

# Bayesian Inference for Brain Activity from Functional Magnetic Resonance Imaging Collected at Two Spatial Resolutions

Andrew S. Whiteman<sup>1</sup>, Andreas J. Bartsch<sup>2</sup> Jian Kang<sup>1</sup> and Timothy D. Johnson<sup>1</sup>

<sup>1</sup>*Department of Biostatistics, University of Michigan School of Public Health, e-mail: awhitem@umich.edu e-mail: \*\* jiankang@umich.edu e-mail: †tdjtdj@umich.edu*

<sup>2</sup>*Department of Neuroradiology, University of Heidelberg, e-mail: \*dr.bartsch@radiologie-bamberg.de*

**Abstract:** Neuroradiologists and neurosurgeons increasingly opt to use functional magnetic resonance imaging (fMRI) to map functionally relevant brain regions for noninvasive presurgical planning and intraoperative neuronavigation. This application requires a high degree of spatial accuracy, but the fMRI signal-to-noise ratio (SNR) decreases as spatial resolution increases. In practice, fMRI scans can be collected at multiple spatial resolutions, and it is of interest to make more accurate inference on brain activity by combining data with different resolutions. To this end, we develop a new Bayesian model to leverage both better anatomical precision in high resolution fMRI and higher SNR in standard resolution fMRI. We assign a Gaussian process prior to the mean intensity function and develop an efficient, scalable posterior computation algorithm to integrate both sources of data. We draw posterior samples using an algorithm analogous to Riemann manifold Hamiltonian Monte Carlo in an expanded parameter space. We illustrate our method in analysis of presurgical fMRI data, and show in simulation that it infers the mean intensity more accurately than alternatives that use either the high or standard resolution fMRI data alone.

**Keywords and phrases:** Imaging statistics, Gaussian process, Bayesian nonparametrics, Data integration, Presurgical fMRI.

## 1. Introduction

Neurosurgery presents a unique set of challenges to the operating surgeon. Treatment of brain tumors, for example, is handled primarily by surgical resection when possible, although gliomas are often infiltrative, in which case it may be impossible to remove them entirely (Jovčevska, Kočevár and Komel, 2013; Stipich, 2015). The neurosurgeon’s goal is typically to resect as much of the tumor as possible while avoiding damage to surrounding healthy and eloquent areas of brain tissue, requiring precise structural and functional information. Although the structure of the human brain shares a gross organization common across individuals, functional neuroanatomy may vary between patients and within regions (e.g. Large et al., 2016), highlighting the need for within-patient precision.

Here we propose a model that leverages the massive amount of spatial data available in individual functional magnetic resonance imaging (fMRI) scans to help guide presurgical planning by identifying functionally relevant brain regions in a patient-specific manner.

Traditionally, electrocortical interference is used to map brain functional organization during surgery (e.g. Cordella et al., 2013), but this procedure is highly invasive, lengthens surgery duration, and cannot be incorporated into presurgical planning (Stippich, 2015). Increasingly, clinicians can also opt to use imaging methods to help inform patient-specific presurgical planning and intraoperative neuronavigation (e.g. Archip et al., 2007; Nimsy et al., 2006; Durnez et al., 2013; Silva et al., 2018). fMRI may be used, for example, to map patient-specific functional areas, but the data come with an inherent trade off. Surgeons would like to collect information that is spatially precise, but the fMRI signal-to-noise ratio (SNR) decreases as spatial resolution increases, potentially making functional mapping more difficult (Bodurka et al., 2007). In practice, modern scanners are equipped to handle a variety of image resolutions by modifying magnetic pulse sequences, so radiologists are in principle able to collect any combination of scans advantageous for presurgical planning.

Our motivating datasets come from two separate fMRI experiments in which preoperative patients performed cognitive tasks chosen to localize brain regions involved in language processing (see sections 2 and 3 for details). Each individual patient was administered their task over two separate scanning runs, collected at different spatial resolutions. Details vary by patient, but in both instances one run was collected at “standard” spatial resolution with voxel (volumetric pixel) dimensions measuring approximately  $3 \times 3 \times 3 \text{ mm}^3$ , and the other was collected at “high” spatial resolution with approximately  $2 \times 2 \times 2 \text{ mm}^3$  voxels. Raw image time series data were preprocessed using standard software (Jenkinson et al., 2012; Woolrich et al., 2001) to yield statistical parametric maps for each spatial resolution that summarized patients’ fMRI activation over time. In this paper, we propose a new Bayesian model to integrate both sources of data, leveraging the anatomical/spatial precision of high resolution fMRI and the SNR of standard resolution fMRI for enhanced within-patient precision. The primary goal of our model is to reduce spatial noise while making inferential statements identifying functional regions at the highest resolution available. Conceptually, we accomplish this goal by modeling the mean intensity function of both data sources as a Gaussian process. Gaussian processes induce a probability measure on a functional space with distribution characterized by a mean and covariance function (Rasmussen and Williams, 2006). Conditional on the covariance function hyperparameters, which we estimate from data, we conduct fully Bayesian inference on the mean function measured at voxel locations in the high spatial resolution image.

In addition to spatial precision, computational complexity is also a major concern since excessive latency between preoperative scanning and a patient’s actual surgery is undesirable. Computation with spatial Gaussian process models typically involve decomposition of an  $n \times n$  matrix, where  $n$  is the number of spatial locations. Between the two image types there are over 200,000 unique

spatial locations in each of our motivating datasets, rendering usual computational approaches to inference intractable in most computing environments. Here, we outline a modification of the typical Hamiltonian Monte Carlo (HMC) algorithm that makes this inference not only feasible but computationally efficient. To do so, we propose a dual resolution mapping prior that generalizes the existing Gaussian predictive process framework (e.g. Seeger, Williams and Lawrence, 2003; Rasmussen and Williams, 2006; Banerjee et al., 2008) to our setting with multiple data sources. Our algorithm further harnesses a parameter expansion idea from Wood and Chan (1994) to sample from the posterior using Riemann manifold Hamiltonian dynamics (Girolami and Calderhead, 2011) in an ultrahigh dimensional parameter space.

Our model is related to existing literature from the field of spatial statistics that consider the “change of support problem” (e.g. Gelfand, Zhu and Carlin, 2001; Fuentes and Raftery, 2005; Berrocal, Gelfand and Holland, 2012). Such models have been used, for example, to combine data from air pollution monitoring sites with simulations from physical models for prediction at unobserved locations and model validation. Studies such as these commonly model conditional relationships between data sources, for example by regressing measured air pollution onto physical model output. Our multi-resolution imaging paradigm is related in the sense that we would like to use standard resolution data to improve inference in high resolution space. This goal, however, is complicated by the fact that high and standard spatial resolution voxels in general only partially overlap with their neighbors in their complementary image (see Fig. 1). We will, however, take a different approach by modeling both sources of data as joint outcomes. Not only does this approach perhaps make more conceptual sense for modeling multiple image types, it permits flexible and natural reconfiguration in response to real world challenges. For example, if only one fMRI resolution or session is available presurgically, the missing data can be removed from the joint outcome. Though we discuss our method exclusively in a functional neuroimaging context, the method can easily generalize to other imaging modalities or indeed to spatial data with mixed supports more broadly.

Whereas the inferential goal of most neuroimaging studies is to identify activated or deactivated brain regions while controlling the family-wise error rate, we take a somewhat different approach given specific presurgical needs. In a neurosurgical context, clinicians are typically more concerned with inaccurate labeling of functionally important tissue as unimportant. To this end, we adopt a decision theoretic rule from previous work to control the ratio of false negative to false positive errors (Liu et al., 2016, 2019; Müller, Parmigiani and Rice, 2006). We show in simulation that our dual resolution method achieves good accuracy for realistic effect sizes. Specifically, our method outperformed single spatial resolution alternatives in terms of both false negative and false positive error rates when the number of discoveries was fixed across methods. Software to fit the dual and single resolution models discussed in this paper to data stored in the NIfTI data standard (Cox et al., 2004) is available online at <https://github.com/asw221/dualres>.

The body of this paper contains descriptions of our motivating clinical datasets

in section 2, and a summary of the method we propose to handle the unique challenges of those data in section 2.1. In sections 2.2 and 2.3, we elaborate on our approach to enable precise estimation and computation in such a large parameter space. We discuss a strategy to conduct inference based on weighted trade offs between false negative and false positive errors in section 2.4. Section 3 reports on analyses of real patient data using our proposed method for dual resolution fMRI. We quantify this method’s performance against single resolution alternative methods in section 4, and present an overall evaluation of our contributions in section 5.

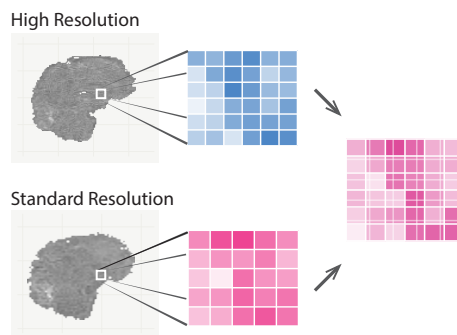


Fig 1: Schematic of aims and difficulties with integration of fMRI data collected at multiple spatial resolutions. Images collected at different resolutions exhibit inherently different levels of noise. We would like to reduce spatial noise while making inferential statements at the highest resolution available, but voxel locations may not align in general.

## 2. Data and methods

We developed the method presented here to analyze single-patient presurgical fMRI data collected at two spatial resolutions. Our first motivating dataset comes from a 62 year old right handed woman—“patient 1”—who presented difficulties with reading, finding, and comprehending words. This patient was subsequently found to have a tumor in her left middle and inferior temporal gyrus. Prior to surgery, the patient was scanned while performing a reading task to map brain areas associated with reading nonfinal embedded clause sentences and language processing. Scans were collected in two separate runs: once at standard  $3 \times 3 \times 3.45 \text{ mm}^3$  resolution ( $64 \times 64 \times 48$  grid), and once at high  $1.8 \times 1.8 \times 2.3 \text{ mm}^3$  resolution ( $120 \times 120 \times 62$  grid).

Our second motivating dataset comes from an 18 year old right handed woman—“patient 2”—who presented after a general seizure. She was subsequently found to have a cavernous malformation (cavernoma) with chronic and subacute hemorrhage (Zabramski type I) in her left temporal lobe close to the transverse temporal gyrus and insular cortex. Cavernomas are a specific type

of arteriovenous malformation without shunting. They contain closely apposed, angiogenetically immature blood vessels, typically with intralesional bleeding residuals. Cavernomas can be treated via microsurgical removal (e.g. Bertalanffy et al., 2002); if left untreated, they may lead to seizures or progressive neurological deficits upon symptomatic micro- or macrohemorrhages. For cavernomas in eloquent locations, presurgical fMRI is considered one option—as with brain tumors—to map the eloquent cortex noninvasively for presurgical planning and intraoperative neuronavigation. Patient 2 was also scanned prior to surgery while performing a language processing task. Her standard resolution data were collected with slightly smaller  $3 \times 3 \times 3.3 \text{ mm}^3$  voxels ( $64 \times 64 \times 48$  grid), and her high resolution data with  $1.8 \times 1.8 \times 2.2 \text{ mm}^3$  voxels ( $120 \times 120 \times 62$  grid). As in this patient, cavernomas typically cause profound T2\*-weighted MR signal loss, with blooming into surrounding brain tissue. Signal loss is caused by abrupt differences in magnetic susceptibility in apposed tissues and is a common occurrence in clinical fMRI (e.g. intratumoral hemorrhages can cause similar dropout). We use this patient’s data to illustrate our model’s capacity to recover an estimate of activation in areas of such fMRI signal loss.

fMRI time series preprocessing without spatial smoothing was performed prior to our analysis using FSL software (Jenkinson et al., 2012) and the FEAT tool (Woolrich et al., 2001). As will become clear in section 2.1, our model imposes smoothness on the image mean function, and so we avoided smoothing the data during preprocessing (beyond the small amount of unavoidable smoothing that can occur when time series images from the two spatial resolutions are motion corrected and registered to one another). Smoothing is an otherwise ubiquitous step in typical fMRI pipelines, but over smoothing is not desirable for presurgical planning applications as it may reduce spatial precision by, for example, smearing activation into adjacent areas when the smoothing kernel is too wide. Data were corrected for motion and temporally high pass filtered, and marginal linear models were fit to the time series data at each voxel to create summary statistic maps of task-related activation.

Preprocessing resulted in one unsmoothed  $z$ -statistic contrast image for each fMRI resolution that summarized task-related activation over the course of each respective scan. We went on to use the generated test statistic maps as outcome data in our subsequent analysis, treating the images as noisy measures of true activation. Although we may find it beneficial to include both spatial and temporal data in our modeling framework in future work, the present model only explicitly represents a spatial process. As such, throughout the rest of this paper we will use “high resolution,” for example, as a stand in for “high *spatial* resolution” etc. In the greater imaging community, however, “resolution” could in general relate to frequency of either spatial or temporal data collection, or both. We give additional details regarding patient data collection and image preprocessing in section 3.

### 2.1. Bayesian dual-resolution mapping

Let  $\mathcal{B}$  denote a generic brain image space, and let  $B_h \subset \mathcal{B}$  and  $B_s \subset \mathcal{B}$  denote the sets of spatial locations in the brain where high and standard resolution functional MRI data are collected, respectively. For reference, the number of voxels  $|B_h| \approx 200,000$  in the high resolution image, and  $|B_s| \approx 50,000$  in the standard resolution image. Each atom  $\mathbf{v} \in \mathcal{B}$  is a three dimensional vector of spatial coordinates relative to some origin point  $\mathbf{v}_0 \in \mathcal{B}$ ; the Euclidean distance between any two points,  $\mathbf{v}, \mathbf{v}' \in \mathcal{B}$  can be represented  $\|\mathbf{v} - \mathbf{v}'\|_2$ , and is typically measured in millimeters. In general, even voxels that overlap between the two image types may not have the same centers, so that the set of points in the intersection  $B_h \cap B_s$  may be empty. Although data at a given voxel is associated with a small volume, we follow common practice and essentially treat that data as observed on location  $\mathbf{v}$  exactly.

Conceptually, we motivate our proposed model as follows. Let  $Y_h(\mathbf{v}_h)$  denote the high resolution imaging outcome at voxel  $\mathbf{v}_h$ , and let  $Y_s(\mathbf{v}_s)$  denote the standard resolution imaging outcome at voxel  $\mathbf{v}_s$ . For the same patient performing the same cognitive task in the same scanner, we make the assumption that  $Y_h(\mathbf{v}_h)$  and  $Y_s(\mathbf{v}_s)$  are realizations from a unifying generative process. Let  $\mathcal{N}(\mu, \sigma^2)$  denote a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . We model the data as jointly Gaussian,

$$\begin{aligned} Y_h(\mathbf{v}_h) &\sim \mathcal{N}(\mu(\mathbf{v}_h), \sigma_h^2), & \mathbf{v}_h &\in B_h \\ Y_s(\mathbf{v}_s) &\sim \mathcal{N}(\mu(\mathbf{v}_s), \sigma_s^2), & \mathbf{v}_s &\in B_s \end{aligned} \quad (1)$$

where  $\mu(\mathbf{v})$  represents the expected intensity of brain activity in voxel  $\mathbf{v} \in \mathcal{B}$ , and  $\sigma_h^2$  and  $\sigma_s^2$  are noise variances in the high and standard resolution images, respectively. Because our data were not smoothed, we modeled noise as a spatially independent and additive process. Given the known phenomenon that SNR increases with voxel volume (e.g. Bodurka et al., 2007), we expect standard resolution images to be less noisy than high resolution images. We therefore adopted a weakly informative prior for the noise variances with the restriction  $\sigma_h^2 > \sigma_s^2$ :

$$\pi(\sigma_h^2, \sigma_s^2) \propto \sigma_h^{-2} \sigma_s^{-2} \mathbb{1}(0 < \sigma_s^2 < \sigma_h^2), \quad (2)$$

where  $\mathbb{1}(\cdot) \in \{0, 1\}$  is the event indicator function ( $\mathbb{1}(\mathcal{A}) = 1$  if  $\mathcal{A}$  occurs, and 0 otherwise).

For functional maps, we are primarily interested in making inferences about the mean intensity function,  $\mu(\cdot)$ , to which we assign a mean zero Gaussian process prior,

$$\mu(\mathbf{v}) \sim \mathcal{GP}(0, K(\mathbf{v}, \mathbf{v}')). \quad (3)$$

In our formulation, the function  $\mu(\cdot)$  captures all of the correlation between voxels and between the two images; conditional on  $\mu(\cdot)$ ,  $Y_h(\mathbf{v}_h)$  and  $Y_s(\mathbf{v}_s)$  are mutually independent across all  $\mathbf{v}_h \in B_h$  and  $\mathbf{v}_s \in B_s$ . We implicitly assume that the two brain images share a real-world coordinate system, and that  $\mu(\mathbf{v})$

is correlated with  $\mu(\mathbf{v}')$  if the distance  $\|\mathbf{v} - \mathbf{v}'\|_2$  is small. A variety of pre-processing techniques have been developed to align 3D images and ensure the former assumption holds with minimal error (e.g. Reuter, Rosas and Fischl, 2010; Jenkinson et al., 2012).

Since anatomical precision is paramount in our application, we would like to conduct inference on  $\mu(\cdot)$  for all locations in  $B_h$ . To facilitate this goal while simultaneously modeling the cross correlation between  $\mu(\cdot)$  evaluated on locations in  $B_h$  and  $B_s$ , we introduce a nonstationary covariance function to map between data sets. For any  $\mathbf{v}, \mathbf{v}' \in \mathcal{B}$ , let,

$$K(\mathbf{v}, \mathbf{v}') = \begin{cases} k(\mathbf{v}, \mathbf{v}') & \text{if } \mathbf{v}' \in B_h \\ w^\top(\mathbf{v})k(B_h, \mathbf{v}') & \text{otherwise,} \end{cases} \quad (4)$$

where  $w(\cdot)$  is a vector of weights in a finite basis (defined below), and  $k(\cdot, \cdot)$  is some positive definite function with range  $\mathbb{R}_{>0}$ . In our application, we take  $k(\cdot, \cdot)$  to be the isotropic radial basis function,

$$k(\mathbf{v}, \mathbf{v}') = \tau^2 \exp(-\psi \|\mathbf{v} - \mathbf{v}'\|_2^\nu), \quad \tau^2, \psi > 0, \quad \nu \in (0, 2], \quad (5)$$

and extend the notation to apply to sets of locations so that  $k(B_h, \mathbf{v}') = [k(\mathbf{v}_h, \mathbf{v}')]_{\mathbf{v}_h \in B_h}$  is a vector in  $\mathbb{R}^{|B_h|}$ . In (5),  $\tau^2 > 0$  is the ‘‘partial sill’’ or marginal prior variance of  $\mu(\cdot)$ , the decay parameter  $\psi > 0$  defines the correlation bandwidth, and  $\nu \in (0, 2]$  is the kernel exponent or smoothness parameter. We define the covariance parameters  $\boldsymbol{\theta} = (\tau^2, \psi, \nu)^\top$ ;  $\psi$  and  $\nu$ , are commonly fixed prior to analyses, but because of the abundance of spatial data in even a single brain image, in practice we recommend estimating these parameters from data (see section 3.1 for details).

By the definition of a Gaussian process,  $\mu(\mathbf{v})$  and  $\mu(\mathbf{v}')$  are jointly multivariate Gaussian distributed for any distinct locations  $\mathbf{v}$  and  $\mathbf{v}'$ . As a result, Gaussian process models promote natural and flexible predictions of values of  $\mu(\cdot)$  at unobserved locations. For arbitrary collections of locations  $U = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathcal{B}$  and  $V = \{\mathbf{v}'_1, \dots, \mathbf{v}'_m\} \subset \mathcal{B}$ , we define  $\mu(U) = [\mu(\mathbf{v}_i)]_{i=1}^n$  as a vector in  $\mathbb{R}^n$ ;  $K(U, \mathbf{v}) = [K(\mathbf{v}_i, \mathbf{v})]_{i=1}^n$  as a vector in  $\mathbb{R}^n$ ; and  $K(U, V) = [K(\mathbf{v}_i, \mathbf{v}'_j)]_{i,j=1}^{n,m}$  as a matrix in  $\mathbb{R}^{n \times m}$ . If, for example,  $V$  is a set of observed locations, and  $U$  are unobserved locations, then conditional on  $\mu(V)$ ,  $\mu(U)$  is multivariate Gaussian distributed with mean  $K(U, V)K(V, V)^{-1}\mu(V)$  and variance  $K(U, U) - K(U, V)K(V, V)^{-1}K(V, U)$ .

Since imaging data is collected on a dense grid we often have no need to predict outcomes at unobserved or non-brain locations, except in cases of signal loss or other artifact. We made use of the kriging or conditional distribution relationships above primarily to define the basis weight function  $w(\cdot)$  to integrate information from both high and standard resolution images. We constructed  $w(\mathbf{v})$  as a sparse approximation of the vector  $K(B_h, B_h)^{-1}k(B_h, \mathbf{v})$  (see section 2.2). This formulation allowed us to leverage the relationship that  $w^\top(\mathbf{v})\boldsymbol{\mu}_h$  approximates the prior conditional expectation of  $\mu(\mathbf{v})$  given  $\boldsymbol{\mu}_h = \mu(B_h)$ . As such, our construction in (4) generalizes a Gaussian predictive process framework (e.g. Seeger, Williams and Lawrence, 2003; Rasmussen and Williams, 2006;

Banerjee et al., 2008) to our setting with multiple data sources by using  $B_h$  as a high-dimensional reference set. In general, within this framework we could have defined the weights  $w(\cdot)$  based on any arbitrary set of knot locations  $B_* \subset \mathcal{B}$ . Since inference at a fine spatial scale typically requires a dense set of knot locations (e.g. Stein et al., 2007), we preferred to define  $w(\cdot)$  based on all of  $B_h$ .

## 2.2. Construction of the covariance weights

Our covariance function in (4) effectively employs kriging methods to map  $\mu(B_h)$  onto the locations in  $B_s$  so that the standard resolution data can still inform  $\mu(B_h)$  in the posterior. Switching to vector notation, let  $\boldsymbol{\mu}_s = \mu(B_s)$ . We express the prior in (3),

$$\pi(\boldsymbol{\mu}_h, \boldsymbol{\mu}_s) = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_h & \mathbf{K}_{h,s} \\ \mathbf{K}_{s,h} & \mathbf{K}_{s,h}\mathbf{K}_h^{-1}\mathbf{K}_{h,s} \end{bmatrix}\right). \quad (6)$$

where  $\boldsymbol{\mu}_h$  and  $\boldsymbol{\mu}_s$  are the means of the high, and standard resolution images, respectively; we denote the marginal prior variance of  $\boldsymbol{\mu}_h$  by  $\mathbf{K}_h = K(B_h, B_h)$ , the prior covariance of  $\boldsymbol{\mu}_h$  and  $\boldsymbol{\mu}_s$  by  $\mathbf{K}_{h,s} = K(B_h, B_s)$ , etc. The obvious difficulty working with (6) directly is that the covariance matrix is large and dense and we need to be able to compute its inverse in order to evaluate the prior. We would like to make inferential statements about  $\boldsymbol{\mu}_h$ , but the dimension of the submatrix  $\mathbf{K}_h$  ( $n_h \approx 200,000$ ) alone is prohibitive on most hardware architectures—such a matrix would require over  $(1.8 \cdot 10^5)^2 \times 32 = 129.6$  Gb of memory just to store in a single precision floating point format. Though the memory requirement could be reduced by storing just the upper or lower triangle, to sample  $\boldsymbol{\mu}_h$  Cholesky decomposition of  $\mathbf{K}_h$  would still require  $\approx 1.9 \cdot 10^{15}$  floating point operations (FLOPs) to compute.

Wood and Chan (1994) proposed a solution to this problem that counter-intuitively relies on embedding the mean field within an even larger parameter space. Briefly, Wood and Chan (1994) proposed an algorithm in which a Toeplitz-family covariance matrix can be embedded within a larger circulant matrix. Then, given only the circulant matrix’s so called *base*, its complex eigenvalues can be computed efficiently via discrete Fourier transform (DFT). Without immediately going into detail, the algorithm in Wood and Chan (1994) can be used to reduce the computational requirement to evaluate and sample from the marginal prior on  $\boldsymbol{\mu}_h$  to  $< 0.01$  Gb and  $\approx 2.1 \cdot 10^9$  FLOPs. Unfortunately, because of the requirement that the original covariance matrix be Toeplitz, since  $B_h$  and  $B_s$  consist of almost entirely non-overlapping locations, this algorithm can feasibly only be used to sample  $\boldsymbol{\mu}_h$  or  $\boldsymbol{\mu}_s$  marginally, and cannot be used to sample the joint vector  $(\boldsymbol{\mu}_h^T, \boldsymbol{\mu}_s^T)^T$ . (Or rather the algorithm *could* be used, but not feasibly. Expanding to an even higher dimensional grid by considering the minimum distance between voxels would result in a circulant matrix base with over  $6.8 \cdot 10^{10}$  elements).

In (6), the covariance matrix has rank of at most  $n_h$ , and the implied conditional density  $\pi(\boldsymbol{\mu}_s | \boldsymbol{\mu}_h)$  is degenerate on  $\mathbf{K}_{s,h}\mathbf{K}_h^{-1}\boldsymbol{\mu}_h$ . By construction, this



formulation maintains the anatomical precision of  $B_h$  while reducing the size of our estimation problem enough so that we can use the Wood and Chan (1994) algorithm for posterior computation. Additionally, by (4), the product  $\mathbf{K}_h^{-1}\mathbf{K}_{h,s}$  is represented by matrix  $\mathbf{W}^\top = [w(\mathbf{v}_s)]_{\mathbf{v}_s \in B_s}$ . To induce sparsity and save computational resources, we defined  $\mathbf{W}$  in terms of neighborhoods of voxels in  $B_h$ . For any  $\mathbf{v} \in \mathcal{B}$ , let  $N_h(\mathbf{v})$  denote a set of locations in  $B_h$  in an  $r$ -neighborhood of location  $\mathbf{v}$ ,  $N_h(\mathbf{v}) = \{\mathbf{v}_h \in B_h : \|\mathbf{v}_h - \mathbf{v}\|_2 \leq r\}$ . If  $N_h(\mathbf{v})$  is empty, then we defined  $w(\mathbf{v}) = \mathbf{0}$ ; otherwise let  $\mathbf{K}_{N_h(\mathbf{v})} = [k(\mathbf{v}_i, \mathbf{v}_j)]_{\mathbf{v}_i, \mathbf{v}_j \in N_h(\mathbf{v})}$ , let  $\mathbf{k}_{N_h(\mathbf{v})} = [k(\mathbf{v}_i, \mathbf{v})]_{\mathbf{v}_i \in N_h(\mathbf{v})}$ , and let  $\tilde{\mathbf{w}} = \mathbf{K}_{N_h(\mathbf{v})}^{-1}\mathbf{k}_{N_h(\mathbf{v})}$  denote a vector with implicit dependence on  $N_h(\mathbf{v})$  where each element corresponds with one location in  $N_h(\mathbf{v})$ . For non-empty  $N_h(\mathbf{v})$ , each element of  $w(\mathbf{v})$  similarly corresponds with one location in  $B_h$ . We defined those elements to be,

$$w_i(\mathbf{v}) = \begin{cases} \tilde{w}_j & \text{if the } j^{\text{th}} \text{ location in } N_h(\mathbf{v}) \text{ corresponds to the } i^{\text{th}} \text{ location in } B_h \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

With  $\mathbf{W}^\top = [w(\mathbf{v}_s)]_{\mathbf{v}_s \in B_s}$ , the product  $\mathbf{W}\boldsymbol{\mu}_h$  can be interpreted as a local kriging approximation of  $\boldsymbol{\mu}_s$  conditional on  $\boldsymbol{\mu}_h$ . Our definition of  $\mathbf{W}$  is conceptually somewhat inspired by work on Nearest Neighbor Gaussian Processes by Datta et al. (2016); Finley et al. (2019). A sensitivity analysis over choice of  $r$  is available in the Supplement.

The matrix  $\mathbf{W}$  can be entirely precomputed given the kernel parameters,  $\psi$ ,  $\nu$ , and a neighborhood radius,  $r$ . Equipped with the matrix  $\mathbf{W}$ , samples from (6) can be drawn by first sampling  $\boldsymbol{\mu}_h \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_h)$ , and then computing  $\boldsymbol{\mu}_s = \mathbf{W}\boldsymbol{\mu}_h$ . In practice we treat  $r$  as a hyperparameter and condition analyses on it. In our data example (see section 3) we took the radius  $r$  to be roughly one FWHM length based on estimated prior covariance and hyperparameters  $\boldsymbol{\theta}$  (section 3.1). This choice was motivated by the desire to keep  $r$  roughly in line with the width of (5) while keeping  $\mathbf{W}$  only modestly expensive to compute: for this choice of  $r$ , typical neighborhood sizes  $|N_h(\mathbf{v})|$  were on the order of 300–700 voxels in patient data. We next elaborate on an efficient posterior computation algorithm for  $\boldsymbol{\mu}_h$ .

### 2.3. Posterior computation

To develop an efficient computational scheme, we embedded  $\boldsymbol{\mu}_h$  on a 3-torus by considering a larger set of locations than simply those in  $B_h$ . Our goal in doing so was to be able to replace expensive matrix operations with computations involving discrete Fourier transformations (DFTs) as we show below. We modify our original notation to treat  $\boldsymbol{\mu}_h$  as complex and write,  $\mathbf{u} = (\boldsymbol{\mu}_h^\top, \check{\mathbf{u}}^\top)^\top$ , where  $\check{\mathbf{u}}$  corresponds with non-brain locations. In our data application, the resulting  $\mathbf{u}$  has about  $8.4 \cdot 10^6$  elements corresponding with an extended grid of dimension  $(256 \times 256 \times 128)$ . Computationally, we treat this extended grid as if it were a part of  $B_h$  in the prior.

To  $\text{var}(\mathbf{u})$ , we assigned a nested block-circulant matrix  $\mathbf{C}$ , where  $\mathbf{K}_h$  is nested within  $\mathbf{C}$ , and the nested block structure arises from working on a 3D grid. As a consequence of this embedding, the dimension of  $\mathbf{u}$  is dramatically greater than that of  $\boldsymbol{\mu}_h$ , but the eigenvalues of  $\mathbf{C}$  are programmatically available via DFT without having  $\mathbf{C}$  fully in memory (e.g. Wood and Chan, 1994; Rue and Held, 2005). We express the prior  $\pi(\mathbf{u}|\boldsymbol{\theta})$  as,

$$\mathbf{u} | \boldsymbol{\theta} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}\mathbf{C}\mathbf{Q}^\top), \quad (8)$$

where  $\mathcal{CN}(\mathbf{m}, \boldsymbol{\Sigma})$  denotes the circularly symmetric complex Gaussian distribution with mean  $\mathbf{m}$  and variance  $2\boldsymbol{\Sigma}$ ; in the prior, the real,  $\mathcal{Re}(\cdot)$ , and imaginary,  $\mathcal{Im}(\cdot)$ , parts of  $\mathbf{u}$  are independent. In equation (8),  $\mathbf{Q}$  is a permutation matrix defined so that we can conveniently stack  $\mathbf{u} = (\boldsymbol{\mu}_h^\top, \check{\mathbf{u}}^\top)^\top$  in a clear order as we have here. To be explicit,

$$\mathbf{Q}\mathbf{C}\mathbf{Q}^\top = \begin{bmatrix} \mathbf{K}_h & \mathbf{C}_{h,\check{\mathbf{u}}} \\ \mathbf{C}_{\check{\mathbf{u}},h} & \mathbf{C}_{\check{\mathbf{u}},\check{\mathbf{u}}} \end{bmatrix}$$

shows the nesting of  $\mathbf{K}_h$  in  $\mathbf{C}$ . Clearly, the prior in (8) implies the same marginal distribution on  $\mathcal{Re}(\boldsymbol{\mu}_h)$  as in (6), so we can equivalently use (8) for posterior computation. From here on, we will drop pre and post multiplication by  $\mathbf{Q}$  for notational convenience as that operation can be replaced by simply rearranging the multiplicand.

Circulant matrix–vector products can be computed efficiently with DFT software. Given the first row or column  $\mathbf{a}$ —the so called base—of a circulant matrix  $\mathbf{A}$  the product  $\mathbf{A}\mathbf{x}$  can be expressed as the discrete convolution  $\mathbf{a} * \mathbf{x}$ . Equivalently, by the discrete convolution theorem,  $\text{DFT}(\mathbf{A}\mathbf{x}) = \text{DFT}(\mathbf{a}) \odot \text{DFT}(\mathbf{x})$ , where  $\odot$  denotes an elementwise or Hadamard product. The principle is the same with a nested block-circulant matrix like  $\mathbf{C}$ : the matrix has a base,  $\mathbf{c}$ , that is efficient to work with using 3D DFTs. In the present case,  $\mathbf{c}$  can be precomputed (see Algorithm 2 in the Supplement) from the original grid dimensions and covariance function  $K(\cdot, \cdot)$ . As with any circulant matrix,  $\mathbf{C}$  can be diagonalized by two Fourier matrices. If  $\mathbf{F}$  denotes a scaled 3D DFT matrix, and  $\mathbf{F}^\text{H}$  its adjugate,  $\mathbf{F}\mathbf{C}\mathbf{F}^\text{H} = \text{diag}(\boldsymbol{\lambda})$ , where  $\boldsymbol{\lambda}$  are the (complex) eigenvalues of  $\mathbf{C}$ . With only the base  $\mathbf{c}$  in memory,  $\boldsymbol{\lambda} = \mathcal{F}^{-1}(\mathbf{c})/N$  can be computed directly, where  $\mathcal{F}^{-1}(\cdot)$  computes the inverse 3D DFT and  $N$  is the number of elements in  $\mathbf{c}$ . With this setup, using the Kahan summation algorithm (Kahan, 1965) for accuracy, we can easily evaluate the quadratic form in the log prior,  $\mathbf{u}^\text{H}\mathbf{C}^{-1}\mathbf{u} = \mathcal{F}^\text{H}(\mathbf{u})(\mathcal{F}(\mathbf{u}) \oslash \boldsymbol{\lambda})$ , where  $\mathcal{F}(\cdot)$  computes the 3D DFT and  $\oslash$  denotes elementwise division. We provide a simple algorithm to construct  $\mathbf{c}$  for any dense 3D grid in the Supplement.

Posterior computation can then proceed by integrating the imaginary and non-brain parts of  $\mathbf{u}$  out of the joint posterior,

$$\begin{aligned} \pi(\mathcal{Re}(\boldsymbol{\mu}_h), \sigma_h^2, \sigma_s^2 | \mathbf{Y}_h, \mathbf{Y}_s, \boldsymbol{\theta}, r) \propto \\ \iint \pi(\mathbf{Y}_h, \mathbf{Y}_s | \mathbf{u}, \boldsymbol{\theta}, \sigma_h^2, \sigma_s^2, r) \pi(\mathbf{u} | \boldsymbol{\theta}) \pi(\sigma_h^2, \sigma_s^2) d\check{\mathbf{u}} d\mathcal{I}m(\boldsymbol{\mu}_h). \quad (9) \end{aligned}$$

Although we have conjugacy for this problem, the posterior variance of  $\mathbf{u}$  is non-circulant since the log likelihood is zero over non-brain locations in the extended grid. We sample from (9), using a blocked Hamiltonian Monte Carlo (HMC) algorithm with full conditional updates for the noise variances,  $\sigma_h^2$  and  $\sigma_s^2$ . HMC relies on several tuning parameters, including the choice of momentum distribution, mass matrix, step size, and number of numerical integration steps (Neal, 2011). While a review of HMC-flavored algorithms and tuning parameter selection is beyond the scope of this paper, we will detail our approach to tuning parameter selection for model (1). Given the other tuning parameters and a target Metropolis-Hastings rate (which we fixed at 65%), we tuned the step size  $\epsilon$  during warm up following the dual averaging method of Hoffman and Gelman (2014). We then fixed  $\epsilon_0$  at the value of  $\epsilon$  on the last burnin iteration, and drew  $\epsilon \sim \text{Uniform}(0.9 \epsilon_0, 1.1 \epsilon_0)$  to induce random integration path lengths (the product  $\epsilon L$ ) during sampling, potentially helping the algorithm escape local modes (Neal, 2011). To inform selection of the number of leapfrog integration steps  $L$ , we performed repeated analyses of patient data (Fig. 2). Results of this experiment suggest  $L = 25$  or  $L = 50$  as practical starting points for best algorithmic efficiency.

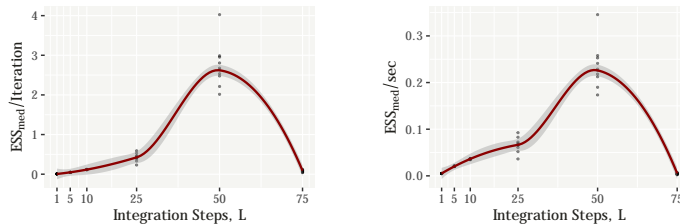


Fig 2: Algorithm efficiency (median ESS per iteration and per second) as a function of integration steps  $L$  in analysis of whole brain patient data.  $ESS$  denotes the effective sample size of elements of  $\boldsymbol{\mu}_h$ . Peak efficiency was estimated around  $L = 50$ ; analyses were replicated 10 times for each value of  $L$ , and were timed on a Thelio System76 desktop with 62 Gb of free RAM and 20 logical cores (3.3 GHz Intel<sup>®</sup> Core<sup>™</sup> i9 processors).

To sample from (9), we draw  $\sigma_h^2$  and  $\sigma_s^2$  from their full conditional distributions, and update  $\mathbf{u}$  through HMC (see Algorithm 1). Our prior places the restriction  $\sigma_h^2 > \sigma_s^2$  so that the posteriors of both nugget variance parameters are truncated inverse Gamma, but we sometimes encountered numerical difficulty sampling these parameters during warm up. As a result, we chose to ignore the restriction on  $\sigma_h^2$  and  $\sigma_s^2$  programmatically, and simply discard posterior samples where the restriction was not satisfied. After warm up, however, we found that even when working with patient data the posterior probability that  $\sigma_h^2 > \sigma_s^2$  was effectively unity, and that we never had to discard or post-process MCMC samples in this way.

To develop HMC for our problem, let  $\mathcal{L}(\mathbf{u}) = \ln \pi(\mathbf{u} | \mathbf{Y}_h, \mathbf{Y}_s, \boldsymbol{\theta}, \sigma_h^2, \sigma_s^2, r)$  rep-

resent the log posterior of  $\mathbf{u}$  conditional on  $\sigma_h^2$  and  $\sigma_s^2$ . Since  $\exp\{\mathcal{L}(\mathbf{u})\}$  is complex Gaussian, we in turn chose a complex Gaussian distribution for HMC momenta. [Girolami and Calderhead \(2011\)](#) suggest exploiting Riemannian geometry in HMC by adapting the algorithm’s mass matrix,  $\mathbf{M}$ , to the local curvature of the log posterior. The authors suggest that taking  $\mathbf{M}$  proportional to the negative Hessian of the log posterior leads to improved algorithmic efficiency in high dimensions, though this approach is not typically feasible when the dimension of  $\mathbf{M}$  is more than a few thousand. In our model, we have that,

$$-\nabla^2 \mathcal{L}(\mathbf{u}) = \begin{pmatrix} \sigma_h^{-2} \mathbf{I} + \sigma_s^{-2} \mathbf{W}^T \mathbf{W} \\ \mathbf{0} \end{pmatrix} + \mathbf{F}^H \boldsymbol{\Lambda}^{-1} \mathbf{F}, \quad (10)$$

where  $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ . In the present case, (10) is ultrahigh dimensional and impossible to work with directly, but by dropping the term involving  $\mathbf{W}^T \mathbf{W}$ , which is dense, and extending  $\sigma_h^{-2} \mathbf{I}$  we can arrive at an alternative choice of mass matrix. Let  $\mathbf{M}(\sigma_h^2)$  denote the matrix-valued function,

$$\mathbf{M}(\sigma_h^2) = \mathbf{F}^H [\boldsymbol{\Lambda}^{-1} + \sigma_h^{-2} \mathbf{I}] \mathbf{F}, \quad (11)$$

which, like  $\mathbf{C}$ , is nested block-circulant, and easy to compute with. Circulant matrices have been used successfully as preconditioners in other gradient-based optimization schemes for imaging problems (e.g. [Fessler and Booth, 1999](#)). If each element on the diagonal of  $\mathcal{R}e(\boldsymbol{\Lambda})$  is strictly greater than zero,  $\mathbf{M}(\sigma_h^2)$  is positive definite and so can be used to define a metric tensor on a Riemannian manifold as in [Girolami and Calderhead \(2011\)](#). Some additional intuition can be gained by considering how (9) relates to a missing data problem. If we were only modeling high resolution data, and those data were observed on the entire extended grid with variance  $\sigma_h^2$ , then (11) would be exactly the negative Hessian of  $\mathcal{L}(\mathbf{u})$ .

---

**Algorithm 1** Riemann manifold HMC for dual-resolution mapping models
 

---

- 1: **procedure** UPDATEMEAN( $\mathbf{u}; \boldsymbol{\lambda}, \sigma_h^2, \epsilon, L$ )
  - 2:   Compute eigenvalues of  $\mathbf{M}(\sigma_h^2)$ :
  - 3:      $\lambda_i^M \leftarrow \sigma_h^{-2} + \lambda_i^{-1}$
  - 4:     Set  $\boldsymbol{\Lambda}_M \leftarrow \text{diag}\{\lambda_i^M\}, i = 1, \dots, \dim(\mathbf{u})$
  - 5:   Sample momentum,  $\mathbf{p} \sim \mathcal{CN}(\mathbf{0}, \mathbf{F}^H \boldsymbol{\Lambda}_M \mathbf{F})$
  - 6:   Compute total energy,  $H \leftarrow -\mathcal{L}(\mathbf{u}) + \frac{1}{2} \mathbf{p}^H \mathbf{F}^H \boldsymbol{\Lambda}_M^{-1} \mathbf{F} \mathbf{p}$
  - 7:   Set  $\mathbf{u}^{\text{new}} \leftarrow \mathbf{u}$
  - 8:   **for**  $l$  in  $1, \dots, L$  **do** ▷ Leapfrog integrator
  - 9:      $\mathbf{p} \leftarrow \mathbf{p} + \frac{\epsilon}{2} \nabla \mathcal{L}(\mathbf{u}^{\text{new}})$
  - 10:      $\mathbf{u}^{\text{new}} \leftarrow \mathbf{u}^{\text{new}} + \epsilon \mathbf{F}^H \boldsymbol{\Lambda}_M^{-1} \mathbf{F} \mathbf{p}$
  - 11:      $\mathbf{p} \leftarrow \mathbf{p} + \frac{\epsilon}{2} \nabla \mathcal{L}(\mathbf{u}^{\text{new}})$
  - 12:   Compute  $H^{\text{new}} \leftarrow -\mathcal{L}(\mathbf{u}^{\text{new}}) + \frac{1}{2} \mathbf{p}^H \mathbf{F}^H \boldsymbol{\Lambda}_M^{-1} \mathbf{F} \mathbf{p}$
  - 13:   Set  $\mathbf{u} \leftarrow \mathbf{u}^{\text{new}}$  with probability  $\alpha = \min\{1, \exp(H - H^{\text{new}})\}$
  - 14:   Return posterior sample,  $\mathbf{u}$
-

#### 2.4. Functional region detection

fMRI detects functionally relevant brain regions by recording changes in oxygenated blood flow (BOLD signal). In a typical study, practitioners identify these regions by thresholding voxelwise statistical summaries in a manner that controls the false discovery rate (e.g. [Genovese, Lazar and Nichols, 2002](#)). For presurgical applications, it is at least as important to limit false negative reports, since errors of this kind may potentially lead to damage of eloquent tissue. To this end, we adapted a decision theoretic approach following previous work ([Müller, Parmigiani and Rice, 2006](#); [Liu et al., 2016, 2019](#)). We consider the loss function,

$$L(\mathbf{m}, \boldsymbol{\delta}) = \sum_i -f(m_i)\delta_i - \{1 - f(m_i)\}(1 - \delta_i) + k_1 f(m_i)(1 - \delta_i) + k_2 \{1 - f(m_i)\}\delta_i + t\delta_i, \quad (12)$$

where  $(k_1, k_2, t)$  are tunable constants, the  $m_i = |\mathcal{R}e(\mu_{h,i})|/\sqrt{\text{var } \mathcal{R}e(\mu_{h,i})}$  are posterior  $t$ -statistic analogs measuring pointwise signal strength in  $\mathcal{R}e(\boldsymbol{\mu}_h)$ , and the  $\delta_i \in \{0, 1\}$  are pointwise statistical decisions (i.e.  $\delta_i = 1$  reports a finding at voxel  $i$ , and  $\delta_i = 0$  otherwise). The function  $f(\cdot)$  can be any monotonically increasing function restricted to  $[0, 1]$ , and is intended to act as a proxy for  $\pi(\delta_i = 1 | \mathbf{Y}_h, \mathbf{Y}_s, \boldsymbol{\theta}, r)$ . Again, following previous work ([Liu et al., 2016, 2019](#)), we take  $f(m) = m/M$ , where  $M = \max_i \{m_i\}$ .

The loss function (12) is composed of five terms, each with a distinct importance:  $-\sum_i f(m_i)\delta_i$  and  $-\sum_i \{1 - f(m_i)\}(1 - \delta_i)$  induce gains for correct discoveries and correct non discoveries, respectively;  $k_1 \sum_i f(m_i)(1 - \delta_i)$  penalizes false negative errors;  $k_2 \sum_i \{1 - f(m_i)\}\delta_i$  penalizes false positive errors; and  $t \sum_i \delta_i$  penalizes the total number of discoveries. Optimal decisions  $\delta_i^*$  minimize the posterior risk and follow,

$$\delta_i^* = \mathbb{1}\{\bar{f}_i \geq (1 + k_2 + t)/(2 + k_1 + k_2)\}, \quad (13)$$

where  $\bar{f}_i$  is the posterior expectation  $\mathbb{E}\{f(m_i) | \mathbf{Y}_h, \mathbf{Y}_s, \boldsymbol{\theta}, r\}$ , and the parameters  $(k_1, k_2, t)$  suggest a threshold based on a trade off between false negative and false positive errors.

Thresholds can be tuned with domain expert guidance and/or varied dynamically, as a single static threshold may not be sufficient for a surgeon's needs (e.g. [Stippich, 2015](#)). As a practical note, in one of our patient data analyses (below), we set  $t = 1$ ,  $k_1 = 12$ , and  $k_2 = 1$  for inference. As per our coauthor and collaborating neuroradiologist's advice, this tuning parameter choice penalizes false negative errors 12 times more heavily than false positive errors. The other patient in our data was somewhat younger and less ill, with no interictal speech or language impairments. Consequentially her  $z$ -statistic images appeared to have a better signal to noise ratio. For this patient, the suggestion was to set  $k_1 = 7$  and use a seven fold penalty ratio (not shown). In both cases, the corresponding activation thresholds were confirmed visually by comparison with results from intraoperative electrocortical interference mapping.

### 3. Patient data analysis

As noted, our first motivating dataset—“patient 1”—comes from a right handed 62 year old woman who presented primarily with difficulties reading (pure alexia). This patient was found to have a large tumor in her left middle and inferior temporal gyrus; following partial surgical resection, the tumor was classified as a glioblastoma multiforme. Our second motivating dataset—“patient 2”—comes from an 18 year old right handed woman who presented after a general seizure. This patient was found to have a relatively large cavernoma adjacent to insular cortex and the transverse temporal gyrus. Both patients were scanned prior to surgery while performing a reading task with a 30 second on/off block design to map brain areas associated with reading and subsequent language processing. The task consisted of silent reading in interleaved blocks of non-final embedded clause sentences (on; eight blocks) and strings of consonants (control; eight blocks).

FMRI data collection and methods have been described previously (Liu et al., 2016). Briefly, the patients were scanned using a 3 Tesla TrioTim scanner (TQ engine, 32 channel head coil; Siemens Medical Solutions, Erlangen) using gradient-echo echo-planar imaging (GE-EPI; 3000 ms repetition time; 30 ms echo time; 0.69 ms echo spacing; GRAPPA acceleration factor 2). High resolution structural T1 weighted MPRAGE and T2 weighted FLAIR scans were also acquired to aid intraoperative neuronavigation and fMRI data preprocessing. The high and standard spatial resolution scans largely followed the same protocols, except that multi-band acceleration was used to increase the spatial resolution of high resolution acquisitions while keeping the temporal resolution the same between protocols (160 volumes were collected for each run).

FMRI time series preprocessing without spatial smoothing was performed prior to our analysis using FSL software (version 6.0.4; Jenkinson et al., 2012) and the FEAT tool (version 6.00; Woolrich et al., 2001). Standard resolution fMRI data were padded by 8 voxels in x and y (resulting in a  $72 \times 72 \times 48$  grid), and high resolution data by 10 voxels in z (resulting in  $120 \times 120 \times 72$  grid). Given standard resolution voxel sizes of  $3 \times 3 \times 3.45 \text{ mm}^3$  (patient 1) and  $3 \times 3 \times 3.3 \text{ mm}^3$  (patient 2), and high resolution voxel sizes of  $1.8 \times 1.8 \times 2.3 \text{ mm}^3$  (patient 1) and  $1.8 \times 1.8 \times 2.2 \text{ mm}^3$  (patient 2) this padding ensured that standard and high resolution data spanned the same field-of-view (FoV) within subject prior to further processing. The difference in effective resolution between the two patients resulted only from different interslice gaps (15% for patient 1 vs. 10% for patient 2; interslice gap was lowered for patient 2 because of a smaller head size). Optimal within-subject alignment of the two runs was then achieved by downsampling the volume used as the target reference for motion correction in the high resolution run and supplying this downsampled image as an alternative reference image for motion correction of the standard resolution time series. Per FSL default, we used the middle volume of the recorded frames (the 80<sup>th</sup> of our 160 volume time series) as the target reference. This volume was downsampled to the gridding of the standard resolution run using FSLeaves (part of FSL).

Data were motion corrected and temporally filtered using a 0.011 Hz high pass filter to remove low frequency drifts, and marginal linear models were fit to the time series data at each voxel to create summary statistic maps of task-related activation. In this last step, task related regressors were convolved with the canonical hemodynamic response function; temporal derivatives of resulting functions were also used as covariates of no interest. Preprocessing resulted in one unsmoothed  $z$ -statistic image for each fMRI resolution that summarized task-related activation over the course of the scans. We went on to use the generated test statistic maps as outcome data in our subsequent analysis, treating the contrast images as noisy measures of true activation.

We fit our model to the patient 1  $z$ -statistic image data to compare relative performance against a set of similar single-resolution alternative methods. With this analysis, our goal was to show how our method can be applied to identify peritumoral activations in patient data and to illustrate potential benefits to inference using combined spatial resolutions. In addition, we fit our model to the  $z$ -statistic images from patient 2 to illustrate the method’s capacity to recover an estimate of activation in regions with signal loss. Signal loss in fMRI data can occur where tissue types with different magnetic field susceptibilities neighbor one another. This is a common problem encountered in presurgical applications, and can potentially lead to exclusion of areas of interest from the analysis (e.g. [Haller and Bartsch, 2009](#); [Stippich, 2015](#)).

### 3.1. Covariance estimation

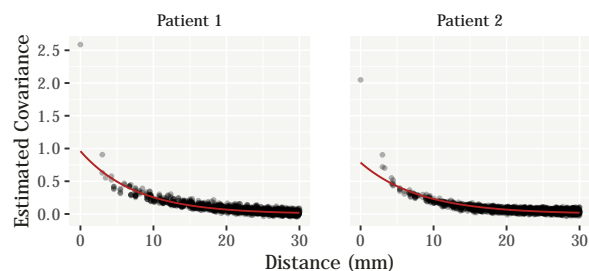


Fig 3: Covariograms show empirical covariances between neighboring standard resolution voxels as a function of distance overlaid with a parametric estimate of the covariance function in red.

We chose to estimate the Gaussian process covariance hyperparameters  $\theta$  in the spirit of empirical Bayes using the method of minimum contrast. Minimum contrast estimation (MCE) originates from [Diggle \(1981\)](#) as a moment estimation approach to spatial modeling. The method seeks to estimate parameters of a function with a known form by minimizing some discrepancy criterion given data. In our case we extracted empirical covariances between voxels at different

distances (“empirical covariogram”). We then selected  $\theta$  to minimize a nonlinear least squares objective over (5), treating the empirical covariogram as pseudo data. The Supplement gives a detailed overview of this procedure for interested readers.

This method is not without difficulty. For example, asymptotic theory suggests that empirical covariogram estimation is biased (e.g. Cressie and Glonek, 1984). Although this bias does not decrease with increased sampling density (“infill asymptotics”), it can be decreased by sampling data over increasing domains (Mardia and Marshall, 1984; Stein, 1999; Zhang, 2004). This point is worth acknowledging because the kriging identity encoded in our prior (6) makes estimation of the correlation relatively important. On the other hand, MCE is computationally efficient and scalable to large datasets, and we found that it produced reasonable estimates of the true covariance function in simulation (see the Supplement). One reason for this may be that with fMRI data we have a tremendous amount of spatial information collected on a dense grid. Although we can only estimate empirical covariances at a fixed set of distances, we typically have tens of thousands of unique pairs of voxels separated by those distances. In Fig. 3, we used the standard resolution images to estimate  $\theta$  as we expect these data to have better SNR and there is no theoretical benefit to adding infill locations as with the high resolution images. In doing so, we make an appeal to the notion of a parent process (Banerjee et al., 2008) for  $\pi\{\mu(\cdot)\}$ , which would be defined such that in the prior  $\text{cov}\{\mu(\mathbf{v}), \mu(\mathbf{v}')\} = k(\mathbf{v}, \mathbf{v}')$  for all  $\mathbf{v}, \mathbf{v}' \in \mathcal{B}$ .

For patient 1, an initial unrestricted estimate of  $\theta$  yielded an estimated kernel exponent of  $\nu \approx 1.25$ ; for improved interpretability we reran our MCE procedure fixing  $\nu = 1$  to yield  $\theta = (0.887, 0.135, 1)^\top$ . Optimization was performed using the COBYLA algorithm (Powell, 1994) as implemented by Johnson in the popular NLOpt library. The resulting covariance function is shown in red in the left panel of Fig. 3, and corresponds to a 10.47 mm full width at half maximum (FWHM) exponential correlation function. This estimate of  $\theta$  was used for all of our analyses of patient 1’s data; correspondingly, we set the neighborhood radius  $r$  to 10.35 mm in analyses of this patient’s data.

Similarly, we estimated  $\theta = (0.785, 0.132, 1)^\top$  for patient 2. The resulting covariance function (also shown in Fig. 3, right) corresponds to a 11.28 mm FWHM exponential correlation function; we set  $r$  to 11 mm for analysis of this patient’s data. In Fig. 3, the exponential kernels appear to fit the empirical covariograms quite well. The points at distances of 0 mm are not outliers but estimates of the “sill” or marginal variance of the  $\mathbf{Y}_s$  which in our model is  $\tau^2 + \sigma_s^2$ . Consequentially, our algorithm constrains  $\tau^2$  to be strictly less than the empirical variance of  $\mathbf{Y}_s$ , or whichever image is used to construct the covariogram (see the Supplement for full details).

### 3.2. Patient 1: Inference on the functional signals

We fit our model to the data from patient 1 described in section 2 with custom software written in C++ that uses the Eigen (Guennebaud et al., 2010)



and FFTW (Frigo and Johnson, 2005) libraries for linear algebra and DFT operations, respectively. For these analysis, we set the number of leapfrog steps  $L = 25$  and ran three independent HMC chains of 4,000 iterations each, discarding the first 1,000 as burnin, and thinning the output to every third iteration thereafter. Univariate Gelman–Rubin statistics (Gelman et al., 1992) were used to evaluate voxelwise convergence of  $\mathcal{R}e(\boldsymbol{\mu}_h)$ . This statistic was  $\leq 1.03$  for every voxel, suggesting approximate convergence. Additionally, trace plots of means from six randomly selected voxels are shown in in the Supplement and show good mixing of the Markov chains.

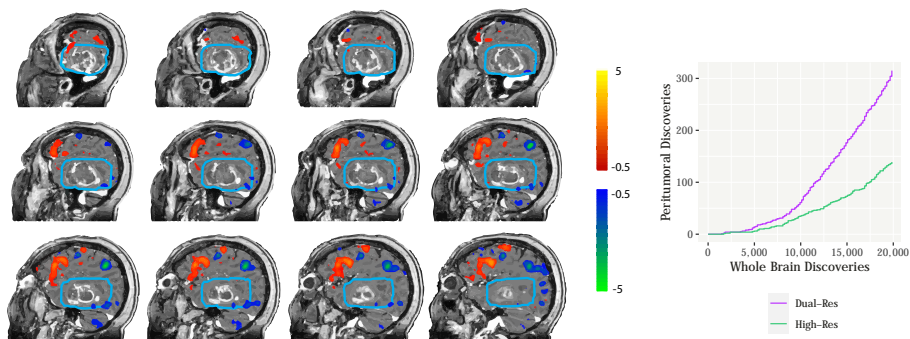


Fig 4: Patient 1: (*Left*) Thresholded posterior mean image shows peritumoral activation identified using our dual resolution method. The tumor is the region of mixed hypo- and hyperintensity in the temporal lobe across slices; the peritumoral region is outlined in cyan. Functional activations are shown in warm colors, and functional deactivations are shown in cool colors, with units on the  $z$ -statistic scale. Activation regions are shown setting  $k_1$  in our decision rule (13) to 17 to enhance the visualization. Slices are shown proceeding lateral-to-medial through the left hemisphere in left-to-right, top-to-bottom order. (*Right*) Cumulative counts of discoveries at varying decision thresholds. Voxelwise discoveries in the peritumoral region plotted against whole brain discoveries for both dual and high resolution methods.

Fig. 4 (*left*) shows posterior mean activation maps for a series of sagittal slices through the patient’s tumor in left temporal lobe. Activations are overlaid on a high resolution, gadolinium enhanced T1-weighted anatomical scan. In the figure, we have circled a peritumoral region that was deemed to determine the surgical access considered. The patient’s tumor can be seen within this circled region in all slices. We compared performance of our dual resolution model (1) against single resolution alternative methods: (i) a related Gaussian process model that only considered the high resolution data, (ii) the same model but considering only standard resolution data (kriging the posterior mean of  $\mathcal{R}e(\boldsymbol{\mu}_s)$  to the locations in  $B_h$ ), and (iii) a method that we term naive data averaging. For the alternative high and standard resolution models, we used a Gaussian process to model the the mean of the data as in (3). For the naive alternative,

we estimated the matrix  $\mathbf{W}$  (defined in section 2.2) from the data and used it to interpolate standard resolution data into the high resolution space. We then treated a simple pointwise average of the high and interpolated standard resolution images—i.e.  $\bar{\mathbf{Y}}_{h,s} = (\mathbf{Y}_h + \mathbf{W}^\top \mathbf{Y}_s)/2$ —as data in the alternative high resolution model (i). This approach is conceptually similar to previous work in this area (Liu et al., 2019). The high resolution method (i) served as our primary comparison point both because of its inherent spatial resolution and because it tended to be the best competing method in our simulations (see section 4).

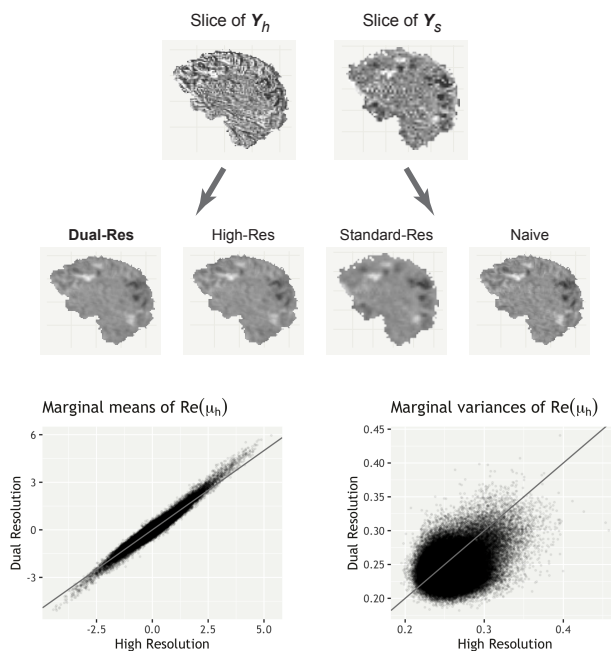


Fig 5: Patient 1: visual comparison of posterior means in a single sagittal slice from four models fit to different combinations of whole brain patient data (*middle*). The (*top*) row of the figure shows the raw data from the same slice at both high and standard resolution. Grayscale intensity is shared across all subfigures. The (*bottom*) row shows a comparison of voxelwise posterior means (*bottom, left*) and variances (*bottom, right*) of the elements of  $\mathcal{R}e(\boldsymbol{\mu}_h)$  estimated using the proposed model and a single (high) resolution alternative. The gray lines show identity relationships for comparison; variances were lower using the dual resolution model in about 72.4% of voxels.

In Fig. 4 (*right*), we show that no matter the threshold applied to whole brain posterior activation maps, our dual resolution method identified at least as many active voxels in the peritumoral region than if we had ignored the standard resolution data. A visual comparison of the posterior mean of  $\mu(\cdot)$  for all four methods is shown in Fig. 5. We chose a single sagittal slice to represent this com-

parison although the analysis was over the whole brain. Qualitatively, posterior means from the dual and high resolution analyses appear substantially sharper than for the standard resolution analysis. At the same time, differences are apparent in the dual and high resolution posterior means, particularly around the edges of areas with high magnitude signal. We also plot voxelwise comparisons of dual and high resolution posterior means and variances of  $\mathcal{R}e(\boldsymbol{\mu}_h)$  in Fig. 5. In the figure, voxels with high signal strength typically had higher magnitude posterior means estimated with the dual resolution model; marginal variances of the  $\mathcal{R}e(\mu_{h,i})$ , moreover, were lower with the dual resolution model in about 72.4% of voxels. With respect to mean image smoothness, we estimated (using our MCE procedure; see section 3.1) the standard resolution posterior mean image had a kernel FWHM of about 17.3 mm, and the high resolution posterior mean image had a kernel FWHM of about 13 mm. Appropriately, the dual resolution posterior mean image had a kernel FWHM between these two, at about 14.4 mm. Relating back to Fig. 1, our initial goal in modeling joint data sources was to reduce noise inherent in the high resolution signal and leverage signal strength from the standard resolution data. Taken all together, these results demonstrate that we have met that goal.

Additional patient 1 model fit and diagnostic evaluations are given in the Supplement. In particular, we evaluated the residual independence approximation present in our model likelihood by running our kernel estimation procedure (see section 3.1) on the model residual images. These analyses suggested that residual correlation decayed to near zero within the smallest voxel dimension widths, leading us to conclude that residual independence was a reasonable approximation in our data. Full results are available in the Supplement.

### 3.3. Patient 2: Recovery of lost signal

Similar to our analysis of patient 1, we fit our dual and single resolution models to the data from patient 2. In this case, we ran five independent HMC chains for each model and set the chain length, burnin, thinning rate, and number of leapfrog steps identically as above. Univariate Gelman–Rubin statistics were  $\leq 1.05$ , again suggesting approximate voxelwise convergence of  $\mathcal{R}e(\boldsymbol{\mu}_h)$ . Our primary goal with this analysis was to illustrate our method’s ability to recover estimates of activation from regions of signal loss.

In this particular data set, the patient’s cavernoma caused a region of GE-EPI signal dropout, with blooming along the left insular and upper temporal lobe (see Fig. 6, *left*). This is a common occurrence in clinical fMRI: brain lesions can induce signal loss and magnetic susceptibility artifacts in gradient echo imaging. The functional signal is not “missing,” per se, but local signal hypointensities can cause image preprocessing software to exclude affected areas from analysis (Haller and Bartsch, 2009). Such was the case here, and we leveraged this data structure to highlight our model’s predictive ability. By adjusting brain/background thresholding in FSL, we created two versions of this patient’s task contrast data: one where data in the core dropout region was completely

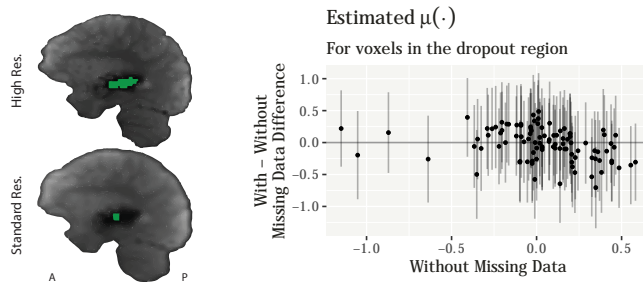


Fig 6: Patient 2: (*Left*) Core regions of fMRI signal loss across the left temporal and insular cortex are highlighted in green on high and standard resolution T2\*-weighted slices. (*Right*) Comparison of the mean parameter for voxels in the core high resolution dropout region. We fit our dual resolution model to parallel versions of the data with and without missingness. The posterior mean estimate of  $\mu(\cdot)$  without missing data is shown on  $x$ -axis, with the difference in the estimates shown on  $y$ -axis. Error bars give  $\pm$  one standard error of the difference estimated across five HMC chains.

masked out of all analyses (with missing data), and one where voxels in this same region were included in all analyses (without missing data). In this case, we were able to create parallel versions of the data for this patient, though in more general practice it may sometimes be difficult to engineer the data without missingness (Haller and Bartsch, 2009).

We fit our models to both versions of the data to compare resulting functional estimates. In the case of missing data, the Gaussian process formulation of our model enables natural prediction of values of  $\mu(\cdot)$  for voxels with missing data. In the high resolution image, the region masked out due to signal loss encompassed exactly 100 voxels, had a maximum length (measured anterior to posterior) of around 25–26 mm, and was on average about 9.4 mm wide (lateral to medial) and 4.5 mm tall (dorsal to ventral). In the low resolution image, masked dropout was limited to only two voxels given default brain/background thresholding in FSL. With typical preprocessing pipelines we would generally expect more signal loss voxels to be excluded from high resolution images. In Fig. 6 (*right*) we show, for our dual resolution method, the correspondence of predicted/estimated voxelwise means in the dropout region for the two data variants. The figure shows excellent correspondence: the Pearson correlation between the predictions and the estimates is 0.673, suggesting our method has a strong capacity for signal recovery in dropout regions of this size. Similar predictions/estimations using only the high resolution data also show good correspondence, but were made with higher variance relative to our dual resolution approach in 81 of the 100 dropout region voxels.

Similarly, across the whole brain, the marginal variances of  $\mathcal{R}e(\mu_h)$  for patient 2 were lower with our dual resolution approach in about 62.4% of voxels (compared to the high resolution model). Moreover, we estimated that dual

resolution posterior mean image smoothness had a kernel FWHM of about 7.5 mm, while the high and standard resolution posterior mean image smoothness FWHMs were about 7.6 mm and 13.5 mm, respectively. Overall, estimation and inference about  $\mathcal{R}e(\boldsymbol{\mu}_h)$  was more similar between our dual and high resolution models here than for the previous patient. Altogether, results from our analysis of the patient 2 data again suggested that our inference benefited from combining information from both spatial resolutions, though the benefit may be less pronounced compared to in patient 1. Based on our single data source high and standard resolution alternative methods, we estimated that for patient 2, the standard resolution data provided only a modest 5.4% improvement in SNR compared to the high resolution data. By contrast, for patient 1, the standard resolution data provided approximately a 139% improvement in SNR, lending context to the above result.

#### 4. Simulation studies

We quantified the advantages of our proposed method with easier to visualize simulations in two dimensions. Our goal in simulation was to evaluate how well the proposed model and alternative methods recovered activation patterns in data. Typical fMRI studies use significance testing as a means to identify functionally relevant brain regions. To mimic this setting, our simulation designs considered active regions embedded within low variance signal (see Fig. 7).

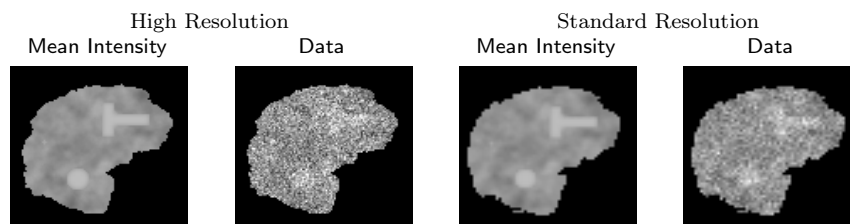


Fig 7: Simulation design example with  $\text{SNR}_h = 0.1$  and  $\text{SNR}_s = 0.2$ . Non-activation smooth signal has marginal variance 0.2 and 6 mm FWHM Exponential correlation; activation signal has mean 2.

##### 4.1. Simulations on 2D grids

Figure 7 illustrates our general approach to data simulation. In the figure, active regions were drawn in a midsagittal plane including a T-shaped region in the frontal lobe, a circular region near the cerebellum, and a four voxel square near the parietal-occipital border. These signals were created by smoothing binary images with a six millimeter full width at half maximum (FWHM) Gaussian kernel, scaling by a factor of two, and thresholding the result at 0.4. We then

embedded the active regions within random draws from a 2D random field with mean zero, marginal variance 0.2, and 6 mm FWHM Exponential or Gaussian correlation functions. We treated the resultant images as true nonzero mean intensity images, with “active” voxels given only by the smoothed T, circle, and square shapes. As per the patient data, we treated voxels in this plane as  $(1.8 \times 1.8)$  mm for high resolution (4,722 voxels total), and as  $(3 \times 3)$  mm for standard resolution (1,853 voxels). With our design, there were exactly 450 active voxels in the high resolution slice (9.5%; see Fig. 7).

We adopted this method to generate “high resolution” mean images, or  $\mathcal{R}e(\boldsymbol{\mu}_h)$  as in section 2.3, and projected  $\mathcal{R}e(\boldsymbol{\mu}_h)$  into “standard resolution” space by multiplying by  $\mathbf{W}$  as in section 2.2 to generate corresponding standard resolution mean images. In all simulation settings,  $\mathbf{W}$  was constructed using the true 6 mm FWHM Exponential or Gaussian background signal correlation functions, and an extent radius  $r$  defined as the distance after which the correlation would drop below 0.05. To simulate observed outcome data, we added independent Gaussian noise to the mean intensity images, modulated the noise variances to control SNRs of the simulated high and standard resolution images, and ran 100 replicates per parameter combination. We took the SNR to be the ratio of the second moment of the mean to the variance of the noise, and set this to be one of  $\{0.1, 0.2\}$  for high resolution images ( $\text{SNR}_h$ ). We parameterized standard resolution noise in terms of the ratio of standard to high resolution SNR ( $\text{SNR}_s:\text{SNR}_h$ ), and set this ratio to one of  $\{1, 2, 4\}$ . In the first case, the standard resolution image would not provide additional signal-to-noise support as it typically would in real data. We considered this a worst case scenario. The latter two settings were chosen so that the standard resolution image provided increasingly large signal-to-noise support, where we expected our dual resolution method to dominate. In our analysis of the patient 1 data, based on the fits of our high and standard single resolution alternative models, we estimated  $\text{SNR}_h \approx 0.18$  and  $\text{SNR}_s \approx 0.44$  based on their posterior means (ratio  $\text{SNR}_s:\text{SNR}_h \approx 2.4$ ).

#### 4.2. Recovery of simulated activation regions in 2D images

As in our patient data analysis, we compared performance of our dual resolution model against single resolution alternatives: models considering only the simulated high or standard resolution data, and an additional setting using a naive average as data. In each simulation, models were conditioned on the true  $\boldsymbol{\theta} = (\tau^2, \psi, \nu)^\top$  used to generate the low variance mean fields. We chose to condition on the true  $\boldsymbol{\theta}$  so as to explicitly focus our simulation results on estimation of and inference on the image mean intensities.

Table 1 presents selected results for estimation and inference quality in our 2D simulation settings. Results are presented predominantly for the setting with  $\text{SNR}_h = 0.1$ , the  $\text{SNR}_s:\text{SNR}_h$  ratio set to two, and an Exponential correlation function to roughly approximate our patient data (also reflected in Fig. 7). We provide a comparison point with the SNR ratio equal to one for additional interest. More extensive results are available in the Supplement. In the table,

<i>Model</i>	<i>Kernel</i>	$SNR_s:SNR_h$	$SNR_h$	<i>MSE</i>	<i>False –</i>
<b>Dual</b>	Exponential	1	0.1	<b>0.20</b>	<b>31.8% (0.4)</b>
High	Exponential	1	0.1	0.23	34.0% (0.5)
Naive	Exponential	1	0.1	0.30	43.6% (0.4)
Std	Exponential	1	0.1	0.47	43.1% (0.6)
<b>Dual</b>	Exponential	2	0.1	<b>0.18</b>	<b>30.6% (0.4)</b>
High	Exponential	2	0.1	0.23	34.0% (0.5)
Naive	Exponential	2	0.1	0.29	42.7% (0.4)
Std	Exponential	2	0.1	0.43	40.6% (0.4)

TABLE 1

*Selected results for estimation and inference quality in 2D simulations. Results for High and Std resolutions do not change across the different SNR ratios, but are repeated to facilitate comparison. Model denotes the image combination used in the analysis, and Kernel gives the correlation pattern of low variance background signal. MSE refers to mean squared error computed over the entire high resolution mean parameter vector; the simulation standard error of this metric was on the order of  $10^{-3}$  for all simulation settings and so was omitted for brevity. False – reports the mean (SE) false negative error rate when the number of discoveries was fixed at 450. One hundred replicates per parameter combination; additional results with different kernel and  $SNR_h$  parameter settings are summarized in the Supplement.*

MSE denotes the mean squared error of the estimated  $\mathcal{R}e(\mu_h)$ , computed over pixels in our simulated high resolution slices. We treat MSE as a measure of estimation quality, and report that in all simulation settings considered, MSE was lowest for the dual resolution models. This result indicates that when the same mean intensity function underlies both high and standard resolution images and the kernel function is estimated accurately, the model that used joint information from both imaging modalities outperformed possible single resolution alternatives. Interestingly, when the background intensity was generated with an Exponential kernel, as in Table 1, the model that used only high resolution data was the second best performer, underscoring the importance of spatial precision in estimation.

We also report false negative rates—the measure of inference we are most concerned with in our framework—for each model in Table 1. In the table, we set parameters  $k_1$ ,  $k_2$ , and  $t$  in our decision rule (13) independently for each model type so as to control the total number of discoveries to exactly 450 (the same as the number of pixels we considered truly active in the simulations). The actual decisions corresponding to these thresholds are shown in Fig. 8 (*right*) for a single representative simulation iteration. We emphasize that thresholds here were chosen as an objective point of comparison across the alternative methods, not by optimizing any kind of inferential criteria. In Fig. 8 (*left*), we show that the dual resolution model would give superior inference for any set of decision rule thresholds that fix the false negative rate at a single value across all methods.

## 5. Discussion

Preoperative fMRI presents many interesting and unique statistical challenges from an applied perspective. Presurgical planning requires spatially precise lo-

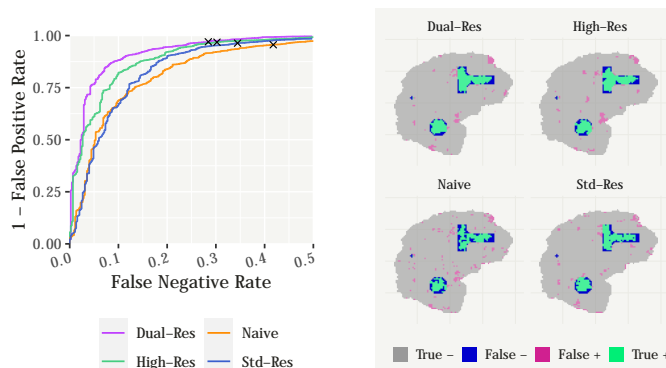


Fig 8: Inference quality in 2D simulations. (*Left*) Receiver operating characteristic (ROC) curves comparing dual and single resolution methods to a naive data averaging approach in a setting that matches the data in Fig. 7. The curves show that for almost any given false negative rate, the dual resolution method can have a uniformly lower false positive rate than alternative single resolution methods. The  $\times$ 's mark the thresholds used to generate the inferential summary on the (*right*). These thresholds limit the total number of discoveries to 450 across all four methods.

calization of patient specific functional neuroanatomy, but the current physical limitations of MR imaging technology lead to reductions in the signal to noise ratio (SNR) with increases in spatial resolution. This inherent limitation has led to the hypothesis that collecting fMRI data at multiple spatial resolutions may result in improved functional region detection; our simulations in the present paper suggest that this may indeed be the case. We have also shown how a simple decision rule can be applied by practitioners to infer about functional regions given some desired trade off between false positive and false negative errors. This is important because neuroradiologists and neurosurgeons may be more concerned with false negative errors, which could lead to resection of functionally relevant tissue in practice.

With our present work, we propose to base inferences about functional regions on a joint model for images collected at each spatial resolution. Modeling high dimensional correlated outcomes can be quite challenging computationally, and the dataset presented the additional burden of integrating two data sources with different spatial support sets. We circumvented this problem with a Gaussian parent process approximation using only the highest collected resolution image's voxel locations as a primary support set, and embedding these locations within a larger, toroidal space, leading to computational gains. As a consequence, our Gaussian process model and related algorithm has very natural extensions to cases with different numbers of data sources. For example, we recognize that not every preoperative plan will rely on collecting both high and standard resolution fMRI data. Our model can easily accommodate the situation where only one



spatial resolution is collected by simply dropping unobserved data from the joint outcome.

Just as easily, our model could accommodate data collected at additional spatial resolutions with minimal added computational cost. In fact, in a different setting, we imagine our method could be used for image based meta-analysis to synthesize results from multiple experimental studies. In such a setting, posterior credible sets—Bayesian analogs of spatial confidence sets from [Bowring et al. \(2019\)](#)—could be used to shift inferential focus back to limiting a family wise false positive error rate. With the recent proliferation of large, multicenter imaging collectives (e.g. [Van Horn and Toga, 2009](#)), we feel this may be a promising area for further applied research.

In our simulations and analysis of patient data we showed that our dual resolution method borrows strength from both data sources to improve inference, especially around the edges of active regions, without sacrificing the spatial resolution of the high resolution data (confer from Figs. 5 and 8). To accomplish this task with patient images, we started from the output of typical single subject fMRI analyses, treating summary statistics from voxelwise marginal time series models as data. [Bowman et al. \(2008\)](#) similarly used summary statistics from voxelwise marginal models as data in a group analysis in an experimental setting. Although their approach and setting was different from ours, the authors also chose not to smooth their data during preprocessing and made a similar independent noise approximation in their model likelihood as we do here ([Bowman et al., 2008](#)). We provide additional evaluation of our independent and homogeneous residual noise approximations in the Supplement, and conclude that the approximations are reasonable in our patient data. While we might eventually like to incorporate available time series information into our model, doing so would only add to computational complexity, and it is unclear to what extent spatial inference would improve as a result. At present, a handful of integrated spatiotemporal models have been developed for fMRI studies, but nearly all of these are intended to be fit to single slice data, not whole brain (e.g. [Penny, Trujillo-Barreto and Friston, 2005](#); [Groves, Chappell and Woolrich, 2009](#); [Lindquist, Loh and Yue, 2010](#)). Only more recently have variational approximations been leveraged to enable whole brain spatiotemporal inference at a reasonable computational cost ([Sidén et al., 2017](#)). In its current form, our work uses summaries of temporal data to enable whole brain inference at a very fine spatial scale, but there may be room to incorporate richer temporal information into our model as part of future study.

Finally, in our work we estimated the Gaussian process hyperparameters  $\theta = (\tau^2, \psi, \nu)^T$  from the data in the spirit of empirical Bayes. We accomplished this goal by minimizing a least squares contrast function over an empirical covariogram estimated from the data. Other approaches to learning these parameters include maximizing the data marginal likelihood (e.g. [Mardia and Marshall, 1984](#)), and fully Bayesian estimation (e.g. [Banerjee et al., 2008](#)). We chose our minimum contrast estimation (MCE) type approach as it is generally more extensible to the size of our dataset. Computing the marginal likelihood would involve inversion of an  $(n \times n)$  matrix where  $n$  is the number of voxels or spatial

locations. Our posterior computation algorithm specifically avoids even constructing such a matrix, which is impossible to store on most computer systems (see section 2.2). Fully Bayesian estimation of  $\theta$  on the other hand is possible, though still computationally demanding. The kernel bandwidth and exponent parameters,  $\psi$  and  $\nu$ , respectively, can be quite slow to update with multiple data sources, and computation time is a concern in a preoperative setting. In contrast, the partial sill variance  $\tau^2$  is straightforward to update in our framework, and an abundance of spatial data make this parameter strongly identifiable. We considered updating  $\tau^2$  by default in our algorithm, but found that it did not dramatically affect spatial inference in our data and sometimes led to slower Markov chain mixing. As a result, we decided to condition inference on fixed  $\theta$  by default in our analyses and consider alternative estimation methods a possibility for future extension.

Conditional on  $\theta$ , our method enables spatially precise inference on whole brain fMRI data collected at multiple spatial resolutions. Despite the very high dimensional nature of our data, our method is computationally efficient enough to be viable for application in presurgical planning. In addition, we have shown through simulation that inference drawn from a joint model using both available data sources can lead to substantial improvement over inference with single resolution alternatives. We hope that this body of work will benefit the presurgical fMRI community, and may find extension in experimental fields by supporting image based meta analysis and results synthesis.

### Acknowledgements

Dr. Andreas J. Bartsch has additional joint appointments with the Department of Neuroradiology at the University of Wuerzburg, Wuerzburg, Germany; with Radiologie Bamberg, Bamberg, Germany; and with the FMRI Centre Department of Clinical Neurology at the University of Oxford, Oxford, United Kingdom. This work was partially supported by NIH R01 DA048993 (Kang and Johnson).

### Supplementary Material

#### Circulant base construction

This section presents a simple algorithm to illustrate circulant matrix base computation for our applications.

---

**Algorithm 2** Compute the base of a circulant matrix associated with a 3D grid

---

```

1: procedure COMPUTECIRCULANTBASE( $\mathbf{d}$ ,  $K(\cdot, \cdot; \boldsymbol{\theta})$ )
2:   Inputs:  $\mathbf{d}$ , original 3D grid dimensions;  $K(\cdot, \cdot; \boldsymbol{\theta})$  covariance function parameterized by
    $\boldsymbol{\theta}$ 
3:   Compute extended grid dimensions,  $d_i^* \leftarrow 2^{\lceil \log_2[2(d_i-1)] \rceil}$  for  $i = 1, 2, 3$ 
4:    $k \leftarrow 0$ ,  $h \leftarrow 1$ 
5:   Find location  $\mathbf{v}_1$  associated with grid position (1, 1, 1)
6:   for  $l$  in  $1, \dots, d_3^*$  do ▷ Column-major order
7:      $j \leftarrow 0$ 
8:     if  $l \leq d_3$  then  $k \leftarrow k + 1$  else  $k \leftarrow k - 1$ 
9:     for  $m$  in  $1, \dots, d_2^*$  do
10:       $i \leftarrow 0$ 
11:      if  $m \leq d_2$  then  $j \leftarrow j + 1$  else  $j \leftarrow j - 1$ 
12:      for  $n$  in  $1, \dots, d_1^*$  do
13:        if  $n \leq d_1$  then  $i \leftarrow i + 1$  else  $i \leftarrow i - 1$ 
14:        Find location  $\mathbf{v}$  associated with grid position  $(i, j, k)$ 
15:        Compute  $c_h \leftarrow K(\mathbf{v}_1, \mathbf{v}; \boldsymbol{\theta})$ 
16:         $h \leftarrow h + 1$ 
17:   Return circulant matrix base,  $\mathbf{c}$ 

```

---

### Covariance function estimation

In this section we detail our procedure to estimate isotropic covariance functions from 3D data; in practice the method could be extended to arbitrary  $n$  dimensional data sources. The methods considered herein are not new but are included for completeness. We also report simulation results using this method to estimate the covariance from small three dimensional images and show that the method has relatively small bias in most simulation settings.

---

**Algorithm 3** Minimum contrast estimation of  $\boldsymbol{\theta}$ : high level overview

---

```

1: procedure ESTIMATEIMAGECOVARIANCE( $\mathbf{Y}$ ,  $k(\cdot; \boldsymbol{\theta})$ ,  $\Theta$ ) ▷ With the argument to  $k(\cdot; \boldsymbol{\theta})$ 
   the Euclidean distance between any two points,  $\|\mathbf{v} - \mathbf{v}'\|$ 
2:   Inputs: Image  $\mathbf{Y}$ ; covariance function  $k(\cdot; \boldsymbol{\theta})$  parameterized by  $\boldsymbol{\theta}$  with feasible region
    $\Theta$ 
3:   Construct  $\mathcal{D} \leftarrow \text{EXTRACTCOVARIANCESUMMARY}(\mathbf{Y})$  ▷ With  $\mathcal{D} = (\mathbf{d}, \hat{\mathbf{c}}, \boldsymbol{\omega})$ 
4:   Return  $\arg \min_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{\dim(\hat{\mathbf{c}})} \omega_i [\hat{c}_i - k(d_i; \boldsymbol{\theta})]^2$ 

```

---

Algorithm 3 outlines our minimum contrast estimation (MCE) procedure at a high level. The algorithm first extracts summary data  $\mathcal{D} = (\mathbf{d}, \hat{\mathbf{c}}, \boldsymbol{\omega})$  from the input data source  $\mathbf{Y}$ , where  $\hat{\mathbf{c}}$  are empirical covariances between elements of  $\mathbf{Y}$  offset by corresponding distances  $\mathbf{d}$ , and  $\boldsymbol{\omega}$  is a set of corresponding weights (defined below in algorithm 4). The algorithm then finds  $\boldsymbol{\theta}$  from within constraint region  $\Theta$  to minimize a weighted least squares contrast between the  $\hat{c}_i$  and  $k(d_i; \boldsymbol{\theta})$ .

With  $k(\cdot; \boldsymbol{\theta})$  taken to be the radial basis function as in (5), for example, the parameters  $\boldsymbol{\theta}$  correspond to the marginal variance  $\tau^2$ , correlation bandwidth  $\psi$ , and exponent  $\nu$ . For this problem, we took the feasible region  $\Theta$  to constrain  $0 < \tau^2 < \hat{c}_0$ ,  $0 < \psi$ , and  $0 < \nu \leq 2$ , where  $\hat{c}_0$  is the empirical variance of  $\mathbf{Y}$ . For

problems we consider, we found that the additional constraint  $\psi \leq \nu$  frequently helped improve estimation.

Construction of  $\mathcal{D}$  using a modified 3D raster scan is outlined in algorithm 4. In the algorithm, empirical covariances between voxels and their neighbors are computed by shifting the  $(i, j, k)$  index of each voxel by the rows of the matrix  $\mathbf{P}$  (which is constructed with the procedure outlined in Algorithm 5). The rows of  $\mathbf{P}$  define a series of perturbations in a dense 3D raster scan. In one dimension, a raster scan might only look ahead one pixel at a time so as to visit each pair of adjacent pixels only once. In two dimensions, the procedure might be defined to look ahead one pixel and look down one pixel for the same reason. In three dimensions, a simple raster might look ahead, down, and to the right by one or more voxels. We designed our procedure to sample local pairs of voxels more densely than this while still only visiting each unique pair once. Briefly, our algorithm “looks ahead” by visiting pairs of voxels within an  $(n_0 \times n_0 \times n_0)$  voxel cube such that the polar and azimuthal angles of the search are between  $[0^\circ, 180^\circ)$ . We further extended this search by adding simple raster scan perturbations out to an  $n_1$  voxel distance. In algorithm 5, we defined  $n_0 = 18$  voxels and  $n_1 = 25$  voxels by default. Our default values encompass a large number of perturbations while limiting the total computation time to a few seconds for full scale brain images.

---

**Algorithm 4** Compute empirical covariance summary data
 

---

```

1: procedure EXTRACTCOVARIANCESUMMARY( $\mathbf{Y}$ ,  $n_0$ ,  $n_1$ )
2:   Inputs: Image  $\mathbf{Y}$  with dimensions  $\mathbf{q} \in \mathbb{R}^3$ . We set  $n_0 = 18$ , and  $n_1 = 25$  by default
3:   Set  $N \leftarrow q_1 \cdot q_2 \cdot q_3$   $\triangleright N$  is the total number of voxels in image  $\mathbf{Y}$ 
4:   Store  $\mathbf{P} \leftarrow \text{IMAGESCANPERTURBATIONS}(n_0, n_1)$   $\triangleright \mathbf{P}$  is an  $(M \times 3)$  matrix of integers
5:   Allocate  $\mathbf{d} \in \mathbb{R}^M$ ,  $\hat{\mathbf{c}} \in \mathbb{R}^M$   $\triangleright \mathbf{d}$ —perturbation distances;  $\hat{\mathbf{c}}$ —empirical covariances
6:    $\mathbf{s}^{ab} \leftarrow \mathbf{0}_M$ ,  $\mathbf{s}^a \leftarrow \mathbf{0}_M$ ,  $\mathbf{s}^b \leftarrow \mathbf{0}_M$   $\triangleright$  Accumulators for sufficient statistics
7:    $\mathbf{r} \leftarrow \mathbf{0}_M$   $\triangleright$  Accumulators for counts of voxel pairs
8:   Compute sufficient statistics for pairs of voxels separated by perturbation distances:
9:   for  $h$  in  $1, \dots, N$  do  $\triangleright$  Outer loop over voxels
10:    Locate grid position  $(i, j, k)$  such that corresponds to voxel  $\mathbf{v}_h$ 
11:    if  $Y_{ijk}$  corresponds to brain data then
12:      for  $m$  in  $1, \dots, M$  do  $\triangleright$  Inner loop over perturbations
13:         $(i', j', k') \leftarrow (i, j, k) + \mathbf{P}_m^\top$ 
14:        if  $Y_{i'j'k'}$  corresponds to brain data then  $\triangleright$  Update sufficient statistics
15:           $s_m^{ab} \leftarrow s_m^{ab} + Y_{ijk} \cdot Y_{i'j'k'}$ 
16:           $s_m^a \leftarrow s_m^a + Y_{ijk}$ ;  $s_m^b \leftarrow s_m^b + Y_{i'j'k'}$ 
17:           $r_m \leftarrow r_m + 1$ 
18:   Compute distances and empirical covariances associated with grid perturbations:
19:   “Locate” voxel  $\mathbf{v}_0$  associated with grid position  $(i, j, k) = \mathbf{0}_3$ 
20:   for  $m$  in  $1, \dots, M$  do
21:     “Locate” voxel  $\mathbf{v}'$  associated with grid position  $\mathbf{P}_m$ 
22:      $d_m \leftarrow \|\mathbf{v}_0 - \mathbf{v}'\|$ 
23:     if  $r_m > 1$  then
24:        $\hat{c}_m \leftarrow (s_m^{ab} - s_m^a s_m^b / r_m) / (r_m - 1)$ 
25:   Set  $\omega_m \leftarrow \#(\mathbf{d} = d_m)$  for  $m$  in  $1, \dots, M$   $\triangleright$  Count of instances of unique elements in  $\mathbf{d}$ 
26:   Set  $\omega_m \leftarrow 1/\omega_m$  if  $\omega_m > 0$  and  $\omega_m \leftarrow 0$  otherwise for  $m$  in  $1, \dots, M$ 
27:   Return  $\mathcal{D} = (\mathbf{d}, \hat{\mathbf{c}}, \boldsymbol{\omega})$ 

```

---

---

**Algorithm 5** Construct matrix  $\mathbf{P}$  of grid index perturbations for minimum contrast estimation procedure.

---

- 1: **procedure** IMAGESCANPERTURBATIONS( $n_0, n_1$ )
  - 2:   Inputs: positive integers  $n_0, n_1, n_0 < n_1$
  - 3:   Construct principal direction matrix  $\mathbf{U} \in \mathbb{R}^{(14 \times 3)}$  such that each element  $U_{ij} \in \{-1, 0, 1\}$ ; the polar angle of each row of  $\mathbf{U}$  is between  $[0^\circ, 180^\circ)$ ; and the azimuthal angle of each row of  $\mathbf{U}$  is between  $[0^\circ, 180^\circ)$ . In our construction,  $\mathbf{U}$  includes a row of all 0's
  - 4:   Construct  $\mathbf{Q} \in \mathbb{R}^{(n_0^3 \times 3)}$  with rows consisting of all 3-element permutations of elements of  $(1, \dots, n_0)$
  - 5:   Compute  $\mathbf{P} \leftarrow \mathbf{Q} * \mathbf{U}$ , where  $*$  denotes the column-wise Khatri-Rao product
  - 6:   **for**  $k$  in  $n_0 + 1, \dots, n_1$  **do**
  - 7:      $\mathbf{P} \leftarrow [\mathbf{P}^\top \ k \ \mathbf{I}_3]^\top$
  - 8:   Remove duplicate rows from  $\mathbf{P}$
  - 9:   Return  $\mathbf{P}$
- 

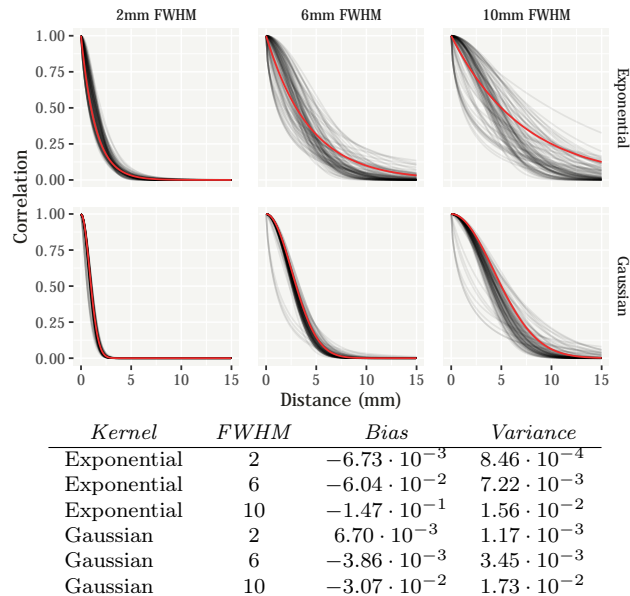


Fig 9: Recovery of the correlation function in small 3D images. Each gray line shows a correlation function estimated in repeated simulation (true correlation functions for each panel shown in red). In the table, *Bias* and *Variance* were computed pointwise and averaged over a dense grid from  $[0, 15]$  (mm).

Fig. 9 presents the results of a simulation assessing the performance of our MCE procedure. We simulated small 3D images on a  $(32 \times 32 \times 16)$  grid, treating voxels as isotropic  $1 \text{ mm}^3$ . In our simulation, we drew mean images from Gaussian processes with either Exponential or Gaussian correlation functions; unit marginal variance; and either two, six, or ten mm full widths at half maxima. Mean images were corrupted with independent Gaussian noise with the signal

to noise ratio set to 0.2 roughly to match our observed patient data and our 2D simulations in section 4.

Since the Gaussian predictive process basis described in section 2.2 relies only on the correlation bandwidth and exponent parameters  $\psi$  and  $\nu$ , the most important measure of estimation success in our setting is recovery of the correlation function, not necessarily estimation of  $\theta$  itself. For any given dataset, the nonlinear least squares objective in algorithm 3 might be multimodal in  $\theta$ , but this is relatively unimportant if the resulting correlation functions at different modes behave similarly.

In Fig. 9, the red line in each panel shows the true correlation function used to generate underlying mean images in simulation, and the corresponding gray lines show estimated correlation functions from 100 repeated simulations. The table below the panels summarizes pointwise bias and variance averaged over a grid of 1,000 equally spaced points from  $[0, 15]$  (mm). In the worst case scenario (10 mm FWHM Exponential correlation function), pointwise mean squared error was on average only about  $3.8 \cdot 10^{-2}$ , and was between  $[1.1, 7.5] \cdot 10^{-2}$  for 95% of points on the grid. Even with relatively small 3D images, these results suggest our MCE procedure can recover the true correlation functions reasonably well.

#### Analysis of patient 1 data: Model diagnostics

In this section we include several of the general attempts we have made to probe Markov chain convergence and model fit in our analysis of patient data.

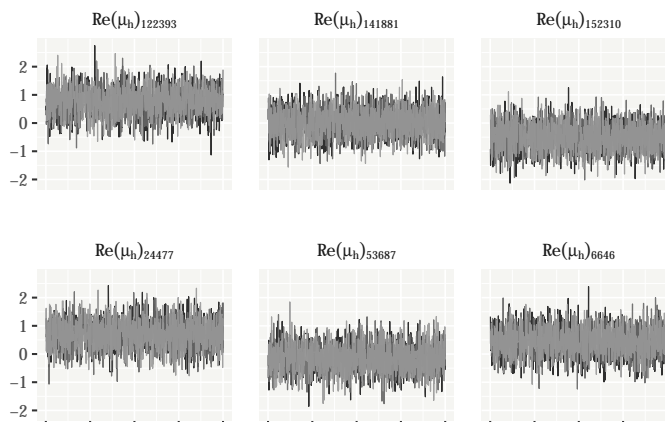


Fig 10: Trace plots for the mean parameter of six random voxels from analysis of patient 1's data with our dual resolution model. Three different HMC chains are overlaid on one another in each subfigure.

As discussed in the main text, Fig. 10 shows trace plots for three chains of Hamiltonian Monte Carlo (HMC) draws of the mean parameter for six random voxels. In all cases we examined, chains appear to show good convergence and mixing. Fig. 11 (*left*) shows empirical covariograms and estimated covariance

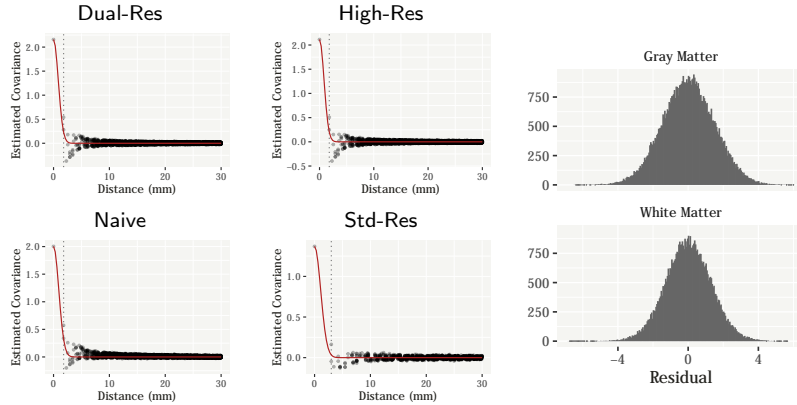


Fig 11: (*Left*) Residual covariograms for each method. The dotted lines show minimum voxel dimensions for each resolution, suggesting that the residual independence approximation is reasonable in these data. (*Right*) Dual resolution method residual histograms roughly separated by gross tissue type. Residuals have modestly higher dispersion in gray matter than in white.

functions for residual images from each method. We found that the estimated residual correlation functions' full widths at half maxima were on the order of the minimum voxel dimensions in all cases. These analyses suggested that residual correlation decayed to near zero within the smallest voxel dimension widths, leading us to conclude that residual independence was a reasonable approximation in our data.

The right panel of Fig. 11 shows histograms of the residuals from our dual resolution method roughly separated by gross tissue type. We chose to parse the residuals in this way due to some concern that a homogeneous residual variance approximation may not be fully justifiable across the whole brain. To construct this figure, we created non-overlapping gray and white matter tissue labels using the FAST program from the FSL software suite (Zhang, Brady and Smith, 2001), though the presence of the tumor complicates this procedure. The figure suggests that residuals had modestly higher dispersion in gray matter (standard deviation = 1.49) than in white (standard deviation = 1.31). If it were not for the tumor, we might ideally only want to analyze gray matter voxels for signs of task-related activation. Given the present context, however, this strategy is not completely possible. As it stands, although it appears homogeneous residual variance may not strictly hold across different tissue types, we do not believe the approximation is so poor as to grossly impact our analyses in a negative way.

We further examined posterior predictive distributions for the data from each voxel in the high resolution image, and compared the distributions against the observed data (analysis not shown). Dual resolution model posterior predictive inverse quantiles for the observed data were roughly uniform, suggesting that

data outliers occurred no more or less frequently than would be expected given the model.

### Analysis of patient 2 data: Sensitivity analysis

In this section we include a brief sensitivity analysis related to the choice of neighborhood size in our dual resolution method.

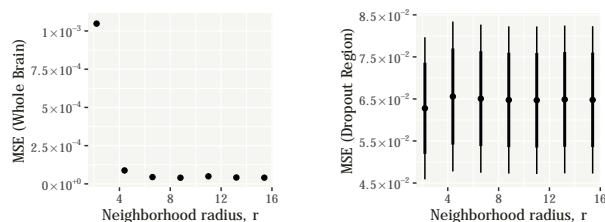


Fig 12: Mean squared error (MSE) of the posterior expectation of  $\mu(\cdot)$  given fixed  $\theta$  but different values of  $r$ . The (*left*) panel shows MSE of  $\mu(\cdot)$  evaluated across the whole brain, while the (*right*) panel shows the predictive MSE for voxels in patient 2’s dropout region. Thick and thin lines give approximate 80% and 95% confidence intervals.

Our dual resolution mapping method relies on a neighborhood radius parameter  $r$  to construct locally kriged samples of  $\mu_s$  given  $\mu_h$  (see section 2.2 in the main text). Conceptually, this construction is somewhat inspired by the so called nearest neighbor Gaussian process (Datta et al., 2016; Finley et al., 2019). In practice, we treat  $r$  as a hyperparameter and condition analyses on it, though it is of interest to understand how the choice of  $r$  affects inference about  $\mu_h$ . In our patient data analyses (sections 3.2 and 3.3 in the main text), we set  $r$  to be approximately equal to the estimated full kernel width at half maximum. This resulted in neighborhood sizes on the order of 300–700 voxels for our patient data.

To explore the influence of  $r$  on estimation and prediction, we fit our dual resolution model to the patient 2 data under several different settings, all for fixed  $\theta$ . As a comparison point, we took the posterior mean of  $\mu(\cdot)$  fit to the data without missingness and conditioned on  $r = 11$  mm. We then compared against the posterior mean of  $\mu(\cdot)$  from repeat analyses of the with-missingness data and varying values of  $r$  (see section 3.3 in the main text for an explanation of the two data sets). For these repeat analyses, we chose values of  $r$  based on multiples of the largest high resolution image voxel dimension (2.2 mm). Fig. 12 summarizes this experiment in terms of the squared error of  $\mu(\cdot)$  averaged over the whole brain (*left*) and voxels in the dropout region (*right*). From these results we conclude that as long as  $r$  is sufficiently large ( $\geq 6.6$  mm or so; corresponding to neighborhood sizes of at least 100–200 voxels), it does not appear to have much influence on posterior estimates.

### 2D simulation results

In this section we give results designed augment those reported in section 4.2



with additional simulation settings. Tables follow the exact format of Table 1 in the main text.

In all cases considered, our dual resolution method had the lowest mean squared error (MSE) and false negative rate. Of potential interest, however, is that the high resolution-only method was the second best performer when data were simulated with a marginal exponential correlation structure (Table 2), and the naive data averaging method was the second best performer when data were simulated with a marginal Gaussian correlation structure (Table 3).

<i>Model</i>	<i>Kernel</i>	$SNR_s:SNR_h$	$SNR_h$	<i>MSE</i>	<i>False –</i>
<b>Dual</b>	Exponential	1	0.1	<b>0.20</b>	<b>31.8% (0.4)</b>
High	Exponential	1	0.1	0.23	34.0% (0.5)
Naive	Exponential	1	0.1	0.30	43.6% (0.4)
Std	Exponential	1	0.1	0.47	43.1% (0.6)
<b>Dual</b>	Exponential	1	0.2	<b>0.17</b>	<b>29.3% (0.4)</b>
High	Exponential	1	0.2	0.20	31.0% (0.4)
Naive	Exponential	1	0.2	0.29	43.0% (0.3)
Std	Exponential	1	0.2	0.43	40.6% (0.4)
<b>Dual</b>	Exponential	2	0.1	<b>0.18</b>	<b>30.6% (0.4)</b>
High	Exponential	2	0.1	0.23	34.0% (0.5)
Naive	Exponential	2	0.1	0.29	42.7% (0.4)
Std	Exponential	2	0.1	0.43	40.6% (0.4)
<b>Dual</b>	Exponential	2	0.2	<b>0.15</b>	<b>28.5% (0.3)</b>
High	Exponential	2	0.2	0.20	31.0% (0.4)
Naive	Exponential	2	0.2	0.29	42.4% (0.3)
Std	Exponential	2	0.2	0.41	40.5% (0.3)
<b>Dual</b>	Exponential	4	0.1	<b>0.16</b>	<b>29.5% (0.3)</b>
High	Exponential	4	0.1	0.23	34.0% (0.5)
Naive	Exponential	4	0.1	0.29	42.3% (0.3)
Std	Exponential	4	0.1	0.41	40.5% (0.3)
<b>Dual</b>	Exponential	4	0.2	<b>0.14</b>	<b>27.9% (0.3)</b>
High	Exponential	4	0.2	0.20	31.0% (0.4)
Naive	Exponential	4	0.2	0.28	42.1% (0.3)
Std	Exponential	4	0.2	0.40	40.8% (0.3)

TABLE 2

*Results for estimation and inference quality in 2D simulations when background signal has an Exponential correlation structure. As in Table 1, results for High and Std resolutions do not change across the different SNR ratios, but are repeated to facilitate comparison. MSE refers to mean squared error computed over the entire high resolution mean parameter vector. False – reports the mean (SE) false negative error rate when the number of discoveries was fixed at 450. One hundred replicates per parameter combination.*

Model	Kernel	$SNR_s:SNR_h$	$SNR_h$	MSE	False –
<b>Dual</b>	Gaussian	1	0.1	<b>0.24</b>	<b>29.8% (0.3)</b>
High	Gaussian	1	0.1	0.28	34.1% (0.3)
Naive	Gaussian	1	0.1	0.25	34.5% (0.3)
Std	Gaussian	1	0.1	0.59	50.0% (0.4)
<b>Dual</b>	Gaussian	1	0.2	<b>0.17</b>	<b>24.2% (0.2)</b>
High	Gaussian	1	0.2	0.21	27.1% (0.2)
Naive	Gaussian	1	0.2	0.19	27.1% (0.2)
Std	Gaussian	1	0.2	0.58	43.5% (0.3)
<b>Dual</b>	Gaussian	2	0.1	<b>0.21</b>	<b>27.5% (0.2)</b>
High	Gaussian	2	0.1	0.28	34.1% (0.3)
Naive	Gaussian	2	0.1	0.24	33.3% (0.3)
Std	Gaussian	2	0.1	0.58	43.5% (0.3)
<b>Dual</b>	Gaussian	2	0.2	<b>0.15</b>	<b>22.8% (0.2)</b>
High	Gaussian	2	0.2	0.21	27.1% (0.2)
Naive	Gaussian	2	0.2	0.18	26.4% (0.2)
Std	Gaussian	2	0.2	0.57	38.6% (0.2)
<b>Dual</b>	Gaussian	4	0.1	<b>0.18</b>	<b>25.0% (0.2)</b>
High	Gaussian	4	0.1	0.28	34.1% (0.3)
Naive	Gaussian	4	0.1	0.24	33.0% (0.3)
Std	Gaussian	4	0.1	0.57	38.6% (0.2)
<b>Dual</b>	Gaussian	4	0.2	<b>0.12</b>	<b>21.1% (0.2)</b>
High	Gaussian	4	0.2	0.21	27.1% (0.2)
Naive	Gaussian	4	0.2	0.18	25.9% (0.2)
Std	Gaussian	4	0.2	0.54	34.8% (0.2)

TABLE 3

Results for estimation and inference quality in 2D simulations when background signal has a Gaussian correlation structure. As in Tables 1 and 2, results for High and Std resolutions do not change across the different SNR ratios, but are repeated to facilitate comparison. MSE refers to mean squared error computed over the entire high resolution mean parameter vector. False – reports the mean (SE) false negative error rate when the number of discoveries was fixed at 450. One hundred replicates per parameter combination.

## References

- ARCHIP, N., CLATZ, O., WHALEN, S., KACHER, D., FEDOROV, A., KOT, A., CHRISOCHOIDES, N., JOLESZ, F., GOLBY, A., BLACK, P. M. et al. (2007). Non-rigid alignment of pre-operative MRI, fMRI, and DT-MRI with intra-operative MRI for enhanced visualization and navigation in image-guided neurosurgery. *NeuroImage* **35** 609–624.
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 825–848.
- BERROCAL, V. J., GELFAND, A. E. and HOLLAND, D. M. (2012). Space-time data fusion under error in computer model output: An application to modeling air quality. *Biometrics* **68** 837–848.
- BERTALANFFY, H., BENES, L., MIYAZAWA, T., ALBERTI, O., SIEGEL, A. M. and SURE, U. (2002). Cerebral cavernomas in the adult. Review of the literature and analysis of 72 surgically treated patients. *Neurosurgical review* **25** 1–53.
- BODURKA, J., YE, F., PETRIDOU, N., MURPHY, K. and BANDETTINI, P. A.

- (2007). Mapping the MRI voxel volume in which thermal noise matches physiological noise—implications for fMRI. *Neuroimage* **34** 542–549.
- BOWMAN, F. D., CAFFO, B., BASSETT, S. S. and KILTS, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* **39** 146–156.
- BOWRING, A., TELSCHOW, F., SCHWARTZMAN, A. and NICHOLS, T. E. (2019). Spatial confidence sets for raw effect size images. *NeuroImage* **203** 116187.
- CORDELLA, R., ACERBI, F., BROGGI, M., VAILATI, D., NAZZI, V., SCHIARITI, M., TRINGALI, G., FERROLI, P., FRANZINI, A. and BROGGI, G. (2013). Intraoperative neurophysiological monitoring of the cortico-spinal tract in image-guided mini-invasive neurosurgery. *Clinical Neurophysiology* **124** 1244–1254.
- COX, R., ASHBURNER, J., BREMAN, H., FISSELL, K., HASELGROVE, C., HOLMES, C., LANCASTER, J., REX, D., SMITH, S., WOODWARD, J. et al. (2004). A (Sort of) new image data format standard: NIfTI-1: WE 150. *NeuroImage* **22**.
- CRESSIE, N. and GLONEK, G. (1984). Median based covariogram estimators reduce bias. *Statistics & Probability Letters* **2** 299–304.
- 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.
- DIGGLE, P. J. (1981). Binary mosaics and the spatial pattern of heather. *Biometrics* 531–539.
- DURNEZ, J., MOERKERKE, B., BARTSCH, A. and NICHOLS, T. E. (2013). Alternative-based thresholding with application to presurgical fMRI. *Cognitive, Affective, & Behavioral Neuroscience* **13** 703–713.
- FESSLER, J. A. and BOOTH, S. D. (1999). Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Transactions on Image Processing* **8** 688–699.
- FINLEY, A. O., DATTA, A., COOK, B. D., MORTON, D. C., ANDERSEN, H. E. and BANERJEE, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *Journal of Computational and Graphical Statistics* **28** 401–414.
- FRIGO, M. and JOHNSON, S. G. (2005). The Design and Implementation of FFTW3. *Proceedings of the IEEE* **93** 216–231. Special issue on “Program Generation, Optimization, and Platform Adaptation”.
- FUENTES, M. and RAFTERY, A. E. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics* **61** 36–45.
- GELFAND, A. E., ZHU, L. and CARLIN, B. P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics* **2** 31–45.
- GELMAN, A., RUBIN, D. B. et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7** 457–472.
- GENOVESE, C. R., LAZAR, N. A. and NICHOLS, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* **15** 870–878.

- GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73** 123–214.
- GROVES, A. R., CHAPPELL, M. A. and WOOLRICH, M. W. (2009). Combined spatial and non-spatial prior for inference on MRI time-series. *NeuroImage* **45** 795–809.
- GUENNEBAUD, G., JACOB, B. et al. (2010). Eigen v3. <http://eigen.tuxfamily.org>.
- HALLER, S. and BARTSCH, A. J. (2009). Pitfalls in fMRI. *European radiology* **19** 2689–2706.
- HOFFMAN, M. D. and GELMAN, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15** 1593–1623.
- JENKINSON, M., BECKMANN, C. F., BEHRENS, T. E., WOOLRICH, M. W. and SMITH, S. M. (2012). Fsl. *NeuroImage* **62** 782–790.
- JOHNSON, S. G. The NLOpt nonlinear optimization package <http://github.com/stevengj/nlopt>.
- JOVČEVSKA, I., KOČEVAR, N. and KOMEL, R. (2013). Glioma and glioblastoma-how much do we (not) know? *Molecular and clinical oncology* **1** 935–941.
- KAHAN, W. (1965). Pracniques: further remarks on reducing truncation errors. *Communications of the ACM* **8** 40.
- LARGE, I., BRIDGE, H., AHMED, B., CLARE, S., KOLASINSKI, J., LAM, W., MILLER, K., DYRBY, T. B., PARKER, A., SMITH, J. et al. (2016). Individual differences in the alignment of structural and functional markers of the V5/MT complex in primates. *Cerebral Cortex* **26** 3928–3944.
- LINDQUIST, M. A., LOH, J. M. and YUE, Y. R. (2010). Adaptive spatial smoothing of fMRI images. *Statistics and its Interface* **3** 3–13.
- LIU, Z., BERROCAL, V. J., BARTSCH, A. J. and JOHNSON, T. D. (2016). Pre-surgical fMRI data analysis using a spatially adaptive conditionally autoregressive model. *Bayesian Analysis (Online)* **11** 599.
- LIU, Z., BARTSCH, A. J., BERROCAL, V. J. and JOHNSON, T. D. (2019). A mixed-effects, spatially varying coefficients model with application to multi-resolution functional magnetic resonance imaging data. *Statistical Methods in Medical Research* **28** 1203–1215.
- MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.
- MÜLLER, P., PARMIGIANI, G. and RICE, K. (2006). FDR and Bayesian multiple comparisons rules.
- NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* **2** 2.
- NIMSKY, C., GANSLANDT, O., BUCHFELDER, M. and FAHLBUSCH, R. (2006). Intraoperative visualization for resection of gliomas: the role of functional neuronavigation and intraoperative 1.5 T MRI. *Neurological Research* **28** 482–487.

- PENNY, W. D., TRUJILLO-BARRETO, N. J. and FRISTON, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage* **24** 350–362.
- POWELL, M. J. (1994). A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis* 51–67. Springer.
- RASMUSSEN, C. E. and WILLIAMS, C. K. (2006). *Gaussian Processes for Machine Learning* **2**. MIT press Cambridge, MA.
- REUTER, M., ROSAS, H. D. and FISCHL, B. (2010). Highly accurate inverse consistent registration: a robust approach. *NeuroImage* **53** 1181–1196.
- RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.
- SEEGER, M., WILLIAMS, C. and LAWRENCE, N. (2003). Fast forward selection to speed up sparse Gaussian process regression Technical Report.
- SIDÉN, P., EKLUND, A., BOLIN, D. and VILLANI, M. (2017). Fast Bayesian whole-brain fMRI analysis with spatial 3D priors. *NeuroImage* **146** 211–225.
- SILVA, M. A., SEE, A. P., ESSAYED, W. I., GOLBY, A. J. and TIE, Y. (2018). Challenges and techniques for presurgical brain mapping with functional MRI. *NeuroImage: Clinical* **17** 794–803.
- STEIN, M. (1999). Interpolation of spatial data. Springer series in statistics.
- STEIN, M. L. et al. (2007). Spatial variation of total column ozone on a global scale. *The Annals of Applied Statistics* **1** 191–210.
- STIPPICH, C. (2015). *Clinical Functional MRI: Presurgical Functional Neuroimaging*. Springer.
- VAN HORN, J. D. and TOGA, A. W. (2009). Multi-site neuroimaging trials. *Current Opinion in Neurology* **22** 370.
- WOOD, A. T. and CHAN, G. (1994). Simulation of stationary Gaussian processes in  $[0, 1]^d$ . *Journal of Computational and Graphical Statistics* **3** 409–432.
- WOOLRICH, M. W., RIPLEY, B. D., BRADY, M. and SMITH, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage* **14** 1370–1386.
- ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99** 250–261.
- ZHANG, Y., BRADY, M. and SMITH, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging* **20** 45–57.