

# Hierarchical Spatio-Temporal Modeling of Resting State fMRI Data



Alessia Caponera, Francesco Denti, Tommaso Rigon,  
Andrea Sottosanti and Alan Gelfand

**Abstract** In recent years, state of the art brain imaging techniques like Functional Magnetic Resonance Imaging (fMRI), have raised new challenges to the statistical community, which is asked to provide new frameworks for modeling and data analysis. Here, motivated by resting state fMRI data, which can be seen as a collection of spatially dependent functional observations among brain regions, we propose a parsimonious but flexible representation of their dependence structure leveraging a Bayesian time-dependent latent factor model. Adopting an assumption of separability of the covariance structure in space and time, we are able to substantially reduce the computational cost and, at the same time, provide interpretable results. Theoretical properties of the model along with identifiability conditions are discussed. For model fitting, we propose a MCMC algorithm to enable posterior inference. We illustrate our work through an application to a dataset coming from the ENKIRS project, discussing the estimated covariance structure and also performing model selection along with network

---

A. Caponera

Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy  
e-mail: [Alessia.caponera@uniroma1.it](mailto:Alessia.caponera@uniroma1.it)

F. Denti

Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy  
e-mail: [f.denti2@campus.unimib.it](mailto:f.denti2@campus.unimib.it)

T. Rigon (✉)

Department of Decision Sciences, Bocconi University, Milan, Italy  
e-mail: [tommaso.rigon@phd.unibocconi.it](mailto:tommaso.rigon@phd.unibocconi.it)

A. Sottosanti

Department of Statistical Sciences, University of Padova, Padua, Italy  
e-mail: [sottosanti@stat.unipd.it](mailto:sottosanti@stat.unipd.it)

A. Gelfand

Department of Statistical Science, Duke University, Durham, North Carolina, USA  
e-mail: [alan@stat.duke.edu](mailto:alan@stat.duke.edu)

© Springer Nature Switzerland AG 2018

A. Canale et al. (eds.), *Studies in Neural Data Science*, Springer Proceedings in Mathematics & Statistics 257, [https://doi.org/10.1007/978-3-030-00039-4\\_7](https://doi.org/10.1007/978-3-030-00039-4_7)

analysis. Our modeling is preliminary but offers ideas for developing fully Bayesian fMRI models, incorporating a plausible space and time dependence structure.

**Keywords** Bayesian factor analysis · Gaussian processes  
Low-rank factorizations · Separable models

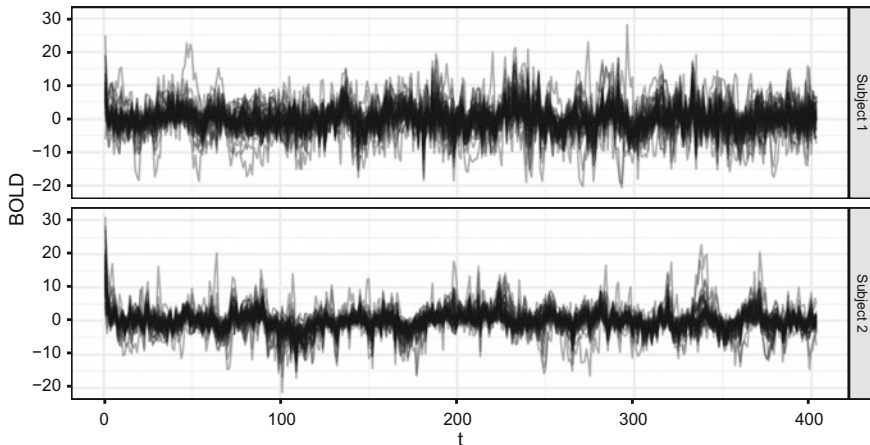
## 1 Introduction

Functional magnetic resonance imaging (fMRI) is an imaging technique which allows the study of human brain activities without being invasive. Such a technique provides a high resolution 3D image reconstruction of a human brain, starting from the blood-oxygen-level dependent (BOLD) signal. The BOLD value is the difference in magnetization between oxygenated and deoxygenated blood, arising from changes in regional cerebral blood flow. In particular, the data at our disposal consist of a collection of BOLD signals obtained from a resting state functional magnetic resonance imaging (RS- fMRI) session. This means that the subjects were not performing any explicit task during the scan. Refer, for instance, to [1–4] for detailed discussions on RS- fMRI data, statistical techniques commonly employed, and medical implications.

From a modeling perspective, what emerges from a RS- fMRI scan is a collection of spatially dependent functional observations. This kind of data collection has encouraged the development of suitable statistical techniques, and indeed, several novel spatio-temporal and dynamic models have been proposed (e.g., [5–10]). Within the Bayesian framework, comprehensive reviews of the main statistical methodologies employed for fMRI data are given in [11, 12].

Since we are proposing a preliminary specification here, we focus on the model of one subject at time, i.e., of a single brain, which is usually referred to as single-subject analysis. Although such an approach does not account for borrowing of information across subjects, it simplifies the modeling process and the related estimation procedures. An early reference to single-subject analysis is given by [13], who propose a general linear model to learn about the blood activity of a single brain. Several contributions appeared afterwards in the context of single-subject modeling and we mention here just a few. For example, the authors of [7] specify a Gaussian random field to capture the spatial correlation, while in [5] the spatial dependence is induced through a hierarchical specification of the parameters.

One of the main goals in the analysis of RS- fMRI data is to study the complex covariance structure between brain regions [3, 14]. In this paper, we propose a Bayesian factor model for fMRI data which is based on the structural assumption of separability. This means that, with regard to the dependence in brain activity across regions, we assume that the covariance structure can be split into two multiplicative components: the spatial and the temporal one. Mainly motivated by the high computational cost that would arise by using a non-separable specification, separability has been employed in several fMRI applications (e.g. [6, 7, 15]). The model presented in this paper benefits from this simplifying assumption, which, in addition,



**Fig. 1** BOLD functional activities for subjects  $i = 1, 2$ , for the  $L$  brain regions of the [16] parcellation and  $t = 1, \dots, 403$ , after the C-PAC pre-processing

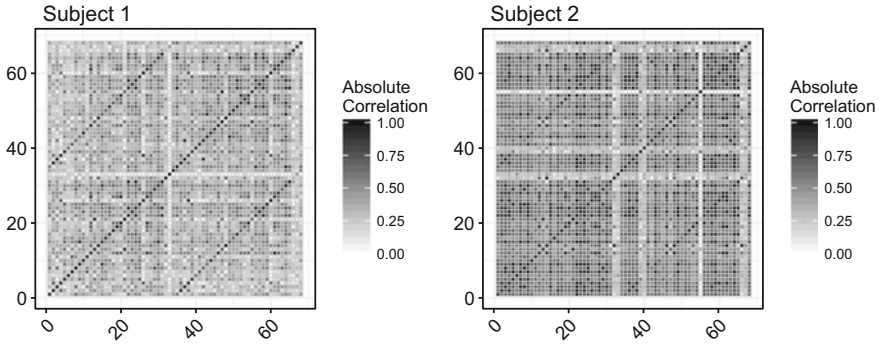
provides interpretable inferential results. This allows the assessment of functional connectivity across brain regions.

The paper is organized as follows. In Sect. 2, we introduce the RS- fMRI dataset at our disposal and we conduct some explorative analysis. In Sect. 3, we specify a single-subject Bayesian factor model for the blood functional activity, which accounts also for temporal dependence. In Sect. 4 we present a Markov Chain Monte Carlo (MCMC) for fitting the proposed model. In Sect. 5, we discuss the performance of our model and we present some empirical results. Concluding remarks are given in Sect. 6.

## 2 The RS- fMRI Dataset

Our dataset comes from the pilot study of the Enhanced Nathan Kline Institute-Rockland Sample project (ENKIRS), which aims at providing a publicly available large sample of multimodal neuroimaging data. Comprehensive information about the project can be found at the link [http://fcon\\_1000.projects.nitrc.org/indi/CoRR/html/nki\\_1.html](http://fcon_1000.projects.nitrc.org/indi/CoRR/html/nki_1.html). From the original multimodal imaging dataset, we retained the BOLD values of two different subjects, which were randomly chosen among the patients. The BOLD values refer to the  $L = 68$  brain regions of the Desikan atlas parcellation [16], equally divided into the left and right hemispheres, discarding the two regions labeled as unknown.

Each measurement is a functional observation composed of  $T = 403$  equally spaced BOLD values, with a lag of approximately 1400 ms, meaning that our dataset comprises two matrices of size  $L \times T$ , for individuals  $i = 1, 2$ . From the original dataset, two BOLD values were discarded since they were missing. The BOLD functional activities, displayed in Fig. 1, are obtained from the raw RS- fMRI scans through

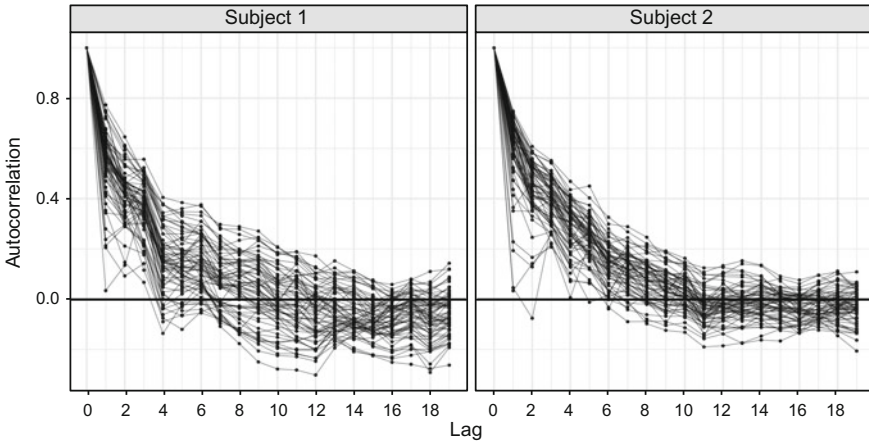


**Fig. 2** Absolute value of the Pearson correlation coefficients among the BOLD functional activities of the 68 brain regions, for subjects  $i = 1, 2$ . Values from 1–34 refer to the left hemisphere, whereas the remaining 35–68 refer to the right hemisphere

an automated pipeline called C-PAC, whose details can be found at <https://fcp-indi.github.io>.

As shown in Fig. 1, the set of regional BOLD functional activities can be regarded as multiple realizations of continuous functions. That is, BOLD, though continuous in time at each region, is evaluated on a finite grid of time  $t = 1, \dots, T$ , while brain regions are specified discretely, being obtained from the Desikan parcellation [16]. There is a considerable statistical literature in spatio-temporal modeling in continuous space and time (see e.g., [17]), particularly in the Bayesian setting [18]. However, data over continuous time and discrete space is rather uncommon in spatio-temporal applications, as pointed out in [19]. In particular, our data should not be modeled, at least in principle, through classical multivariate time series models, since the BOLD activities are continuous in time. Refer for instance to (Chap. 1, [20]) about the use of continuous models for functional observations. Moreover, our data cannot be modeled via standard functional data analysis techniques because some sort of dependence across brain regions is expected. There is need for general modeling methodology for the analysis of this type of RS- fMRI data. We aim to partially fill this gap by introducing a simple spatio-temporal model in the continuous-time and discrete-space framework. Then, we apply it to the RS- fMRI data.

As already mentioned, one of the main goals in the analysis of RS- fMRI data is to study the functional connectivity, e.g. dependence, between brain regions [3, 14]. A simple approach consists in computing the Pearson correlation coefficients between BOLD functional activities, treating the BOLD values as if they were independent over time [2]. The correlation coefficients for subjects  $i = 1, 2$  are shown in Fig. 2. We argue that this strategy, although useful in an explorative phase, could lead to misleading inferential results, for instance revealing fictitious relationships which are due to temporal dependence. In fact, BOLD functional activities are characterized by a non-negligible amount of autocorrelation, as evidenced in Fig. 3, suggesting that Pearson coefficients should be, at the very least, interpreted with care. Nonetheless, correlation matrices like those in Fig. 2 provide an interpretable picture of the BOLD



**Fig. 3** Autocorrelation functions for brain regions  $l = 1, \dots, 68$  and for subjects  $i = 1, 2$

functional connectivity. Additionally, dichotomized versions of these correlations are often at the basis of network analyses for functional connectivity [21, 22]. We aim to preserve this simple structure by seeking a model that naturally leads to an alternative estimate of such a correlation structure, but also takes into account the temporal component.

Some additional difficulties arise when trying to model the spatial component of our RS- fMRI dataset. In particular, areally referenced temporal processes typically rely on some notion of distance, or neighborhood, between different regions, whose definition crucially impacts the results of the analysis. However, given the natural complexity of the brain morphometry, unavoidable questions about the choice of such a distance can be raised [23]. Although there is some evidence that connectivity among brain regions rapidly decays as a function of the Euclidean distance [23, 24], this is often a crude approximation. For instance, as shown in Fig. 2, high levels of connectivity characterize symmetric pairs of brain regions, which are far apart in terms of Euclidean distance. In order to avoid potential misspecification issues, we do not rely on any notion of physical distance between brain regions. Thus, the spatial structure of the BOLD functional activities is reconstructed entirely from the data, without imposing the brain morphometry. This is not to say that there is no potential information in terms of proximity of regions. If a suitable measure reflecting the foregoing caveats were developed, it would provide valuable information and it could improve estimation performance.

### 3 Modeling and Theory

#### 3.1 Low-Rank Multivariate Processes

Consistent with the discussion in Sect. 2, we propose a hierarchical model for RS-fMRI data which (i) accounts for both the spatial and temporal aspects, specifying Gaussian processes for time, and latent factor models for the spatial dimension; (ii) allows a simple interpretation of the functional connectivity among brain regions in terms of a suitable covariance matrix; (iii) avoids misspecification issues by placing few assumptions on the spatial structure. Again, we focus on single-subject models, which means that the two different individuals  $i = 1, 2$  are treated separately and independently, having in common only the model structure. For the sake of notation, we omit the subject index  $i = 1, 2$  and we describe a model for a generic brain. We also note that, with a single subject, we cannot build a regression explanation of response since we cannot include individual level covariate information.

We aim to describe the joint behavior of realizations from a  $L$ -dimensional stochastic process, i.e., the collection of the BOLD functional activities. Formally, we will denote the  $L$ -dimensional stochastic process as  $\mathbf{B}(t)$ , whose entries are the BOLD functional activities  $B_l(t)$ , for  $l = 1, \dots, L$ . We assume a customary additive error structure, that is

$$\mathbf{B}(t) = \mathbf{Z}(t) + \boldsymbol{\epsilon}(t), \quad (1)$$

where  $\boldsymbol{\epsilon}(t)$  is a  $L$ -dimensional pure error, and  $\mathbf{Z}(t)$  is a  $L$ -dimensional process, which we refer to as the mean process. Specifically,  $\boldsymbol{\epsilon}(t)$  is a Gaussian white noise process with variance  $\sigma^2$ , whose entries are independent Gaussian random variables over time and brain regions.<sup>1</sup> Notice that no intercept term is included in specification (1), since, as shown in Fig. 1, the dataset is centered around zero during the C-PAC pipeline.

The overarching goal of our contribution is to infer functional connectivities among brain regions [3]. Therefore, consistent with the available literature and with the descriptive analysis in Sect. 2, the components of the  $L$ -dimensional process  $\mathbf{Z}(t)$  should not be modeled as independent realizations. Moreover, the high dimensionality of our data calls for parsimonious representations, which can be obtained for instance via low-rank approximations. As a notable example, covariance regression models [25, 26] address similar issues and assume that the mean process  $\mathbf{Z}(t)$  can be decomposed as follow:

$$\mathbf{Z}(t) = \mathbf{A}(t)\mathbf{V}(t), \quad (2)$$

where  $\mathbf{A}(t)$  is a  $L \times K$  time varying factor loading matrix, and  $\mathbf{V}(t)$  is a  $K$ -dimensional vector whose entries are independent Gaussian processes—called latent factors in this context. The dimensionality reduction is performed by fixing some

---

<sup>1</sup>Richer modeling might allow heterogeneity in variances, e.g., across regions but we do not consider that here.

$K \ll L$ , e.g.  $K = 3$  or  $K = 5$ . The covariance regression model of [26] is a flexible model for multidimensional stochastic processes, having large support and a familiar interpretation in terms of Bayesian factor models, for any fixed time. Moreover, with discrete time, the covariance regression model could be formally related to the class of dynamic latent factor models (Chap. 10 [27]).

### 3.2 A Time-Dependent Latent Factor Model

We simplify (2) and we set  $A(t) = A$ , that is, the matrix  $A$  is now constant over time and

$$\mathbf{Z}(t) = A\mathbf{V}(t). \quad (3)$$

Hence,  $A$  is a  $L \times K$  factor loading matrix. Although such an assumption reduces the global flexibility of the covariance regression model in Eq. (2), it allows the factor loading matrix  $A$  to be interpreted as a simple measure of dependence, e.g., connectivity, among brain regions, a key feature of our analysis.

In Sect. 5 we provide some empirical support for the factorization (3) as a reasonable assumption for modeling RS- fMRI data while in Sect. 6 we discuss possible extension to the non-stationary case. Additionally, decomposition (3) formally relates our model to the class of latent factor models, which have been used as a dimension reduction tool for instance in genomic applications [28, 29]. Thus, our model can be regarded as a time-dependent extension of a Bayesian factor model, in which the latent factors  $\mathbf{V}(t)$  are independent random functions of time rather than independent draws.

Gaussian processes [30] are a flexible class of stochastic processes to provide random realizations within the space of functions over a specified domain. Therefore, they are a suitable candidate for modeling the time-dependent latent factors  $\mathbf{V}(t)$ . We suppose that the components of  $\mathbf{V}(t)$  are independent and identically distributed Gaussian processes  $\text{GP}(0, \kappa\rho(t, t'))$ , with zero mean and correlation function  $\rho(t, t')$ . As we will discuss in Sect. 3.3, for identifiability purposes we assume  $\kappa = 1$ . Independence among the Gaussian latent factors and the restricted factorization (3) imply a multivariate Gaussian distribution for the mean process  $\mathbf{Z}(t)$  evaluated at a fixed time  $t_0$ , that is

$$\mathbf{Z}(t_0) \sim N_L(0, \boldsymbol{\Sigma}_A), \quad \text{for any fixed } t_0, \quad (4)$$

with  $\boldsymbol{\Sigma}_A = AA^\top$ , which does not depend on time. The role of  $A$  is now clearer since it can be viewed as the square root of the covariance matrix  $\boldsymbol{\Sigma}_A$ . We remark that  $\boldsymbol{\Sigma}_A$  is singular, being of rank  $K \ll L$ . In turn, this implies that  $\mathbf{Z}(t_0)$  for any fixed  $t_0$  would be a degenerate multivariate Gaussian, lying in a subspace of dimension  $K$ . Factorization (3) effectively induces dependence among the components of the mean process, e.g. among brain regions, but implicitly enforces some form of stationarity, since the spatial dependence structure is constant over time. Such an assumption is

discussed in depth for instance in [31], who suggest that it might be worth looking at non-stationary models to obtain a more complete picture of the phenomenon. However, as argued by [31] themselves, stationarity is also convenient in order to prevent the model from becoming vastly more complex. The stationary temporal dependence of our model can be appreciated by observing the covariance matrix between  $\mathbf{Z}(t)$  and  $\mathbf{Z}(t')$ , that is, the so-called cross-covariance matrix

$$\text{Cov}(\mathbf{Z}(t), \mathbf{Z}(t')) = \rho(t, t') \boldsymbol{\Sigma}_A, \quad t \neq t', \quad (5)$$

whose limit as  $|t - t'| \rightarrow 0$  is  $\boldsymbol{\Sigma}_A = \mathbf{A}\mathbf{A}^\top$ , consistently with Eq. (4). Moreover, for  $t \neq t'$  we have  $\text{Cov}(\mathbf{B}(t), \mathbf{B}(t')) = \text{Cov}(\mathbf{Z}(t), \mathbf{Z}(t'))$ . Thus, the cross-covariance in (5) has an appealing interpretation: dependence between BOLD values is multiplicatively adjusted according to temporal proximity.

In practice, we observe the BOLD functional activities only over a finite grid of times  $t = 1, \dots, T$ ; we denote with  $\mathcal{Z}$  the  $L \times T$  matrix containing the values of  $\mathbf{Z}(t)$  over this time grid. Also, let  $\mathcal{B}$  be a  $L \times T$  observed data matrix having entries  $B_l(t)$ , for  $t = 1, \dots, T$ . We can re-express the model of Eqs. (1), (3) and (4) in terms of matrix Gaussian distributions [32], evaluated over the finite time grid:

$$(\mathcal{B} \mid \mathcal{Z}, \sigma^2) \sim \mathcal{N}_{L,T}(\mathcal{Z}, \sigma^2 I_{L \times L}, I_{T \times T}), \quad (6)$$

$$(\mathcal{Z} \mid \mathbf{A}) \sim \mathcal{N}_{L,T}(0, \boldsymbol{\Sigma}_A, \boldsymbol{\Sigma}_T), \quad (7)$$

where  $\boldsymbol{\Sigma}_T$  denotes the Gram-matrix obtained by evaluating the covariance functions  $\rho(t, t')$  over the finite grid  $t = 1, \dots, T$ . Notice that the stationarity assumption translates into a separability assumption in the finite-dimensional setting, since we have

$$(\text{vec}(\mathcal{Z}) \mid \mathbf{A}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_A. \quad (8)$$

This convenient separability result is well described in the spatial literature on multivariate spatial processes: see for instance [18]. Factorization  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_A$  has relevant benefits: it provides a parsimonious representation of the covariance matrix  $\boldsymbol{\Sigma}$  and it facilitates numerical computations. Notice that under the separability assumption the marginal distribution of the rows of  $\mathcal{Z}$  is a multivariate Gaussian with covariance matrix  $\boldsymbol{\Sigma}_T$  and, symmetrically, the columns of  $\mathcal{Z}$  follow a multivariate Gaussian with covariance  $\boldsymbol{\Sigma}_A$ . In other words, the dependence structure over time does not depend on the brain regions, and vice versa.

### 3.3 Identifiability

Without further restrictions, the model described in Eqs. (6) and (7) is not identified. There are two sources of non-identifiability which can be handled by imposing some constraints on the parameters. Notice that  $\mathbf{A}$  appears in Eq. (7) only in terms of



product with its transpose. This means, for instance, that for any orthogonal matrix  $\mathbf{Q}$  such that  $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}_{K \times K}$  we get

$$\mathbf{A}\mathbf{A}^\top = \mathbf{A}\mathbf{Q}\mathbf{Q}^\top\mathbf{A}^\top = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top,$$

where  $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{Q}$ . Thus, it is not possible to discriminate between a model with parameter  $\mathbf{A}$  and another with parameter  $\tilde{\mathbf{A}}$ , since they lead to exactly the same likelihood. Thus, we let  $\mathbf{A}$  to be lower triangular with positive diagonal, as commonly done in coregionalization models in spatial statistics [18]. To avoid confusions: since  $\mathbf{A}$  is a  $L \times K$  rectangular matrix, by lower triangular with positive diagonal we mean that the elements  $a_{lk}$  of  $\mathbf{A}$  are such that  $a_{lk} = 0$  for  $k > l$  and  $a_{kk} > 0$  for  $k = 1, \dots, K$ . Thanks to the Cholesky decomposition for positive semi-definite matrices, under these assumptions the matrix  $\mathbf{A}$  is a Cholesky factor, which uniquely identifies  $\Sigma_{\mathbf{A}} = \mathbf{A}\mathbf{A}^\top$ .

The second source of non-identifiability concerns the scale of the covariance matrices in Eq.(7). For any positive constant  $c \in \mathbb{R}^+$  we have that

$$\Sigma_{\mathbf{A}} \otimes \Sigma_{\mathbf{T}} = (c\Sigma_{\mathbf{A}}) \otimes \left(\frac{1}{c}\Sigma_{\mathbf{T}}\right) = \tilde{\Sigma}_{\mathbf{A}} \otimes \tilde{\Sigma}_{\mathbf{T}},$$

which leads again to non-identifiability. To overcome this difficulty, we set the trace  $\text{Tr}(\Sigma_{\mathbf{T}})$  equal to some constant, which can be easily obtained by letting the scaling parameter of the covariance function  $\kappa$  to be equal to one. Under these constraints, the model is fully identified. As an alternative, one could impose the first, or the last, diagonal entry to be equal to one.

### 3.4 Prior Specification

We conduct inference within the Bayesian framework and therefore we need to specify prior distributions for both  $\mathbf{A}$  and  $\sigma^2$ . In the latter case, we choose an inverse gamma prior for the residual variance, that is

$$\sigma^{-2} \sim \text{Ga}(a_\sigma, b_\sigma), \tag{9}$$

with  $a_\sigma, b_\sigma > 0$  some fixed hyperparameters. In the former case, we can equivalently deal with the coefficients in  $\mathbf{A}$  or with the covariance matrix  $\Sigma_{\mathbf{A}}$ , since they are in a one-to-one correspondence. We formulate the prior distribution in terms of the coefficients in  $\mathbf{A}$ : we let its elements  $a_{lk}$ , for  $l = 1, \dots, L$  and  $k = 1, \dots, K$ , to be independently distributed as follow

$$\begin{aligned}
a_{lk} &\stackrel{\text{iid}}{\sim} N(0, \gamma^2), & l = 1, \dots, L, & \quad k = 1, \dots, K, & \quad k < l, \\
a_{kk}^2 &\stackrel{\text{ind}}{\sim} \gamma^2 \chi_{K-k+1}^2, & k = 1, \dots, K, & & \\
a_{lk} &= 0, & \text{otherwise,} & & 
\end{aligned} \tag{10}$$

for some variance hyperparameter  $\gamma^2 > 0$ . By employing specification (10), we automatically deal with the identifiability constraints of Sect. 3.3.

## 4 Posterior Inference

Posterior inference cannot be conducted in closed form. We need to turn to simulation-based fitting techniques to obtain samples from the posterior distribution. Generally, we would prefer to work with a marginal specification in order to reduce the dimensionality of the problem as much as possible before doing any computation. The *normal-normal* conjugacy enables marginalization of Eq. (6) over  $\mathcal{Z}$ , leading to the following Gaussian model, no longer having a factorized specification

$$(\text{vec}(\mathcal{B}) \mid \mathbf{A}, \sigma^2) \sim N(0, \mathbf{C}), \tag{11}$$

where  $n = L \times T$ , and  $\mathbf{C} = \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_A + \sigma^2 I_{n \times n}$ . The covariance matrix in Eq. (11) is diagonally dominant and thus invertible, allowing to ignore singularity issues that would arise when considering  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_A$  alone. In Appendix A we describe a simple Metropolis-Hastings (M-H) model fitting algorithm with multivariate Gaussian random walk which is based on the marginal specification (11) and is sufficient to guarantee a satisfying mixing. For this purpose, it is convenient to parametrize the residual variance  $\sigma^2$  on the logarithmic scale, i.e.,  $\tau = \log \sigma^2$ . Computational details concerning the M-H sampler are provided in Sect. 4.1, where we describe how to exploit the separability assumption for fast computations.

Suppose we are able to draw posterior samples for  $\mathbf{A}$  and  $\tau$ , for instance by using the M-H in Algorithm 1 (in Appendix). Then, predicting the new BOLD values at a new time and brain region, conditionally on the data, is relatively simple and can be obtained by means of the so-called kriging equations (e.g., [18]), Chap. 2. We remark that kriging the BOLD values is not of direct interest in the analysis of RS-fMRI data. However, as we will discuss in Sect. 5, this procedure is useful to conduct model assessment in terms of out-of-sample prediction performance. Let  $\mathcal{B}_0$  be the  $L_0 \times T_0$  matrix of unobserved BOLD values over a new grid of time values with length  $T_0$  for some subset of  $L_0$  brain regions from the original  $L$  regions. We are interested in finding the predictive distribution

$$p(\mathcal{B}_0 \mid \mathcal{B}) = \int p(\mathcal{B}_0 \mid \mathcal{B}, \theta) p(\theta \mid \mathcal{B}) d\theta, \tag{12}$$

where we have defined  $\theta = (\text{vec}(\mathbf{A}), \tau)$ . The conditional distribution  $p(\text{vec}(\mathbf{B}_0) | \mathbf{B}, \theta)$  is available in closed form, being a multivariate Gaussian distribution

$$\begin{aligned} (\text{vec}(\mathbf{B}_0) | \mathbf{B}, \theta) &\sim \mathbf{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \\ \boldsymbol{\mu}_0 &= \tilde{\mathbf{C}}_0^\top \mathbf{C}^{-1} \text{vec}(\mathbf{B}), \\ \boldsymbol{\Sigma}_0 &= \mathbf{C}_0 - \tilde{\mathbf{C}}_0^\top \mathbf{C}^{-1} \tilde{\mathbf{C}}_0, \end{aligned} \quad (13)$$

where  $\mathbf{C}$  is the covariance matrix of  $\text{vec}(\mathbf{B})$  given the parameters,  $\mathbf{C}_0$  is the covariance matrix of  $\text{vec}(\mathbf{B}_0)$ , and finally  $\tilde{\mathbf{C}}_0$  represent the cross-covariance matrix between  $\text{vec}(\mathbf{B})$  and  $\text{vec}(\mathbf{B}_0)$ . Thus, draws from the predictive distribution in (12) can be obtained by composition sampling, by first drawing posterior values for  $\mathbf{A}$  and  $\tau$  and then by sampling from the multivariate Gaussian distribution in (13).

#### 4.1 Computational Difficulties

Some useful matrix identities can be exploited to reduce the computational burden both for the M-H algorithm and for Eq. (13). We start by inspecting the log-posterior distribution of the marginal model (11) which is equal, up to an additive constant, to the following quantity

$$\mathcal{L}(\mathbf{A}, \tau; \mathbf{B}) = -\frac{1}{2} \log |\mathbf{C}| - \frac{1}{2} \text{vec}(\mathbf{B})^\top \mathbf{C}^{-1} \text{vec}(\mathbf{B}) + \log p(\mathbf{A}) + \log p(\tau), \quad (14)$$

where  $p(\mathbf{A})$  and  $p(\tau)$  denote the probability density functions of the priors for  $\mathbf{A}$  and  $\tau = \log \sigma^2$ , respectively. During the MCMC chain the log-posterior is evaluated several times and therefore it is crucial to maintain computations as fast as possible. A potential computational bottleneck is represented by the inverse of the matrix  $\mathbf{C} = \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_A + \sigma^2 \mathbf{I}_{n \times n}$ , which in our case is a  $n \times n$  matrix, with  $n = T \times L$ . In the separable case, thanks to the properties of the Kronecker product, this issue can be attenuated by exploiting the following decomposition of the inverse of  $\mathbf{C}$ , being equal to

$$\mathbf{C}^{-1} = (\mathbf{U}_T \otimes \mathbf{U}_A)(\boldsymbol{\Lambda}_T \otimes \boldsymbol{\Lambda}_A + \sigma^2 \mathbf{I}_{n \times n})^{-1}(\mathbf{U}_T \otimes \mathbf{U}_A)^\top, \quad (15)$$

where  $\boldsymbol{\Sigma}_T = \mathbf{U}_T \boldsymbol{\Lambda}_T \mathbf{U}_T^\top$  and  $\boldsymbol{\Sigma}_A = \mathbf{U}_A \boldsymbol{\Lambda}_A \mathbf{U}_A^\top$  are the spectral decompositions of the matrices. Detailed calculations leading to (15) are given in Appendix A. The above spectral decompositions are relatively cheap in our context. Notice also that the decomposition of  $\boldsymbol{\Sigma}_T$  has to be computed only once, since it does not depend on unknown parameters in our formulation. More importantly, the matrix  $(\boldsymbol{\Lambda}_T \otimes \boldsymbol{\Lambda}_A + \sigma^2 \mathbf{I}_{n \times n})^{-1}$  is diagonal, and can be, therefore, inverted directly.

Decomposition (15) allows easy evaluation of the log-determinant of  $\mathbf{C}$ , which is given as a simple function of the previously obtained eigenmatrices  $\boldsymbol{\Lambda}_T$  and  $\boldsymbol{\Lambda}_A$ ,

$$\log |\mathbf{C}| = \log |\mathbf{A}_T \otimes \mathbf{A}_A + \sigma^2 \mathbf{I}_{n \times n}| = \sum_{i=1}^n \log (\lambda_i + \sigma^2),$$

where  $\lambda_i$  is the  $i$ -th entry of the diagonal matrix  $\mathbf{A}_T \otimes \mathbf{A}_A$ , for  $i = 1, \dots, n$ . Notice that, as long as  $K \ll L$ , the covariance matrix  $\mathbf{\Sigma}_A$  is not full rank, meaning that some of the eigenvalues  $\lambda_i$  are exactly equal to zero.

Leveraging the decomposition (15) and using some simple properties of the Kronecker product, we can express the quadratic form  $\text{vec}(\mathbf{B})^\top \mathbf{C}^{-1} \text{vec}(\mathbf{B})$  in (14) as follows:

$$\text{vec}(\mathbf{B})^\top \mathbf{C}^{-1} \text{vec}(\mathbf{B}) = \text{vec}(\mathbf{U}_A^\top \mathbf{B} \mathbf{U}_T)^\top (\mathbf{A}_T \otimes \mathbf{A}_A + \sigma^2 \mathbf{I}_{n \times n})^{-1} \text{vec}(\mathbf{U}_A^\top \mathbf{B} \mathbf{U}_T).$$

This drastically reduces the computational burden, since it avoids storage in memory of very large  $n \times n$  matrices. With similar reasoning, the kriging Eq. (13) can also be obtained quite cheaply, adopting fast algorithms for products between matrix involving Kronecker products, implemented for instance in the `klin` R package [33]. The code used in the paper is made available at the link <https://github.com/tommasorigon/StartUpResearch>.

## 5 Data Analysis

### 5.1 Model Checking

We now apply the spatio-temporal model presented in Sect. 3 to the RS- fMRI dataset. However, before proceeding with the interpretation of the results, it is crucial to check the adequacy of the fit to the data to assess the plausibility of the proposed model (Chap. 6, [34]). We measure the goodness of fit of our model by means of out-of-sample predictions, DIC indices, and by direct graphical inspection.

In performing Bayesian inference, we employ the priors described in Sect. 3.4, which require the specification of some tuning parameters. The hyperparameter  $\gamma^2$  controls the prior variability of the coefficients in  $\mathbf{A}$ . By choosing  $\gamma^2 = 100$  we incorporate vague prior information into the model. Following a similar rationale, we let the parameters of the residual variance  $\sigma^2$  to be equal to  $a_\sigma = b_\sigma = 1$ , which induces a fairly noninformative prior for the residual variance.

As discussed in Sect. 3, the temporal component is controlled by the Gaussian processes in  $\mathbf{V}(t)$ , which in turn are characterized by their correlation function  $\rho(t, t')$ . Depending on the choice of such a function, the latent processes  $\mathbf{V}(t)$  could behave quite differently. An extreme example consists in setting  $\rho(t, t') = \mathbb{1}(t = t')$ , with  $\mathbb{1}(\cdot)$  denoting the indicator function, which would imply that the processes  $\mathbf{V}(t)$  are independent over time, and the model in Eqs. (6) and (7) reduces to a simple Bayesian factor model. Instead, by letting  $\rho(t, t') = \exp\{-\psi|t - t'|\}$ , with  $\psi = 3 \times 10^{-2}$ , we introduce temporal dependence favoring stationary and quite regular paths for

**Table 1** For different values of  $K = 1, 3, 5$ , and for subjects  $i = 1, 2$ , the DIC index, the total number of parameters and the out-of-sample root mean squared error (RMSE) are reported. For each individual, it is also shown the out-of-sample RMSE of a random forest model. For both the subjects, the bold values represent the best model according to each index; in both cases, the lower the better

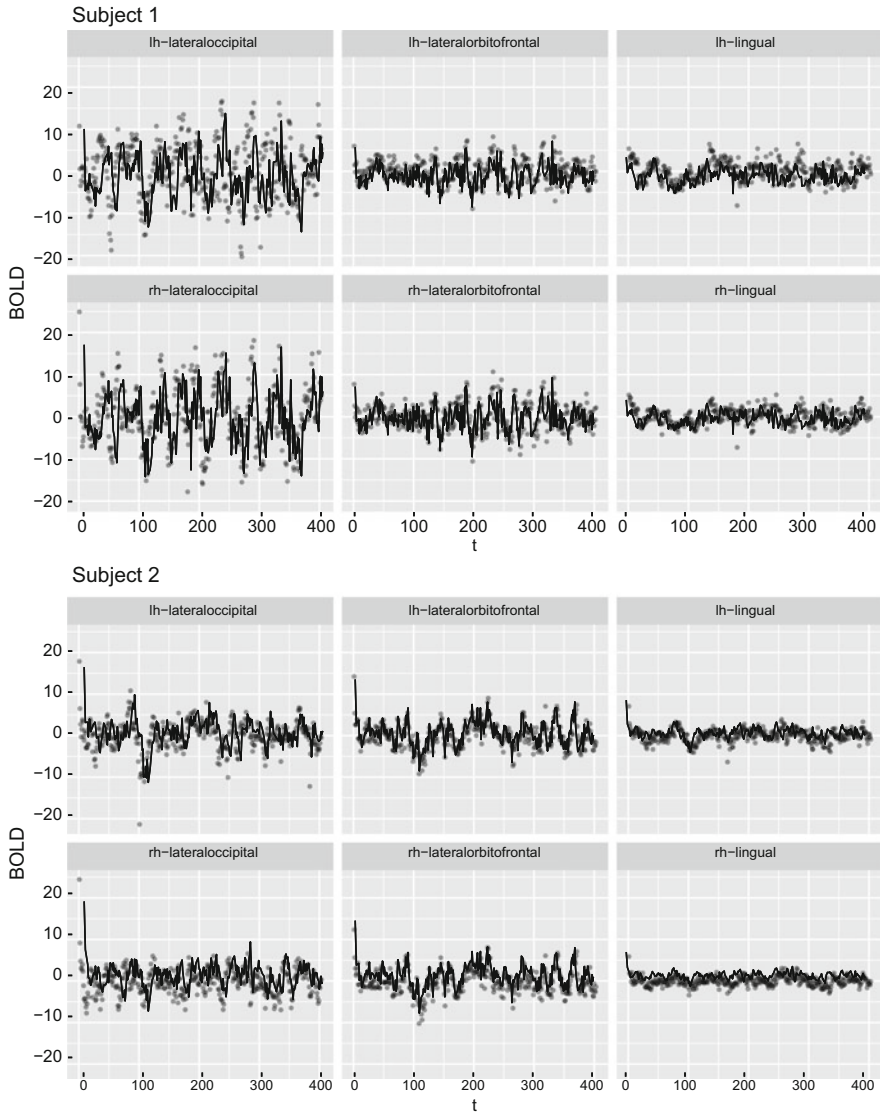
	$K$	DIC	Total # of parameters	RMSE
Subject 1	1	65988.81	69	3.63
	3	60204.04	202	3.34
	5	<b>54800.08</b>	331	<b>3.17</b>
	Random forest	—	—	3.21
Subject 2	1	48108.30	69	2.26
	3	35941.24	202	2.06
	5	<b>31555.90</b>	331	<b>2.00</b>
	Random forest	—	—	2.24

the latent processes  $\mathbf{V}(t)$ . In fact, such a correlation function implicitly induces a continuous-time first order autoregressive process for each element of  $\mathbf{V}(t)$ , with autocorrelation coefficient equal to  $\exp\{-3 \times 10^{-2}\} \approx 0.97$ , which favors fairly regular latent trajectories.

Finally, the number of latent processes  $K$  has to be carefully selected, since its choice critically impacts the computational performance. Indeed, the number of parameters grows linearly as a function of  $K$  and therefore overly complex models become harder to fit using the M-H algorithm. More sophisticated and efficient approaches might mitigate this issue, and a brief discussion is given in Sect. 6. We set  $K = 5$  mainly because of these practical considerations, but we provide below some empirical evidence which offers some reassurance that a model based on such a choice is sufficiently flexible to capture the brain connectivity structure of our data.

To assess whether our model leads to reasonable inferential conclusions and to discriminate between competing models, we conduct some posterior checks, obtaining measures of out-of-sample accuracy as well as the DIC indices [35]. The original dataset is split in two parts: the first one is used for estimation and it comprises the 75% of randomly selected columns of  $\mathcal{B}$ , i.e., different time instants, selected at random, while the remaining 25% is used as a test set to compute for instance the out-of-sample root mean squared error (RMSE).

In Table 1 we compare our model with alternatives involving a smaller number of latent processes, e.g., with  $K = 1$ , or  $K = 3$ , showing that for both the subjects we obtain improved accuracy and lower DIC indices with  $K = 5$ . The out-of-sample predictions are obtained by means of the kriging Eq. (13), after plugging in the MAP estimate. Formally, this is incorrect and may lead to the underestimation of the predictive uncertainty. More correctly, we should take the average of the kriged estimates over posterior realizations of the parameters. However, the latter procedure turns out to be computationally too expensive, so we adopt the aforementioned plug-in alternative. Posterior samples for all the competing models and both the subjects are



**Fig. 4** Scatter plots of the BOLD values for subjects  $i = 1, 2$ , over the time grid  $t = 1, \dots, 403$ , for 6 selected brain regions. Three of these regions are located in the left hemisphere (lh-lateraloccipital, lh-lateralorbitalfronal, lh-lingual), while the others are their symmetric correspondent of the right hemisphere (rh-lateraloccipital, rh-lateralorbitalfronal, rh-lingual). The solid lines represent the predicted values obtained by means of the kriging Eq. (13), after plugging-in the MAP estimate

obtained using the M-H algorithm, which rely on a Gaussian random walk proposal, each with its own covariance matrix. These matrices, one for each competing model in Table 1, have been carefully tuned essentially by trial and error, to ensure good mixing and quick convergence of each MCMC chain. For each MCMC chain we retain 250,000 thinned samples from a chain of 5,000,000 iterations, after a burn-in period of about 100,000 draws. The trace plots show no evidence against convergence and a decent mixing.

As shown in Table 1, we further compare our model with a benchmark method for regression, random forests [36], in which the BOLD response values are fitted as a function of time and brain regions. Although the latter method is specifically designed to provide accurate predictions of response values, our proposal seems to have better out-of-sample performance.

Finally, in Fig. 4 we graphically explore the predictive performance of our model by comparing the original BOLD values with their predictions. For illustrative reasons we displayed only few brain regions, but we remark that the other cases present similar patterns. The graphs of Fig. 4 further corroborate the reasonableness of our proposal, which is able to capture the main trends and the differences in variability of the BOLD values among brain regions.

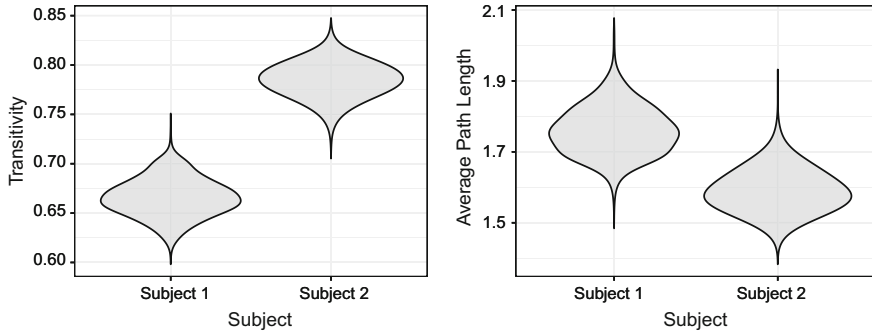
We remark that, in order to reduce the computational burden, the DIC indices of Table 1 and the results in Sect. 5.2 are also based on this 75% partition of the observations, which we believe well-represents the whole dataset.

## 5.2 Network Analysis

In neurological applications it is common practice to explore functional connectivity networks exploiting graph theoretical approaches. As summarized in [21], the typical pipeline of the analysis of structural and functional brain networks consists of the following steps: the identification of the brain regions of interests, the estimation of a continuous measure of association between regions, the application of a threshold to generate a binary adjacency matrix, and the computation of network indices on the obtained undirected graph. In our case, the regions of interests are those obtained from the Desikan parcellation [16], whereas a continuous measure of association can be obtained from the covariance matrix  $\Sigma_A$ , appropriately standardized. Following [22], we define a  $L \times L$  binary adjacency matrix  $\mathbf{G}$  as the truncation of a correlation matrix, that is

$$[\mathbf{G}]_{ll'} = \mathbb{1}([\text{Cor}(\Sigma_A)]_{ll'} > \text{threshold}), \quad \text{for } l \neq l', \quad (16)$$

and  $[\mathbf{G}]_{ll} = 0$  for  $l = 1, \dots, L$  and  $l' = 1, \dots, L$ , where threshold is a constant between 0 and 1, and  $\text{Cor}(\Sigma_A)$  denotes the correlation matrix obtained by standardizing  $\Sigma_A$ . The covariance matrix has a direct interpretation, but the precision matrix, i.e., its inverse, might be considered as well. The choice of the threshold is crucial in determining  $\mathbf{G}$  but, unfortunately, there are no general guidelines. Indeed,



**Fig. 5** Posterior distributions (violin plots) of the transitivity indices and the average path lengths, for subjects  $i = 1, 2$ , evaluated on a graph  $G$  with threshold = 0.8

for any value of the threshold, we could obtain a graph having different sparsity and network properties. To mitigate this issue, we explored a range of plausible thresholds [21] and we noticed that in our setting the inferential conclusions are insensitive to moderate variations of the threshold.

Given the threshold, the adjacency matrix  $G$  is a random quantity whose posterior distribution can be easily approximated using the output of the MCMC. In particular, it is possible to quantify the uncertainty of any network characteristic one could be possibly interested in. Among several alternatives, a relevant network index is the so called clustering coefficient, also known as transitivity in the statistical literature, or fraction of transitive triples, which is a measure of global cohesion of the graph  $G$ . Another index which provides a measure of global connectivity of a given graph is the average path length, defined as the average minimal distance between two brain regions.

We expect these indicators to be negatively correlated in our application: broadly speaking, a high number of transitive triples suggests that two brain regions require a low number of step to be connected. We refer to [37] for the formal definition of these indices and their theoretical properties. In Fig. 5 we reported the posterior distributions of these measures for both of the subjects. We see substantial differences. In particular, subject 2 presents a much higher functional activity compared to subject 1, in terms of both indices. Understanding the qualitative reasons for such a marked distinction between the two subjects is beyond the aim of this paper. Nonetheless, we remark that our proposal was able to capture the differential traits of the two brains, thus providing a tool for detecting differences in functional connectivity and for quantifying the related uncertainty.



## 6 Discussion

In this paper we proposed a spatio-temporal Bayesian factor model for the analysis of RS- fMRI data. Both for interpretational and computational reasons, we employed a separable structure. We discussed how to obtain posterior inference using a M-H algorithm, providing also some technical details that could speed up computations. Finally, we applied our model to a real RS- fMRI dataset and we provided an example.

Although the model we described is designed for a single-subject analysis, it could be extended to the multi-subject case adding a further layer in the hierarchical specification of Sect. 3. One possibility is to borrow information across individuals assuming exchangeable prior distributions for the subject-specific covariance matrices  $\Sigma_A^{(i)}$ . In particular, if we let  $(\Sigma_A^{(i)} | V) \sim \text{InverseWishart}(K, V)$  independently and identically distributed, we could then induce dependence across subjects by placing a hyperprior distribution for  $V$ , which in turn could be interpreted as the baseline covariance structure, common to all the individuals. Additionally, in the multi-subject setting it might be possible to explore the effect of individual covariates on functional connectivity, which we did not attempt, having considered only two subjects.

Another possible extension, already mentioned in Sect. 3.2, could be the implementation of a dynamic model. This would take us to a non separable model by specifying a factor loading matrix  $A(t)$  that also evolves in time. This issue is examined in depth, in a different applied context, by [25, 26]. To capture the evolution of  $A(t)$ , avoiding, at the same time, naive approaches with poor performances, they use independent Gaussian processes with unit variance as a set of basis functions. Thus, the factor loading matrix  $A(t)$  would itself be a time-varying random function, implying that  $\Sigma_A(t) = A(t)A^T(t)$ , for any fixed  $t$ . As a consequence, the evaluation of the adjacency matrix in Eq. (16) for each correlation matrix would generate a dynamic network.

Generalizing our model beyond separability can be done in several other ways. For instance, one could assume that the latent Gaussian processes in  $V(t)$  are independent but not identically distributed, and characterized by different correlation functions  $\rho_k(t, t')$ . This would imply a more sophisticated and non stationary covariance structure for the mean process  $Z(t)$ . Both the above settings are arguably more realistic [31], but unfortunately they do not lead to the simple interpretation which follows from our separable model.

Besides the difficulties in interpretation that could arise from the above generalizations, the main challenge is on the computational side. The algorithm for posterior inference we described in Sect. 4 can be improved in several different directions. For instance, the default prior setting and the parameter expansion strategy of [38] could be adapted to our framework to provide better mixing. In multi-subject scenarios, or whenever the number of brain regions is massive, and therefore MCMC computations are prohibitive, one could attempt deterministic model fitting approximations like variational Bayes. In the context of fMRI data this approach was developed by [9], and it could possibly be adapted to our model.

**Acknowledgements** We are grateful to Greg Kiar and Eric Bridgeford from NeuroData at Johns Hopkins University, who graciously pre-processed the raw DTI and R-fMRI imaging data available at [http://fcon\\_1000.projects.nitrc.org/indi/CoRR/html/nki\\_1.html](http://fcon_1000.projects.nitrc.org/indi/CoRR/html/nki_1.html), using the pipelines NDMG and C-PAC. The authors are also thankful to the organizers of StartUp Research for coordinating such a stimulating event.

## 7 Computational Details

In this appendix we describe a simple Metropolis-Hastings for posterior inference. The algorithm is summarized in Algorithm 1. Additionally, we derive the identity in Eq. (15). Because of orthogonality, we have that  $\mathbf{U}_T \mathbf{U}_T^\top = \mathbf{I}_{T \times T}$  and  $\mathbf{U}_A \mathbf{U}_A^\top = \mathbf{I}_{L \times L}$  and recall also that the matrices  $\mathbf{\Lambda}_T$  and  $\mathbf{\Lambda}_A$  are diagonal, containing the eigenvalues of  $\mathbf{\Sigma}_T$  and  $\mathbf{\Sigma}_A$ , respectively. Exploiting the spectral decompositions of  $\mathbf{\Sigma}_T$  and  $\mathbf{\Sigma}_A$  and the basic properties of the Kronecker product, we get

$$\begin{aligned} \mathbf{C} &= (\mathbf{U}_T \mathbf{\Lambda}_T \mathbf{U}_T^\top) \otimes (\mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^\top) + \sigma^2 \mathbf{I}_{n \times n} \\ &= (\mathbf{U}_T \otimes \mathbf{U}_A) (\mathbf{\Lambda}_T \otimes \mathbf{\Lambda}_A) (\mathbf{U}_T \otimes \mathbf{U}_A)^\top + \sigma^2 \mathbf{I}_{n \times n}. \end{aligned}$$

Then, we can write the identity matrix  $\mathbf{I}_{n \times n} = (\mathbf{U}_T \mathbf{U}_T^\top) \otimes (\mathbf{U}_A \mathbf{U}_A^\top) = (\mathbf{U}_T \otimes \mathbf{U}_A) (\mathbf{U}_T \otimes \mathbf{U}_A)^\top$ . Rearranging the above equation, we get

---

### Algorithm 1: Metropolis-Hastings algorithm for posterior inference

---

**begin**

Let the matrix  $\mathbf{\Sigma}_{\text{Metropolis}}$  be a tuning parameter and let the superscript ( $r$ ) denote the value of the corresponding parameter at the  $r$ -th step of the MCMC chain.

**for**  $r$  from 1 to  $R$  **do**

**Step [1].** Sample a proposed value  $\theta^* = (\text{vec}(\mathbf{A}^*), \tau^*)$  from a multivariate Gaussian

$$\theta^* \sim \mathbf{N}(\theta^{(r)}, \mathbf{\Sigma}_{\text{Metropolis}}),$$

having care that the upper triangular values of  $\mathbf{A}^*$  should be equal to zero. This can be accomplished either by ignoring them in the MCMC step, or forcing the corresponding elements of the  $\mathbf{\Sigma}_{\text{Metropolis}}$  matrix to have zero variance. **Step [2].** Set the acceptance probability  $\alpha$  equal to

$$\alpha = \min \left\{ 1, \exp \left\{ \mathcal{L}(\mathbf{A}^*, \tau^*; \mathbf{B}) - \mathcal{L}(\mathbf{A}^{(r)}, \tau^{(r)}; \mathbf{B}) \right\} \right\},$$

where  $\mathcal{L}(\mathbf{A}, \tau; \mathbf{B})$  denotes the log-posterior distribution (14), up to an additive constant.

**Step [3].** With probability  $\alpha$ , accept the proposed value  $\theta^*$  and set

$$\mathbf{A}^{(r+1)} \leftarrow \mathbf{A}^* \quad \text{and} \quad \tau^{(r+1)} \leftarrow \tau^*.$$


---

$$C = (U_T \otimes U_A)(\Lambda_T \otimes \Lambda_A + \sigma^2 I_{n \times n})(U_T \otimes U_A)^T,$$

from which decomposition of  $C^{-1}$  in Eq. (15) follows directly.

## References

1. Biswal, B., Yetkin, F.Z., V.M., H., Hyde, J.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**(4), 537–541 (1995). <https://doi.org/10.1002/mrm.1910340409>
2. Lee, M.H., Smyser, C.D., Shimony, J.S.: Resting state fMRI: a review of methods and clinical applications. *Am. J. Neuroradiol.* **34**(10), 1866–1872 (2013). <https://doi.org/10.3174/ajnr.A3263>
3. Poldrack, R.A., Mumford, J.A., Nichols, T.E.: *Handbook of Functional MRI Data Analysis*. Cambridge University Press, Cambridge (2011)
4. Smitha, K., Akhil Raja, K., Arun, K., Rajesh, P., Thomas, B., Kapilamoorthy, T., Kesavadas, C.: Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks. *Neuroradiol. J.* **30**(4), 305–317 (2017). <https://doi.org/10.1177/1971400917697342>
5. Bowman, F.D., Caffo, B., Bassett, S.S., Kilts, C.: A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* **39** (2008). <https://doi.org/10.1016/j.neuroimage.2007.08.012>
6. Hartvig, N.G.: A stochastic geometry model for functional magnetic resonance images. *Scand. J. Stat.* **29**(3), 333–353 (2002). <https://doi.org/10.1111/1467-9469.00294>
7. Quirós, A., Diez, R.M., Gamerman, D.: Bayesian spatiotemporal model of fMRI data. *NeuroImage* **49**(1), 442–456 (2010). <https://doi.org/10.1016/j.neuroimage.2009.07.047>
8. Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E., Breakspear, M., Friston, K.J.: Nonlinear dynamic causal models for fMRI. *NeuroImage* **42**(2), 649–662 (2008). <https://doi.org/10.1016/j.neuroimage.2008.04.262>
9. Zhang, L., Guindani, M., Versace, F., Engelmann, J.M., Vannucci, M.: A spatiotemporal non-parametric Bayesian model of multi-subject fMRI data. *Ann. Appl. Stat.* **10**(2), 638–666 (2016). <https://doi.org/10.1214/16-AOAS926>
10. Zhang, L., Guindani, M., Versace, F., Vannucci, M.: A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage* **95**, 162–175 (2014). <https://doi.org/10.1016/j.neuroimage.2014.03.024>
11. Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M.: Bayesian analysis of neuroimaging data in FSL. *NeuroImage* **45** (2009). <https://doi.org/10.1016/j.neuroimage.2008.10.055>
12. Zhang, L., Guindani, M., Vannucci, M.: Bayesian models for functional magnetic resonance imaging data analysis. *Comput. Stat.* **7**(1), 21–41 (2015). <https://doi.org/10.1002/wics.1339>. *Wiley Interdisciplinary Reviews*
13. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**(4), 189–210 (1994). <https://doi.org/10.1002/hbm.460020402>
14. Craddock, R.C., Jbabdi, S., Yan, C., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P.: Imaging human connectomes at the macroscale. *Nat. Methods* **6**, 524–539 (2013). <https://doi.org/10.1038/nmeth.2482>
15. Erhardt, E.B., Allen, E.A., Wei, Y., Eichele, T., Calhoun, V.D.: Simtb, a simulation toolbox for fmri data under a model of spatiotemporal separability. *NeuroImage* **59**(4), 4160–4167 (2012). <https://doi.org/10.1016/j.neuroimage.2011.11.088>

16. Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006). <https://doi.org/10.1016/j.neuroimage.2006.01.021>
17. Cressie, N., Wikle, C.K.: *Statistics for Spatio-Temporal Data*. Wiley, New York (2011)
18. Banerjee, S., Carlin, B.P., Gelfand, A.E.: *Hierarchical Modeling and Analysis for Spatial Data*, 2nd edn. CRC Press, Boca Raton (2014)
19. Quick, H., Banerjee, S., Carlin, B.P.: Modeling temporal gradients in regionally aggregated California asthma hospitalization data. *Ann. Appl. Stat.* **7**(1), 154–176 (2013). <https://doi.org/10.1214/12-AOAS600>
20. Ramsay, J., Silverman, B.W.: *Functional Data Analysis*. Springer, New York (2005)
21. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009). <https://doi.org/10.1038/nrn2575>
22. van den Heuvel, M., Stam, C.J., Boersma, M., Hulshoff Pol, H.E.: Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *NeuroImage* **43**, 528–539 (2008). <https://doi.org/10.1016/j.neuroimage.2008.08.010>
23. Alexander-Bloch, A.F., Vértes, P., Stidd, R., Lalonde, F., Clasen, L., Rapoport, J., Giedd, J., Bullmore, E.T., Gogtay, N.: The anatomical distance of functional connections predicts brain network topology in health and schizophrenia. *Cereb. Cortex* **23**(1), 127–138 (2013). <https://doi.org/10.1093/cercor/bhr388>
24. Salvador, R., Suckling, J., Coleman, M., Pickard, J., Menon, D., Bullmore, E.: Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* **15**, 1332–1342 (2005). <https://doi.org/10.1093/cercor/bhi016>
25. Durante, D., Scarpa, B., Dunson, D.B.: Locally adaptive factor processes for multivariate time series. *J. Mach. Learn. Res.* **15**, 1493–1522 (2014). <http://jmlr.org/papers/v15/durante14a.html>
26. Fox, E., Dunson, D.: Bayesian nonparametric covariance regression. *J. Mach. Learn. Res.* **16**, 2501–2542 (2015). <http://jmlr.org/papers/v16/fox15a.html>
27. Prado, R., West, M.: *Time Series. Modeling, Computation and Inference*. CRC Press (2010)
28. Carvalho, C.M., Chang, J., Lucas, J.E., Nevins, J.R., Wang, Q., West, M.: High-dimensional sparse factor modeling: applications in gene expression genomics. *J. Am. Stat. Assoc.* **103**(484), 1438–1456 (2008). <https://doi.org/10.1198/016214508000000869>
29. West, M.: Bayesian factor regression models in the large p, small n paradigm. In: J. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West (eds.) *Bayesian Statistics*, vol. 7, pp. 733–742. Oxford University Press (2003)
30. Rasmussen, C.E., Williams, C.K.I.: *Gaussian Processes for Machine Learning*. The MIT Press (2006)
31. Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D.: Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* **24**, 663–676 (2014). <https://doi.org/10.1093/cercor/bhs352>
32. Dawid, A.P.: Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika* **68**(1), 265–274 (1981). <https://doi.org/10.2307/2335827>
33. Papp, T.K.: klin: Linear Equations with Kronecker Structure R package version 2007-02-05 (2012). <https://CRAN.R-project.org/package=klin>
34. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: *Bayesian Data Analysis*, 3rd edn. CRC Press (2014)
35. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A.: Bayesian measures of model complexity and fit. *J. R. Stat. Society. Ser. B: Stat. Methodol.* **64**(4), 583–616 (2002). [10.1111/1467-9868.00353](https://doi.org/10.1111/1467-9868.00353)
36. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
37. Kolaczyk, E.D.: *Statistical Analysis of Network Data*. Springer, New York (2009). <https://doi.org/10.1007/978-0-387-88146-1>
38. Ghosh, J., Dunson, D.B.: Default prior distributions and efficient posterior computation in Bayesian factor analysis. *J. Comput. Graph. Stat.* **18**(2), 306–320 (2009). <https://doi.org/10.1198/jcgs.2009.07145>