carlo (MCMC) and assign priors to parameters. Given that $\mathcal{K}$ is the Matern correlation function, we define $\boldsymbol{\Phi}_u = \{\rho_u, \nu_u\}$ as the range and smoothness parameter of the residual dependence, and $\boldsymbol{\Phi}_\beta = \{\rho_\beta, \nu_\beta\}$ as the range and smoothness parameter of the mean dependence. We give priors to these parameters: $\log \rho_u$ and $\log \rho_\beta$ follow a normal distribution with mean 0 and standard deviation 1; $\log \nu_u$ and $\log \nu_\beta$ follow a normal distribution with mean -1 and standard deviation 1; σ_β^{-2} and σ_m^{-2} follow a gamma distribution with shape 0.01 and rate 0.01, which are conjugate priors allowing Gibbs sampling. The coefficients $\boldsymbol{\beta}$ are also updated using Gibbs sampling because their full conditional distributions are Gaussian distributions (see Web Appendix C for complete details).

The computational bottleneck of the Cholesky decomposition process model is factoring the large $n \times n$ covariance matrix of the residual dependence and mean dependence (eg, Heaton *et al.*, 2019). We address this problem using Vecchia's method (Vecchia, 1988), a local likelihood approximation that approximates the joint density of spatial variables as a product of conditional densities. Let ω be an arbitrary Gaussian process. The approximate joint density is $p[w(\boldsymbol{s}_1), \dots, w(\boldsymbol{s}_n)] = \prod_{i=1}^{n} p[w(\boldsymbol{s}_i)|w(\boldsymbol{s}_k), \boldsymbol{s}_k \in N(\boldsymbol{s}_i)]$, where $N(\boldsymbol{s}_i)$ is a set of neighboring locations of $\boldsymbol{s}_i$ (Datta *et al.*, 2016). This reduces the computational complexity from $\mathcal{O}(n^3)$ to $\mathcal{O}(nq^3)$, where $q \ll n$ is the largest size of $N(\boldsymbol{s})$. This approximation is implemented for t_{ikl}, $\log t_{ikk}$, and $\boldsymbol{\beta}$, where lexicographical order of locations on the regular spatial grid is used and $N(\boldsymbol{s}_i)$ is the following q locations with larger ranks. A sensitivity analysis is presented in Section 4 to investigate the impact of the tuning parameter q on the Cholesky decomposition process model.

4 | SIMULATION

In this section, we first investigate the performance of the Cholesky decomposition process model under data generated from either the spatial Wishart process or Cholesky decomposition process model, demonstrating that the Cholesky decomposition process model produces reliable results under different geostatistical settings. Also, since we apply Vecchia's approximation for fast computation, we conduct a sensitivity analysis to investigate the impact of q on parameter estimation.

For both models, we generate the synthetic DTs on 20×20 grids with spacing of 1 between adjacent grid points. To mimic a real DTI study, $N = 10$ subjects are simulated with drug-use indicator $x_{i,drug} \in \{0, 1\}$ and normalized age $x_{i,age} \in \mathbb{R}^+$. The simulation study involves 50 replications. For each replication, there are five drug users ($x_{i,drug} = 1$) and five nondrug users ($x_{i,drug} = 0$), and $x_{i,age}$ is generated by a positive half-normal distribution (Leone *et al.*, 1961) with mean 0 and variance 1 to mimic an age distribution that is bounded below by entrance criteria. For each replication, all the coefficients $\boldsymbol{\beta}$ are generated from a spatial Gaussian process with variance $\sigma_\beta^2 = 0.01$ and correlation function $\mathcal{K}(\boldsymbol{s}, \boldsymbol{s}'|\boldsymbol{\Phi}_\beta)$. We have $\beta_{int,kk} = 0$, $\forall \boldsymbol{s}$ and $\beta_{int,kk} = 0$, $\forall \boldsymbol{s}$ for the intercepts. We have $\beta_{drug,kk} = .5$, for $\boldsymbol{s} \in \mathcal{S}$, $\beta_{drug,kk} = 0$, for $\boldsymbol{s} \notin \mathcal{S}$ $\beta_{drug,kl} = 0$, $\forall \boldsymbol{s}$, for coefficients associated with drug. We also have $\beta_{kk,age} = .25$, $\forall \boldsymbol{s}$ and $\beta_{kl,age} = .25$, $\forall \boldsymbol{s}$ for coefficients associated with age. The Gaussian process mean for three covariates simulates a

TABLE 1 Asymptotic ($m = 50$) simulation results for spatially-varying coefficients with the data generated from the Cholesky decomposition process model or the spatial Wishart process model. The results are summarized in terms of mean absolute deviation of posterior mean estimates, 95% posterior coverage, and Monte Carlo standard deviation. The values are averaged over replications, voxels (n), and covariates (d)

		MAD		Coverage 95%		MCSD	
Parameter	q	**CDP**	**SWP**	**CDP**	**SWP**	**CDP**	**SWP**
β_{11}	10	.10	.10	95%	93%	.05	.05
	50	.10	.10	95%	95%	.05	.05
	Standard	.10	.09	95%	96%	.05	.06
β_{22}	10	.10	.10	94%	95%	.05	.05
	50	.10	.10	94%	95%	.05	.06
	Standard	.10	.09	91%	95%	.05	.05
β_{33}	10	.10	.10	94%	94%	.05	.05
	50	.11	.10	93%	96%	.05	.05
	Standard	.11	.09	93%	97%	.05	.06
β_{21}	10	.11	.10	97%	99%	.08	.07
	50	.11	.10	97%	97%	.08	.08
	Standard	.11	.10	93%	97%	.07	.08
β_{31}	10	.11	.10	95%	99%	.07	.07
	50	.11	.10	95%	95%	.07	.08
	Standard	.11	.10	95%	99%	.07	.08
β_{32}	10	.11	.10	95%	99%	.08	.08
	50	.11	.10	95%	99%	.08	.08
	Standard	.11	.10	95%	97%	.07	.08

Abbreviations: CDP, Cholesky decomposition process model; MAD, mean absolute deviation; MCSD, Monte Carlo tandard eviation; SWP, spatial Wishart process model.

scenario that drug has an effect on certain regions of the brain and increasing age may affect the whole brain. In all replications, we simulate the data with $\rho_u = \rho_\beta = 2$, $\nu_u = \nu_\beta = 0.5$. The degrees of freedom are set to $m = 50$ ($m = 3$ and $m = 30$ are in Web Appendix D). To investigate if Vecchia's approximation with different q affects the model performance, we set $q = 10, 50$ and compare it to the model without Vecchia's approximation. For each replication, we collect 5000 MCMC samples after discarding 2000 samples as burn-in.

The simulation results in terms of mean absolute deviation of posterior mean estimates, 95% posterior coverage, and Monte Carlo standard deviation. The mean absolute deviation of posterior mean estimates is defined as $|\mathbb{E}[\theta|.] - \theta|$ where θ is the true value and $\mathbb{E}[\theta|.]$ is the posterior mean; the 95% posterior coverage is defined as the empirical percentage that the true value is in the 95% posterior; the Monte Carlo standard deviation is defined as $\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\theta^{(t)} - \mathbb{E}[\theta|.])^2}$ where $\theta^{(t)}$ is the tth MCMC sample and there are totally T MCMCM samples. The results of $m = 50$ are summarized in Tables 1 and 2; the results of $m = 30$ are summarized in Tables D.1 and D.2 of Web Appendix D and are similar to those with $m = 50$. The values in Table 1 are averaged over replications, voxels (n), and covariates (d). We find that Vecchia's approximation is acceptable because the computational times are 6, 11, and 35 hours for models with 10 neighbors, 50 neighbors, and without Vecchia's approximation and the mean absolute deviation is nearly identical for all the three methods.

We further conduct simulations with $m = 3$ to study the performance when the asymptotic approximation in Theorem 3 is violated. The results of $m = 3$ (Tables D.3 and D.4 of Web Appendix D) have inflated coverage when data are generated from the SWP model. Combining these results, we conclude that if and only if m is large, then the CDP model provides a reasonable approximation to the SWP model.

In addition to small m, a concern with our method is that it assumes continuity and stationarity over space, and thus, that it may perform poorly in the presence of (crossing) fiber tracts. To examine the robustness of the proposal method in these cases, Web Appendix E conducts a simulation with (crossing) fiber tracts. We find that for the cases considered here, the proposed method is able to estimate the population mean DTI in the presence of (crossing) fiber tracts. This agrees with Fuglstad *et al.* (2015) who find that spatial models are often robust to nonstationarity.

TABLE 2 Asymptotic ($m = 50$) simulation results for spatial parameters with the data generated from the Cholesky decomposition process model or the spatial Wishart process model. The results are summarized in terms of mean absolute deviation of posterior mean estimates, 95% posterior coverage, and Monte Carlo standard deviation. The values are averaged over replications

		MAD		Coverage 95%		MCSD	
Parameter	q	**CDP**	**SWP**	**CDP**	**SWP**	**CDP**	**SWP**
$\rho_u = 2$	10	0.17	0.20	98%	98%	0.15	0.19
	50	0.17	0.20	96%	96%	0.16	0.19
	Standard	0.10	0.14	98%	96%	0.16	0.20
$\nu_u = 0.5$	10	0.033	0.033	98%	98%	0.035	0.027
	50	0.033	0.033	98%	98%	0.032	0.029
	Standard	0.022	0.022	97%	96%	0.040	0.031
$\rho_\beta = 2$	10	0.13	0.13	98%	98%	0.20	0.25
	50	0.13	0.13	98%	98%	0.20	0.24
	Standard	0.10	0.13	98%	98%	0.21	0.25
$\nu_\beta = 0.5$	10	0.038	0.038	96%	96%	0.040	0.050
	50	0.038	0.038	96%	98%	0.044	0.052
	Standard	0.038	0.058	98%	96%	0.043	0.053

Abbreviations: CDP, Cholesky decomposition process model; MAD, mean absolute deviation; MCSD, Monte Carlo standard deviation; SWP, spatial Wishart process model.

Next, we compare the performance of the Cholesky decomposition process model and the univariate spatially-varying coefficient model (Gelfand *et al.*, 2003). In clinical studies and neuroimaging, the most interesting covariate effect is the drug-use effect ($x_{i,drug}$) (Brick and Erickson, 1998). The six coefficients comprehensively but not concisely describe the local covariate effects, which may not be interpretable to clinicians who prefer scalar quantities (eg, fractional anisotropy). However, since the six coefficients capture the covariate effects without information loss, our method can accurately project the information onto any clinically meaningful scalar quantity. One of the useful quantities is fractional anisotropy, projecting a p.d. matrix onto [0,1], defined as $f_{FA}(\boldsymbol{A}) = \sqrt{\frac{1}{2}}\frac{\sqrt{(\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_3-\lambda_1)^2}}{\sqrt{\lambda_1^2+\lambda_2^2+\lambda_3^2}}$, where $\{\lambda_1, \lambda_2, \lambda_3\}$ are the eigenvalues of a diffusion tensor $\boldsymbol{A}$ (Ennis and Kindlmann, 2006). To demonstrate this, we estimate the treatment effect of cocaine use on each voxel $\boldsymbol{s}$ in terms of fractional anisotropy, denoted as $\delta_{FA}(\boldsymbol{s}) = f_{FA}(\boldsymbol{\Sigma}^{(1)}(\boldsymbol{s})) - f_{FA}(\boldsymbol{\Sigma}^{(0)}(\boldsymbol{s}))$. Assuming $\boldsymbol{X}_i'$ as the covariates excluding drug use, the term $\boldsymbol{\Sigma}^{(d)}(\boldsymbol{s}) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\boldsymbol{\Sigma}(\boldsymbol{s})|\boldsymbol{X}_i', x_{drug} = d]$ is to describe the averaged (over subjects) mean matrix at voxel $\boldsymbol{s}$ under drug-use status $d \in \{0, 1\}$. We use a Monte Carlo outcome regression estimator (Rotnitzky *et al.*, 1998) to estimate $\delta_{FA}(\boldsymbol{s})$, defined as

$$\widehat{\delta}_{FA}(\boldsymbol{s}) = \frac{1}{N}\sum_{i=1}^{N}\big(\mathbb{E}[f_{FA}[\boldsymbol{\Sigma}_i(\boldsymbol{s})]|\boldsymbol{x}_{i,drug} = 1, rest] - \mathbb{E}[f_{FA}[\boldsymbol{\Sigma}_i(\boldsymbol{s})]|\boldsymbol{x}_{i,drug} = 0, rest]\big), \quad (6)$$

where the expectation can be empirically obtained by MCMC samples.

Since the spatial matrix-variate methods in terms of coefficients estimation have consistent results, we simply use the results of the Cholesky decomposition process model with $q = 10$ and $m = 50$. We plot the posterior means of $\delta_{FA}(\boldsymbol{s})$ in Figure 2, combining all voxels $\boldsymbol{s}$ and replications. We compare it to the univariate spatially-varying coefficient model (Gelfand *et al.*, 2003) with logit transformation of fractional anisotropy as responses and its associated Monte Carlo-based outcome regression estimator (Rotnitzky *et al.*, 1998)

$$\widehat{\delta}_{FA}(\boldsymbol{s}) = \frac{1}{N}\sum_{i=1}^{N}\big(\mathbb{E}[f_{logit}^{-1}[y_i(\boldsymbol{s})]|\boldsymbol{x}_{i,drug} = 1, rest] - \mathbb{E}[f_{logit}^{-1}[y_i(\boldsymbol{s})]|\boldsymbol{x}_{i,drug} = 0, rest]\big), \quad (7)$$

where $y_i(\boldsymbol{s})$ is response and f_{logit} is the logit transformation function. In Figure 2, the Cholesky decomposition process model produces more precise estimates for $\delta_{FA}(\boldsymbol{s})$ for $\boldsymbol{s} \in \mathcal{S}$ with smaller uncertainties, revealing that utilizing the whole matrix information plays a key role in detecting covariate effects. This claim is further verified in real data analysis (see Figure 4).

5 | APPLICATION: COCAINE USER DATA

In this section, we apply the model to a data set of cocaine users (Ma *et al.*, 2017). The data are provided by the

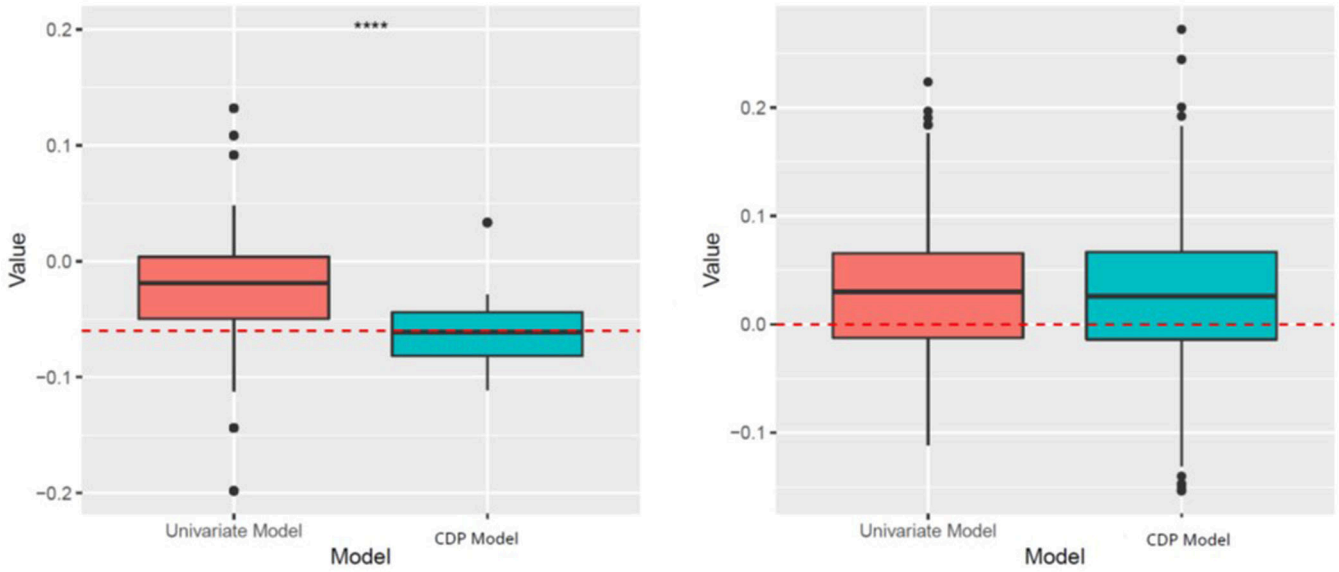

FIGURE 2 The left panel is $\delta_{FA}(\boldsymbol{s})$ for $\boldsymbol{s} \in \mathcal{S}$. The right panel is $\delta_{FA}(\boldsymbol{s})$ for $\boldsymbol{s} \notin \mathcal{S}$. Estimates of $\delta_{FA}(\boldsymbol{s})$ produced by the Cholesky decomposition process model and the univariate model. The red dashed lines are the true values. $\mathcal{S}$ is a set of spatial locations inside a 4×4 region in the middle of the image. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

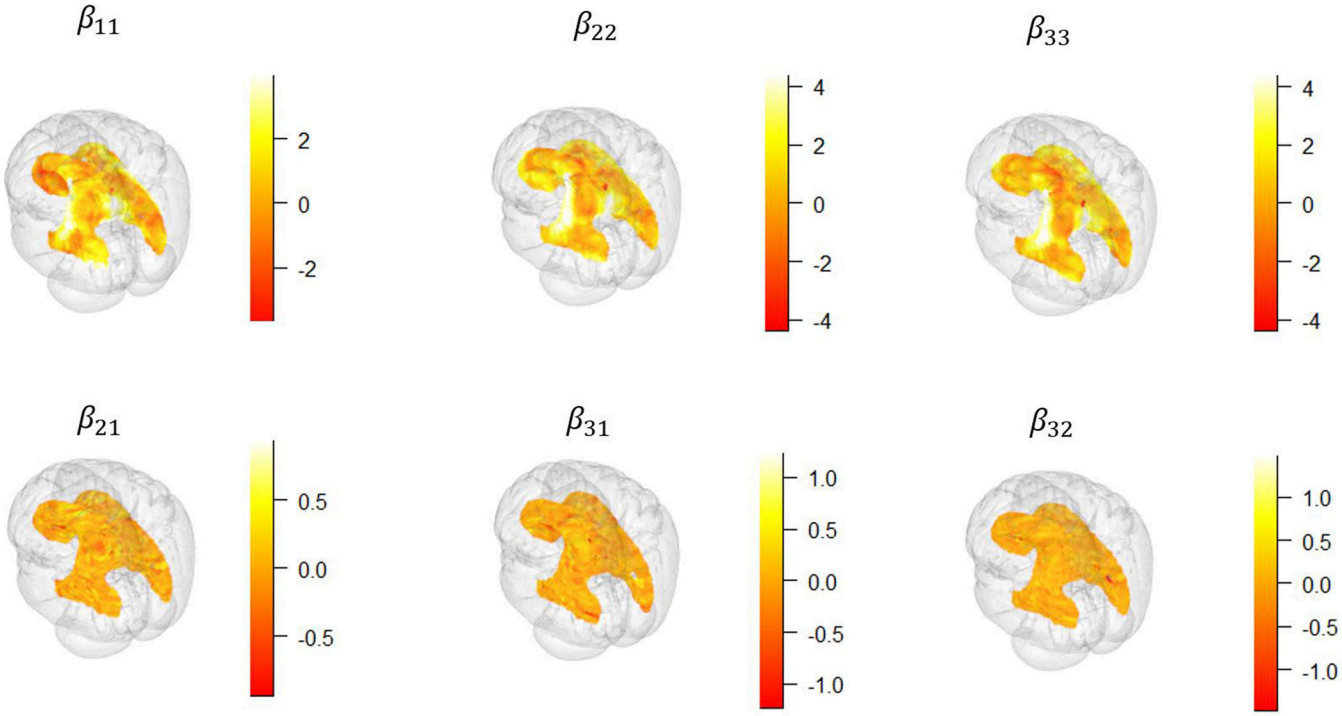

FIGURE 3 The covariate effects of cocaine use on DT expressing by the posterior z-scores of six spatially-varying coefficients. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

Institute for Drug and Alcohol Studies of Virginia Commonwealth University (VCU). Eleven cocaine users and eleven non-cocaine users participated in this study. Besides their cocaine-use status, their age and education years are also recorded. Following the conventions in cocaine use studies (eg, Lane *et al.*, 2010; Ma *et al.*, 2017), we focus on the corpus callosum, a brain region that plays important roles such as transferring motor, sensory, and cognitive

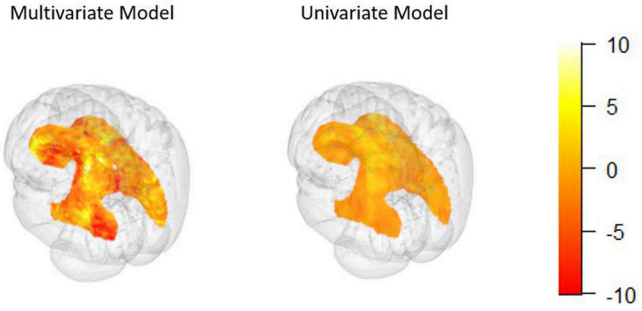

FIGURE 4 The left panel is the posterior z-scores of δ_{FA} based on the Cholesky decomposition process model. The right panel is the posterior z-scores of δ_{FA} based on the univariate model. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

information between the brain hemispheres. This region contains 15 273 voxels.

We first fit the Cholesky decomposition process model to the data, in order to investigate the covariate effects and spatial dependence among voxels. We set the design matrix $\boldsymbol{X}_i$ as

$[1, x_{i,drug}, x_{i,age}, x_{i,edu}, x_{i,handedness}, x_{i,gender}]$, representing the intercept, drug-use ($x_{i,drug} = 1$ if subject i is a cocaine user, otherwise $x_{i,drug} = 0$), age, education years, handedness ($x_{i,handedness} = 1$ if subject i is a left-handed, otherwise $x_{i,handedness} = 0$), and gender ($x_{i,gender} = 1$ if subject i is female, otherwise $x_{i,gender} = 0$). All continuous covariates are standardized. We set $q = 50$ for Vecchia's approximation and 8000 MCMC samples are collected after 3000 samples as burn-in. We apply the proposed stationary model and note that our simulations suggest a degree of robustness to the stationarity assumption; however, for applications with severe nonstationarity, a more sophisticated model could be applied such as those discussed in Section 6.

To understand the spatial dependence of the DTs, we study the posterior density of the spatial dependence parameters ρ_u, ν_u, ρ_β, and ν_β, where the posterior mean estimates are 3.034, 1.75, 3.17, and 1.51, respectively, and the 95% credible regions are [2.81,3.19], [1.39,1.93], [2.89,3.38], and [1.28,1.67], respectively. These results indicate spatial dependence. Therefore, these spatial parameters are well identified by the large multisubject data set. The approximation of the Cholesky decomposition process to the spatial Wishart process holds if $m = \sigma_m^{-2}$ is large. For our data, the posterior 95% interval of $m = \sigma_m^{-2}$ is [33.13,33.89] and the off-diagonal coefficients are close to zeros. Both support the asymptotic approximation. We also explored prior sensitivity by refitting the model with the uniform prior ranging from 0 to 1000 on the spatial range and smoothness. The 95% credible regions for ρ_u, ν_u, ρ_β and ν_β are [2.84,3.32], [1.14,1.73], [2.61,3.21], and [1.11,1.70], respectively, which are similar to the original fit, and so we conclude that the results are not sensitive to the prior.

To test for significance and quantify uncertainties, the covariate effects expressed by their posterior z-scores (Louis, 1984) are visualized. The posterior z-scores are defined as $\frac{\mathbb{E}[\theta|.]}{SD[\theta|.]}$, where $\mathbb{E}[\theta|.]$ is the posterior mean and $SD[\theta|.]$ is the posterior standard deviation. The posterior z-scores of the cocaine use are in Figure 3; the posterior z-scores of the other covariates are in Web Appendix F (Figure F.1-4). All the covariates have effects on certain brain regions. The cocaine use is a covariate whose diagonal coefficients have many absolutely large posterior z-scores located at some regions. We also note that the corpus callosum is strongly oriented in the left-right direction. This may be a reason why the diagonal coefficients $\boldsymbol{\beta}_{kk}$, which can be interpreted as the influence of the covariates on the diffusion in the coordinate directions, capture most of the variations.

Furthermore, we use the estimator in (7) to quantify the drug-use effect on fractional anisotropy $\delta_{FA}(\boldsymbol{s})$. Figure 4 (left) provides the posterior z-scores of $\widehat{\delta}_{FA}(\boldsymbol{s})$ based on the Cholesky decomposition process model and Figure 4 (right) provides the posterior z-scores of the Bayesian outcome regression estimator based on the univariate spatially-varying coefficients model (Gelfand *et al.*, 2003) with the logit transformation of the DTs' fractional anisotropy as responses. The Cholesky decomposition process model provides a more definitive region where cocaine use has a strong effect. This demonstrates the advantages of the matrix-variate modeling over univariate modeling. The regions of differences are located at the splenium, a component at the posterior end of the corpus callosum, indicating group differences between cocaine users and noncocaine users. This result is consistent with previous clinical studies on cocaine use (eg, Lane *et al.*, 2010).

6 | CONCLUSION

In this paper, we propose geostatistical modeling for p.d. matrices with applications to DTI. The spatial Wishart process as a random field for spatially dependent p.d. matrices offers a useful and elegant approach. A bottleneck of the spatial Wishart process model is that the probability density function is intractable. We alleviate this problem by proposing the Cholesky decomposition process model that is composed of Gaussian processes and is asymptotically equivalent to the spatial Wishart process model, which brings the statistical and computational benefits brought from Gaussian processes.

An important area of future work is to formally deal with crossing fiber tracts. Although our simulation study suggests that the proposed method can recover the pop-

ulation mean DTI even in the presence of fiber tracts, adapting the method to this important case may lead to improvements. If the tracts are known *a priori*, than one could propose separate spatial models in each tract. A more sophisticated approach would be to build on spatial modeling for stream networks (Ver Hoef *et al.*, 2006). These methods define a covariance kernel that weighs "river distance" and "spatial distance," which could be adapted to account for "within tract" and "across tract" distances. If the tracts are not known in advance, one may consider a treed Gaussian process (Gramacy and Lee, 2008) that simultaneously searches for homogeneous subregions and models spatial dependence in the subregions.

ACKNOWLEDGMENTS

The authors are thankful to the Editor, the Associate Editor, and reviewers, whose constructive comments improved the manuscript. They also acknowledge the Institute for Drug and Alcohol Studies at Virginia Commonwealth University for providing the motivating cocaine use data set, and support from NIH grants U54DA038999, R01DE024984, and P30CA016059.

DATA AVAILABILITY STATEMENT

The clinical research data (Ma *et al.*, 2017) can be requested from Institute for Drug and Alcohol Studies, Virginia Commonwealth University (VCU), Richmond, VA, USA (https://idas.vcu.edu/). Restrictions apply to the availability of these data.

ORCID

Zhou Lan https://orcid.org/0000-0001-5881-2827
Dipankar Bandyopadhyay https://orcid.org/0000-0001-5421-1725

REFERENCES

Blumenson, L. and Miller, K. (1963) Properties of generalized Rayleigh distributions. *The Annals of Mathematical Statistics*, 34, 903–910.

Brick, J. and Erickson, C.K. (1998) *Drugs, the Brain, and Behavior: The Pharmacology of Abuse and Dependence*. New York: Haworth Medical Press.

Cressie, N. (1992) Statistics for spatial data. *Terra Nova*, 4, 613–617.

Datta, A., Banerjee, S., Finley, A.O. and Gelfand, A.E. (2016) Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111, 800–812.

Dryden, I.L., Koloydenko, A. and Zhou, D. (2009) Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, 3, 1102–1123.

Ennis, D.B. and Kindlmann, G. (2006) Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 55, 136–146.

Fuglstad, G.-A., Simpson, D., Lindgren, F. and Rue, H. (2015) Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, 14, 505–531.

Gelfand, A.E., Diggle, P., Guttorp, P. and Fuentes, M. (2010). *Handbook of Spatial Statistics*. Boca Raton, FL: CRC Press.

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.

Gelfand, A.E., Schmidt, A.M., Banerjee, S. and Sirmans, C. (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13, 263–312.

Gramacy, R.B. and Lee, H.K.H. (2008) Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103, 1119–1130.

Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., et al. (2019) A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24, 398–425.

Karagiannidis, G.K., Zogas, D.A. and Kotsopoulos, S.A. (2003) An efficient approach to multivariate Nakagami-m distribution using Green's matrix approximation. *IEEE Transactions on Wireless Communications*, 2, 883–889.

Kent, J.T. (1989) Continuity properties for random fields. *The Annals of Probability*, 7, 1432–1440.

Kuo, P.-H., Smith, P.J. and Garth, L.M. (2007) Joint density for eigenvalues of two correlated complex Wishart matrices: characterization of Mimo systems. *IEEE Transactions on Wireless Communications*, 6, 3902–3906.

Lan, Z., Reich, B.J. and Bandyopadhyay, D. (2021) A spatial Bayesian semiparametric mixture model for positive definite matrices with applications to diffusion tensor imaging. *The Canadian Journal of Statistics*, *Early View*, 3902–3906. https://doi.org/10.1002/cjs.11601.

Lane, S.D., Steinberg, J.L., Ma, L., Hasan, K.M., Kramer, L.A., Zuniga, E.A., et al. (2010) Diffusion tensor imaging and decision making in cocaine dependence. *PLoS One*, 5, e11591.

Lee, H.N. and Schwartzman, A. (2017) Inference for eigenvalues and eigenvectors in exponential families of random symmetric matrices. *Journal of Multivariate Analysis*, 162, 152–171.

Leone, F., Nelson, L. and Nottingham, R. (1961) The folded normal distribution. *Technometrics*, 3, 543–550.

Liu, B., Qiu, X., Zhu, T., Tian, W., Hu, R., Ekholm, S., Schifitto, G. and Zhong, J. (2016) Spatial regression analysis of serial DTI for subject-specific longitudinal changes of neurodegenerative disease. *NeuroImage: Clinical*, 11, 291–301.

Lo, C.-Y., Wang, P.-N., Chou, K.-H., Wang, J., He, Y. and Lin, C.-P. (2010) Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in Alzheimer's disease. *Journal of Neuroscience*, 30, 16876–16885.

Louis, T.A. (1984) Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association*, 79, 393–398.

Ma, L., Steinberg, J.L., Wang, Q., Schmitz, J.M., Boone, E.L., Narayana, P.A. and Moeller, F.G. (2017) A preliminary longitudinal study of white matter alteration in cocaine use disorder subjects. *Drug and Alcohol Dependence*, 173, 39–46.

Martín-Fernández, M., Westin, C.-F. and Alberola-López, C. (2004) 3d bayesian regularization of diffusion tensor mri using multivariate Gaussian Markov random fields. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 351–359.

Oksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Berlin; Heidelberg: Springer Science & Business Media.

Pourahmadi, M. (2007) Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance–correlation parameters. *Biometrika*, 94, 1006–1013.

Rotnitzky, A., Robins, J.M. and Scharfstein, D.O. (1998) Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, 93, 1321–1339.

Schwartzman, A. (2006). *Random ellipsoids and false discovery rates: Statistics for diffusion tensor imaging data*. PhD thesis, Stanford University.

Schwartzman, A. (2016) Lognormal distributions and geometric averages of symmetric positive definite matrices. *International Statistical Review*, 84, 456–486.

Smith, P.J. and Garth, L.M. (2007) Distribution and characteristic functions for correlated complex Wishart matrices. *Journal of Multivariate Analysis*, 98, 661–677.

Soares, J.M., Marques, P., Alves, V. and Sousa, N. (2013) A Hitchhiker's guide to diffusion tensor imaging. *Frontiers in Neuroscience*, 7, 31.

Spence, J.S., Carmack, P.S., Gunst, R.F., Schucany, W.R., Woodward, W.A. and Haley, R.W. (2007) Accounting for spatial dependence in the analysis of SPECT brain imaging data. *Journal of the American Statistical Association*, 102, 464–473.

Vecchia, A.V. (1988) Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50, 297–312.

Ver Hoef, J.M., Peterson, E. and Theobald, D. (2006) Spatial statistical models that use flow and stream distance. *Environmental and Ecological statistics*, 13, 449–464.

Viraswami, K. (1991) *On multivariate gamma distributions*. Master's thesis, McGill University.

Wu, G.-R., Stramaglia, S., Chen, H., Liao, W. and Marinazzo, D. (2013) Mapping the voxel-wise effective connectome in resting state fMRI. *PLoS One*, 8, e73670.

Xue, W., Bowman, F.D. and Kang, J. (2018) A Bayesian spatial model to predict disease status using imaging data from various modalities. *Frontiers in Neuroscience*, 12, 184.

Yuan, Y., Zhu, H., Lin, W. and Marron, J.S. (2012) Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74, 697–719.

Zhu, H., Chen, Y., Ibrahim, J.G., Li, Y., Hall, C. and Lin, W. (2009) Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *Journal of the American Statistical Association*, 104, 1203–1212.

Zhu, T., Hu, R., Tian, W., Ekholm, S., Schifitto, G., Qiu, X. and Zhong, J. (2013) Spatial regression analysis of diffusion tensor imaging (spread) for longitudinal progression of neurodegenerative disease in individual subjects. *Magnetic resonance imaging*, 31, 1657–1667.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 4 and 5, and R code for implementation are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Lan Z, Reich BJ, Guinness J, Bandyopadhyay D, Ma L, Moeller FG. Geostatistical modeling of positive-definite matrices: An application to diffusion tensor imaging. *Biometrics*. 2022;78:548–559. https://doi.org/10.1111/biom.13445