2023

# ON BAYESIAN METHODS AND FUNCTIONAL REGISTRATION OF FMRI

XiaoXuan Wang
wang8f1@uwindsor.ca

# ON BAYESIAN METHODS AND FUNCTIONAL REGISTRATION OF FMRI

XiaoXuan Wang

# ON BAYESIAN METHODS AND FUNCTIONAL REGISTRATION OF FMRI

by

Xiaoxuan Wang

A Major Research Paper
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
In Partial Fulfillment of the Requirements
for the Degree of Master of Science
at the University of Windsor

Windsor, Ontario, Canada

2023

# ON BAYESIAN METHODS AND FUNCTIONAL REGISTRATION OF FMRI

by

Xiaoxuan Wang

APPROVED BY:

---

M.Belalia
Department of Mathematics and Statistics

---

S.Nkurunziza, Advisor
Department of Mathematics and Statistics

May 8, 2023

# Author's Declaration of Originality

I hereby certify that I am the sole author of this major paper and that no part of this major paper has been published or submitted for publication.

I certify that, to the best of my knowledge, my major paper does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my major paper, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my major paper and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my major paper, including any final revisions, as approved by my major paper committee and the Graduate Studies office, and that this major paper has not been submitted for a higher degree to any other University or Institution.

# Abstract

The application of functional magnetic resonance imaging (fMRI) has greatly improved our comprehension of the human brain and behaviour. However, after anatomical alignment, there remains large inter-individual variability in brain anatomy and functional localization, which is one of the obstacles to conducting group studies and performing group-level inference. This major paper addresses this problem by applying a new method (Bayesian Functional Registration) to decrease misalignment in functional brain systems between people by spatially transforming each subject's functional data into a common reference map. The proposed approach allows us to assess differences in brain function across subjects. It also creates a framework that integrates feature- and intensity-based data and enables inference of the transformation parameters using posterior samples. Next, we evaluate the method using the data from a study of the correspondence of categorical and feature-based representations of music in the human brain. Finally, the proposed approach shows an increased sensitivity for group-level inference compared with the standard method, which uses the registration estimation toolbox in Matlab.

# Acknowledgments

This major paper is the culmination of a long and arduous journey-one that I have only been able to make due to the dedicated support I have received from so many people along the way.

The deepest and sincerest gratitude goes to my supervisor Dr. Sévérien Nkurunziza for his continuous and invaluable guidance throughout my research. It is a great honor and privilege that I was given the opportunity to work under his supervision. I would like to thank him for his patience, support, empathy, and great sense of humor. In addition, I am extremely grateful to my parents for their emotional and financial support during my education and life journey.

My final thanks go to those who have directly or indirectly supported me in completing my major paper. Sincere wishes to you, and may everything be pleasant and pleasant to you.

# Contents

# List of Figures

# Chapter 1

# Introduction

Magnetic resonance imaging (MRI) or functional magnetic resonance imaging (fMRI) is a kind of technique that measures brain activity by detecting changes associated with blood flow (Huettel et al., 2009; Rinck, 2019)[1][2]. The cerebral blood flow is associated with neuronal activation in this technique. This is because when an area of the brain is used, the blood flow to that area also increases. Recently, fMRI has contributed a lot to our understanding of the neurophysiological basis of human behavioural patterns. The major factor behind these advancements has been the development of more sophisticated statistical and computational methods, as well as the ability to record brain data with finer spatial and temporal resolutions. As a result of these developments, researchers can now go deeper into the relationship between the brain and behaviour. At the same time, these advancements help us to create models of functional brain representations which relate to behaviour, performance, clinical status, and prognosis at the population level (Wang et al., 2021)[3]. However, in these

contexts, the differences in functional brain anatomy constitute a major obstacle to efficiently applying emerging analytic techniques. Specifically, the activation patterns of different subjects have different locations. As a result of overcoming this barrier, brain models will be created at the population level more efficiently. The capability of performing hypothesis tests at the group level will be extended, which can be used to predict behaviour better and utilize fMRI's full spatial resolution.

An fMRI study would generally require each participant to receive one or more stimuli while being monitored at hundreds of time points. In order to create multivariate time series data, approximately 100,000 spatial positions (voxels) at each time point are recorded to calculate the subject's blood oxygenation level-dependent response (BOLD) (Wang et al., 2015)[3]. To accurately perform statistical analysis over all participants, every voxel must be located in the same brain structure. However, since each brain is unique in terms of size and structure, it does not occur naturally. In order to achieve this purpose, as in Wang et al. (2021)[3], every subject's brain is normalized to a stereotaxic space prior to all analyses. A high-resolution anatomical scan that is spatially aligned with the fMRI data is typically employed in this process. Standard practices involve nonlinear transformations to warp anatomical scans of individual participants into anatomically-based reference spaces, such as the "Montreal Neurologic Institute" (MNI) space. Once the data have been transformed, they are placed in MNI space, enabling them to be compared across subjects. (Lindquist et al., 2008; Ombao et al., 2016)[4][5].

However, this method does not account for residual differences in brain anatomy

or the distribution and location of functional regions around anatomical landmarks in all individuals. It has been reported that the primary visual cortex can differ between individuals in size and position compared to other anatomical landmarks by up to twofold (Rademacher et al., 1993; Amunts et al., 2000)[6][7], and the same is true for sulcal sites (Thompson et al.,1996)[8]. However, sometimes normalization cannot address inter-individual variations. As introduced in Wang et al. (2021)[3], the cingulate sulcus has structural dimorphism (Vogt et al., 1995)[9]; there exists a considerable fraction of individuals with a double cingulate sulcus, resulting in an inability to align the anatomy with the current methods. Further, functional localization still has a significant difference even after anatomical alignment (Wang et al., 2021)[3]. For example, the location of the motion-sensitive area MT of the visual system may vary by more than $2cm$ (Duncan et al.,2009)[10]. It has been shown that both the size of the fusiform face region, as well as the position and size of the lateral occipital cortex, varies greatly between individuals and are located in a variety of anatomical sulcal landmarks (Iordan et al., 2016; Allison et al., 1999; McCarthy et al., 1999)[11][12][13]. As a result of these facts, the normalization methods used in the regular process are ineffective.

Instead of anatomically based brain spaces, recent methods have mapped each brain into a functional population-level reference space. The “hyper alignment” approach is the first such model (Haxby et al., 2011)[14]. In this context, the neural representational space spanning the voxels in a local neighbourhood is used to represent brain activity patterns in response to stimuli and other cognitive events.

Hyper alignment aligns the representational geometry between subjects by rotating each participant's local voxel-wise activity patterns using a Procrustes transformation. This methodology is similar in mathematical structure to Canonical Correlation Analysis (CCA). As a result of the development of time-series data analysis techniques, other research methods have been developed to align subjects while watching movies. It has been demonstrated that functional time series alignment (Sabuncu et al., 2010)[15] aligns voxels among individuals using a 2D "rubber sheet" warping method, which maximizes correlations across subjects at the individual level (Hasson et al., 2004)[16]. Another technique is functional connectivity alignment (Conroy et al., 2013)[17], which minimizes the Frobenius norm of the difference between a subject's connectivity matrix and a reference matrix using a shape-preserving penalty function.

However, while this body of research has great potential for changing functional activation patterns, it also has a number of disadvantages (Haxby et al., 2011)[14]. First, functional connectivity-based approaches require a substantial amount of resting state data. Additionally, hyper-alignment requires patients to view a substantial movie (up to two hours) in order to be aligned. As a second point, the type of functional patterns that can be appropriately aligned may be determined by the selection of the reference data. For example, audio-visual representations are well aligned with movie reference data as opposed to prefrontal and limbic networks (Wang et al., 2021)[3]. In addition, functional connectivity measures have been shown to be more effective when applied to the limbic cortex but less effective when applied to

the object and semantic representations. It should be noted that these approaches have not been used to investigate important clinical functions like pain or emotion. Third, some recent studies in this field have also made some progress but have not incorporated an explicit spatial model (Nenning et al., 2017)[18].

Following the work of Wang et al. (2021)[3], we focus on the local registration of brain activation according to subject-specific functional activation maps. These maps represent the results of several voxel-wise models that measure the activation in the BOLD signal that generates from the stimulus of interest. They are spatial maps corresponding to the regression coefficient corresponding to the predictor modelled by the stimulus (for more detail, see Chapter 3). Because of the significant spatial correlation, these coefficients estimated from a normal-error linear model can be well described using a Gaussian process (Wang et al., 2021)[3]. All subject-specific activation maps should be in the same anatomical space in order to ensure that voxel-specific values are directly comparable across subjects. As previously stated, there is still a substantial amount of individual variation in functional brain anatomy even after anatomical registration. Hence, the data need to be registered in a common functional space by a secondary functional registration (Wang et al., 2021)[3]. In order to perform secondary-level registration, the floating maps will be warped (or transformed) to the predefined reference map. Additionally, the reference map can be a particular subject-specific activation map or a group-averaged map.

Specifically, we use the generalized Bayes method introduced by Wang et al. (2021)[3], which allows us to crop the subject-specific activation maps to enable them

to fit specific regions of interest (ROIs). This method implements a flexible general loss function-based pseudo-likelihood technique to measure and correct the relative misalignment between the floating and reference maps. The pseudo-likelihood used here represents how similar the reference map is to the warped floating map. It can be defined as the sum of squared differences (SSD) or any other forms based on different situations. Prior distributions are first estimated by matching the features of the activation maps (defined by local peaks). After that, the priors are assigned to the parameters corresponding to the misalignment. Moreover, the intensity correction term simultaneously accounts for the difference in activation level between the reference map and the floating map. Since the images are discrete and defined on a regular lattice (Banerjee et al., 2014; Cressie and Wikle, 2015)[19][20], it is necessary to use Gaussian process-based Kriging interpolation techniques to realign the images. Finally, the parameters of the model are inferred based on the draws from the posterior distributions. To do this, we use the Markov Chain Monte Carlo (MCMC) algorithm implemented in Rstan (Carpenter et al., 2017)[21] to simulate the posterior distributions.

The structure of the rest of this major paper is as follows: Chapter 2 describes the generalized Bayesian framework and the algorithms used for functional image registration. Chapter 3 introduces the data set and data preprocessing. We also evaluate the proposed method in this chapter and make a comparison with the standard method. The standard method is implemented by using the registration estimation toolbox in Matlab. The major paper concludes with a discussion in Chapter 4.

# Chapter 2

# Methods

Image registration is a process of establishing a relationship between the voxels in one image (the floating map) and their corresponding spatial locations in another image (the reference map)(Wang et al., 2021)[3]. In order to define the problem specifically, suppose there exists a set of voxel locations $\mathbf{S} = (s_1, s_2, \ldots, s_v)^T$, which respectively located at floating map $\mathbf{Y} = (Y(s_1), Y(s_2), \ldots, Y(s_v))^T$ and reference map $\mathbf{R} = (R(s_1), R(s_2), \ldots, R(s_v))^T$. In addition, a transformation operator $T$ is used here to describe the mapping (e.g., translation, scaling or rotations), which would be used to align floating map $\mathbf{Y}$ with the reference map $\mathbf{R}$.

Let $\mathbf{w}$ stand for the collection of parameters that define the transformation function $T$, which is defined as $T(\cdot, \mathbf{w})$. Our objective is to estimate $T$ via $\mathbf{w}$ with the appropriate quantification of uncertainty. After that, according to the posterior Markov Chain Monte Carlo (MCMC) samples derived from the Bayesian framework, we can then determine credible intervals for the transformation parameters.

For the traditional Bayesian solution, it is generally necessary to have a likelihood function that depicts the connection between the floating map and the reference map. Based on Wang et al. (2021)[3], we can choose several forms of the likelihood function. The first one is $\mathbf{R} \sim_d F(\mathbf{Y}(T), \theta)$, $\mathbf{R}(s) = b\mathbf{Y}(T(s, \mathbf{w})) + \mathbf{e}(s)$, for $s \in \mathbf{S}$. However, this is not desirable in our context because the observed floating map is expected to be as noisy as the reference map or even noisier. In addition, since this approach generates numerous generative models for the shared reference map $\mathbf{R}$, it becomes problematic when dealing with multiple floating maps $\mathbf{Y}_i$. For the second one, we can select the likelihood function that changes the responsibilities of $\mathbf{Y}$ and $\mathbf{R}$. To be more specific, this model-based approach can be denoted as: $\mathbf{Y}(s) = b\mathbf{R}\left(T_{rv}\left(s, \mathbf{w}'\right)\right) +$ error, where $T_{rv}\left(s, \mathbf{w}'\right)$ means the reverse transformation operator, mapping $\mathbf{R}$ to $\mathbf{Y}$ with parameter $\mathbf{w}'$. This method successfully addresses the problems above. However, it also has some drawbacks. For example, the model fit above does not incorporate the subject-specific features that are not present in the reference map. It is important to note that this switched approach does not offer a direct method for registering the floating image with the reference image. The third likelihood function form considers inverse consistency criteria to solve the problems in the second choice. As in Wang et al. (2021)[3], in order to apply this method, we need to utilize two loss functions, where one defining similarity between $\mathbf{R}$ and $\mathbf{Y}(T)$, and the other defining similarity between $\mathbf{Y}$ and $\mathbf{R}\left(T_{rv}\right)$. It will simultaneously estimate the forward and reverse transformation operator $T$ and $T_{rv}$ with the restriction that their composition is the identity operator. However, since this method uses two different generative models

$\mathbf{Y} \mid \mathbf{R}(T_{rv})$ and $\mathbf{R} \mid \mathbf{Y}(T)$ for the same set of data $(\mathbf{R}, \mathbf{Y})$, it is not easy to make it amenable to a fully Bayesian treatment. Due to the reasons above, we forego a fully model-based formulation in favour of a more adaptable generic loss function-based strategy introduced by Wang et al. (2021)[3]. This strategy can measure and correct the relative misalignment between individual map $\mathbf{Y}$ and reference map $\mathbf{R}$ at the same time. We go into further of this approach in the next section.

## 2.1 Generalized Bayes Framework for Registration

Based on the Generalized Bayes Framework introduced by Wang et al. (2021)[3], we firstly consider a loss-function which only relative to the forward transformation, denoted as $\ell(\mathbf{R}, \mathbf{Y}(T), \mathbf{w})$, where $\mathbf{Y}(T) = (Y(T(s_1)), \ldots, Y(T(s_v)))^T$, here $\mathbf{w}$ denotes all the possible parameters in the forward transformation and loss-function $\ell$. Given the loss-function $\ell$, suppose $\pi(\mathbf{w})$ is the prior of parameters, a posterior distribution which does not depend on a complete probability distribution model for the data but simply a loss function is given by:

$$\pi(\mathbf{w} \mid data) = \frac{\exp\left(-\ell(\mathbf{R}, \mathbf{Y}(T), \mathbf{w})\right)\pi(\mathbf{w})}{\int \exp\left(-\ell(\mathbf{R}, \mathbf{Y}(T), \mathbf{w})\right)\pi(\mathbf{w})d\mathbf{w}} \tag{2.1}$$

The posteriors above are called Gibbs posteriors or generalized posteriors. Their posterior summaries are called Laplace Type Estimators (LTE) in Chernozhukov and Hong (2003)[25]. According to Wang et al. (2021)[3], the multiple registrations to a single template are compatible using this loss-function technique. The total loss is calculated by adding the losses of each floating map when working with numerous

floating maps. Hence, by separating loss functions from likelihoods, the generalized Bayes method of registration overcomes the incompatibility (described above) of a model-based approach for registering multiple floating maps.

Furthermore, as in Wang et al. (2021)[3], we use the loss function defined below. It uses the sum-of-squared differences (SSD) between the floating map $\mathbf{Y}$ and reference map $\mathbf{R}$.

$$\ell(\mathbf{R}, \mathbf{Y}, b, \phi, \mathbf{w}) = \frac{1}{\phi^2}\|\mathbf{R} - b\mathbf{Y}(T(\cdot, \mathbf{w}))\|_{L_2}^2 \tag{2.2}$$

$$\ell\left(\mathbf{Y}, \mathbf{R}, b', \phi, \mathbf{w}'\right) = \frac{1}{\phi^2}\left\|\mathbf{Y} - b'\mathbf{R}\left(T_{rv}\left(\cdot, \mathbf{w}'\right)\right)\right\|_{L_2}^2 \tag{2.3}$$

Where $\mathbf{w}$ and $\mathbf{w}'$ denotes the parameters of forward and reverse transformation operators, respectively, $\phi$ is the scaling parameter used to control the learning rate from the loss function and is supposed to be equal either for reverse or forward loss. Also, the difference in intensities between the wrapped map $\mathbf{Y}$ and reference map $\mathbf{R}$ is corrected by a factor $b$ and $b'$. Additionally, to satisfy the inverse consistency requirement proposed by Christensen and Johnson (2001)[26] and Johnson and Christensen (2002)[27], as in Wang et al.(2021)[3], we extend the loss function as below:

$$\begin{aligned}\ell_{ic}\left(\mathbf{Y}, \mathbf{R}, b, b', \phi, \mathbf{w}, \mathbf{w}'\right) = \ell(\mathbf{R}, \mathbf{Y}, b, \phi, \mathbf{w}) + \ell\left(\mathbf{Y}, \mathbf{R}, b', \phi, \mathbf{w}'\right) \\ + \lambda_{rv}\left\|T_{rv}(T(\cdot)) - Id(\cdot)\right\|_F\end{aligned} \tag{2.4}$$

The penalty term of inverse consistency is the last term in (2.4), which limits forward and reverse transformation composition to somewhat near identity. In addition, the hyper-parameter $\lambda_{rv}$ is used to balance the intensity loss and inverse consistency penalty. Before we implement this new technique, we should first determine how spatial interpolation is carried out on the discrete images to get $\mathbf{Y}(T)$ used by the loss

function (Wang et al., 2021)[3]. In addition, we should also choose the appropriate transformation operator, reference map and hyper-parameters and the prior parameters used in the posterior. We first introduce how to obtain spatial interpolation in the next section.

## 2.2 Kriging Interpolation

As the images are defined on a regular lattice, the discreteness leads to utilizing Gaussian process-based kriging interpolation techniques, a point-level spatial prediction (Banerjee et al., 2014; Cressie and Wikle, 2015)[19][20] while realigning images. Hence, we assume a Gaussian process (GP) prior for the latent template to capture its spatial features so that we can obtain warped image $\mathbf{Y}(T)$ via kriging interpolation. It is important to note that the wrapped images $\mathbf{Y}(T)$ need to be evaluated for a specific value of the transformation parameters $\mathbf{w}$ when the parameters are updated via optimization or sampling using the loss function.

Kriging is an interpolation method based on a Gaussian process governed by prior covariances, also called Gaussian process regression, originated in geostatistics and is known as Kriging in statistics. With appropriate assumptions about the prior, kriging can provide the best linear unbiased prediction (BLUP) at unsampled locations (Chung et al., 2019)[28]. The problem is one of the usual optimal spatial prediction: given observations of a random field $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_V))'$, what is the best predictor of variable $Y$ at a site $s_0$ where it has not been observed.

Before we introduce the principle of kriging interpolation used in Wang et al.

(2021)[3], there are several elements of point-referenced modelling that need to be illustrated first. Suppose the mean of our spatial process was denoted as $\mu(s) = \mathbf{E}(Y(s))$ and the associated variance of this spatial process $Y(s)$ exists for all $s \in D$. As a Gaussian process, $\mathbf{Y} = (Y(s_1), \ldots, Y(s_n))^T$ follows a multivariate normal distribution for any set of $\{s_1, s_2, \ldots, s_n\}$ $(\forall n \geq 1)$.

Except strictly stationary (or strong stationary) and weak stationary (or second-order stationary), there is another type which is called intrinsic stationary. The definitions of the first two types of stationary are shown in Appendix A. Suppose $\mathbf{E}[Y(s+h) - Y(s)] = 0$, and define:

$$\mathbf{E}[Y(s+h) - Y(s)]^2 = \mathbf{Var}(Y(s+h) - Y(s)) = 2\gamma(h). \tag{2.5}$$

The equation above makes sense only when the mean and variance of $Y(s+h) - Y(s)$ solely depend on $h$. If this is the case, we say the process is intrinsically stationary. The function $2\gamma(h)$ is then called the variogram, and $\gamma(h)$ is so called semi variogram (Cressie and Wikle, 2015). Kriging relies on the semi variogram. In simple terms, semivariograms quantify autocorrelation because it graphs out the variance of all pairs of data according to distance. In general, when the distance between two points $Y(s+h)$ and $Y(s)$ is short (or say $\|h\|$ is small), we would expect there exists more similarity between them. When distance $\|h\|$ increases, we would expect less similarity between $Y(s+h)$ and $Y(s)$.

As in Cressie and Wikle (2015)[20], the covariance relationship between the values of the process at any two locations can be summarized by a covariance function $C$ or $\mathbf{C}$, where $\mathbf{Cov}(Y(s+h), Y(s)) = C(h)$, or say $\mathbf{Cov}(Y(s_i), Y(s_j)) = \mathbf{C}(s_i, s_j)$.

Therefore, based on (2.5), the relationship between variogram $2\gamma(h)$ and covariance is shown below:

$$\begin{aligned} 2\gamma(h) &= \mathbf{Var}(Y(s+h) - Y(s)) \\ &= \mathbf{Var}(Y(s+h)) + \mathbf{Var}(Y(s)) - 2\mathbf{Cov}(Y(s+h), Y(s)) \\ &= C(0) + C(0) - 2C(h) \\ &= 2[C(0) - C(h)]. \end{aligned} \tag{2.6}$$

Hence,

$$\gamma(h) = C(0) - C(h). \tag{2.7}$$

Another important related concept is isotropy. The variogram is said to be isotropic if the semi variogram function $\gamma(h)$ relies solely on the separation vector's length $\|h\|$; that is, if $\gamma(h)$ is a real-valued function of a univariate argument and can be written as $\gamma(\|h\|)$. If not, we say it is anisotropic.

Isotropic variograms are popular due to their readability and simplicity, especially since there are several reasonable and simple parametric forms of semi-variograms to choose from. As in Wang et al. (2021)[3], we use the popular exponential parametric form, which is shown below, denote $\|h\|$ as $d$:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi d)) & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

It can also be written as $C$ scale:

$$C(d) = \lim_{u \to \infty} \gamma(u) - \gamma(d) = \tau^2 + \sigma^2 - \left[\tau^2 + \sigma^2(1 - \exp(-\phi d))\right] = \sigma^2 \exp(-\phi d)$$

Thus,

$$C(d) = \begin{cases} \tau^2 + \sigma^2 & \text{if } d = 0 \\ \sigma^2 \exp(-\phi d) & \text{if } d > 0. \end{cases} \tag{2.9}$$

We introduce a simple example to illustrate the principle of kriging interpolation. As in Banerjee et al. (2014), we first take a formal look at kriging in the context of the Gaussian process. Given observations of a random field $\mathbf{Y} = (Y(s_1), ..., Y(s_V))^T$. If the covariate values $X = (x(s_1), ..., x(s_V))^T$ and $\mathbf{x}_0$ (the covariate values at new site $s_0$) are available for incorporation into the analysis, the procedure is often referred to as universal kriging. The universal kriging model has the following form:

$$\mathbf{Y} = X\beta + \epsilon, \text{where } \epsilon \sim \mathcal{N}_V(0, \Sigma) \tag{2.10}$$

$\mathbf{Y} = (Y(s_1), Y(s_2), ..., Y(s_V))^T$ is a $V \times 1$ dimensional vector that includes the responses $Y(s_i)$, where $i = 1, ..., V$. As a particular case of universal kriging (Cressie and Wikle, 2015)[20], $X$ is a vector of ones of length $V$ and is usually denoted as $\mathbf{1}$. $\beta$ in (2.10) is a scalar. $\epsilon$ is a $V$-dimensional vector which is the error. In addition, $\Sigma = \mathbf{C}(s_i, s_j) = \sigma^2 \exp(-\phi\ d_{ij})$, which is a $V \times V$ dimensional covariance matrix with $i = 1, ..., V, j = 1, ..., V$, where $d_{ij} = ||s_i - s_j||$ is the distance between $s_i$ and $s_j$. For the response $Y(s_0)$ and covariate value $\mathbf{x}_0$ at the new site $s_0$, the model can be written as $Y(s_0) = \mathbf{x}_0\beta + \epsilon_0$, where $\epsilon_0 \sim \mathcal{N}(0, \Sigma_0)$, $\Sigma_0 = \mathbf{C}(s_0, s_0) = C(0) = \tau^2 + \sigma^2$.

As in Cressie and Wikle (2015)[20], we now pose our prediction problem as follows: we seek the function $h(\mathbf{y})$ that minimizes the mean-squared prediction error,

$$\mathbf{E}[(Y(s_0) - h(\mathbf{y}))^2 | \mathbf{y}], \tag{2.11}$$

where $\mathbf{y}$ is the data we already observed. By adding and subtracting the conditional mean $\mathbf{E}[Y(s_0)|\mathbf{y}]$ inside the square. Since the expectation of the cross-product term equals zero, after grouping terms and squaring, we obtain the following:

$$\mathbf{E}\left[\left(Y(s_0) - h\left(\mathbf{y}\right)\right)^2 \middle| \mathbf{y}\right] = \mathbf{E}\left\{\left(Y(s_0) - \mathbf{E}[Y(s_0)|\mathbf{y}]\right)^2 \middle| \mathbf{y}\right\} + \left\{\mathbf{E}[Y(s_0)|\mathbf{y}] - h\left(\mathbf{y}\right)\right\}^2.$$

But, since the second term on the right-hand side is nonnegative, we have:

$$\mathbf{E}\left[\left(Y(s_0) - h\left(\mathbf{y}\right)\right)^2 \middle| \mathbf{y}\right] \geq \mathbf{E}\left\{\left(Y(s_0) - \mathbf{E}[Y(s_0)|\mathbf{y}]\right)^2 | \mathbf{y}\right\}.$$

Equality holds if and only if $h\left(\mathbf{y}\right) = \mathbf{E}[Y(s_0)|\mathbf{y}]$. Hence, the predictor $h(\mathbf{y})$ minimizes the error is the conditional expectation of $Y(s_0)$ given the data $\mathbf{y}$, which is $\mathbf{E}[Y(s_0)|\mathbf{y}]$.

Next, we show how to get the posterior mean of $Y(s_0)$. From a Bayesian point of view, this $h\left(\mathbf{y}\right)$ is just the posterior mean of $Y(s_0)$. At the same time, this posterior mean is linear under a Gaussian process model assumption (Cressie and Wikle, 2015)[20]. Based on the classical theory of multivariate normal, suppose $Y_1$ is a $m$-column vector, and $Y_2$ is a $(n-m)$ column vector. Similarly, let $\mu = (\mu_1, \mu_2)^T$ and let $\sum = \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}$ with $\sum_{21} = \sum_{12}^T$. Suppose that

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}_n \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right), \Sigma_{21} = \Sigma_{12}^T,$$

Then, the conditional distribution $Y_1 \mid Y_2$ is normal and has the following mean and

variance:

$$\mathbf{E}\left[Y_1 \mid Y_2\right] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(Y_2 - \mu_2\right),$$

$$\mathbf{Var}\left[Y_1 \mid Y_2\right] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Based on the framework above, we have $Y_1 = Y(s_0)$ (the point we need to estimate at new site $s_0$) and $Y_2 = (Y(s_1), \ldots, Y(s_v))^T$. Let $\mathbf{K}_0$ be a $V-$ column vector where $ith$ component is $\mathbf{C}(s_0, s_i), i = 1, 2, ..., V$. Let $\mathbf{C}$ be a $V \times V$-matrix where $(i, j)$-component is $\mathbf{C}(s_i, s_j), i = 1, 2, ..., V, j = 1, 2, ..., V$. We then have:

$$\Sigma_{12} = \mathbf{K}_0^T,$$

$$\Sigma_{22} = \mathbf{Var}(Y) = \mathbf{C}.$$

As in Banerjee et al. (2014)[19], we substitute these values into $\mathbf{E}(Y_1|Y_2 = Y)$ above and then obtain

$$\mathbf{E}\left[Y(s_0) \mid Y\right] = \mathbf{x}_0^T\beta + \mathbf{K}_0^T\mathbf{C}^{-1}(Y - X\beta),$$

where $X$ is a vector of ones of length $V$, $\mathbf{K}_0$ is a $V \times 1$ dimensional vector and $\mathbf{C}$ is a $V \times V$ matrix with $(i, j)^{th}$ entries $\mathbf{C}(s_i, s_j)$. We remark that this solution supposes we have actually observed the covariate value $\mathbf{x}_0 = x(s_0)$ at new site $s_0$. In practice, the model parameter like $\beta$ is unknown, it must be estimated from the data. Hence, by using the usual weighted least squares estimator $\hat{\beta}$ of $\beta$, where $\hat{\beta} = \left(X^T\mathbf{C}^{-1}X\right)^{-1}X^T\mathbf{C}^{-1}Y$, which is a scalar, we can then obtain $h(\mathbf{y})$ as below:

$$\widehat{h\left(\mathbf{y}\right)} = \mathbf{x}_0^T\frac{\mathbf{1}^T\mathbf{C}^{-1}Y}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}} + \mathbf{K_0}^T\mathbf{C}^{-1}\left(Y - \mathbf{1}\frac{\mathbf{1}^T\mathbf{C}^{-1}Y}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}}\right).$$

However, in our context, the activation maps $\mathbf{Y}$ are observed on a predefined set of $V$ voxels $(s_1, \ldots, s_v)^T$, the warped images $\mathbf{Y}(\mathbf{T})$ at the set of locations $\mathbf{T}(V)$

are acquired through spatial interpolation. Here, $\mathbf{T}(V) = (T(s_1), ..., T(s_V))^T$. This means we need to estimate several new points jointly rather than a single point above. Hence, based on Rodolphe (2014)[29], for the joined prediction at many points, we can use similar formulas but make some changes. First, we should note $Y_1$ now change to $(Y(T(s_1)), ..., Y(T(s_V)))^T$ (the new points we need to obtain via interpolation) and $Y_2 = (Y(s_1), \ldots, Y(s_v))^T$. Second, $\Sigma_{12} = \mathbf{C}(T(s_i), s_j) = \mathbf{K}^T, (i = 1, ..., V, j = 1, ..., V)$, is a $V \times V$ dimensional matrix. Third, we need to assume we have actually observed the covariate values $\mathbf{X}_0 = (x(T(s_1)), ..., x(T(s_V)))$ at new sites $(T(s_1), ..., T(s_V))^T$. Then based on these changes, we can obtain $\mathbf{E}[Y_1 \mid Y_2]$ as

$$\mathbf{E}[Y_1 \mid Y_2] = \mathbf{X}_0^T \hat{\beta} + \mathbf{K}^T \mathbf{C}^{-1}(Y - X\hat{\beta}). \tag{2.12}$$

As in Banerjee et al. (2014)[19] and Wang et al. (2021)[3], $\mathbf{X}_0^T$ and $X$ are two $V \times 1$ dimensional vectors of ones. $\hat{\beta} = \left(X^T \mathbf{C}^{-1} X\right)^{-1} X^T \, \mathbf{C}^{-1} Y$, is still a scalar. $\mathbf{C}$ and $\mathbf{K}$ are $V \times V$ matrices with $(i, j)^{th}$ entries $\mathbf{C}(s_i, s_j)$ and $\mathbf{C}(s_i, T(s_j))$ respectively. Finally, by replacing the $\hat{\beta}$, $\mathbf{X}_0$ and $X$ in (2.12), we can get the kriging interpolator as below:

$$\widehat{\mathbf{Y}(T)} = \frac{\mathbf{1}^T \mathbf{C}^{-1} Y}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \mathbf{1} + \mathbf{K}^T \mathbf{C}^{-1} \left( Y - \frac{\mathbf{1}^T \mathbf{C}^{-1} Y}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \mathbf{1} \right) \tag{2.13}$$

Because we choose the exponential covariance kernel, i.e., $\mathbf{C}(s_i, s_j) = \sigma^2 \exp\{-\rho \| s_i - s_j \|\}$, the spatial parameters $\rho$ and $\sigma$ in the covariance kernel can be estimated with other parameters together via the Bayesian framework.

## 2.3 Transformation Operator

In general, affine transformations include, for instance, translation, scaling, homothety, similarity, reflection, rotation, shear mapping, and compositions of them in any order or combination. Following Cederberg (2001)[30], we apply a similarity transformation operator $T$ to the generic framework for Bayesian registration suggested by the loss function (2.4). As a special case of the general 2D affine transformation, transformation operator $T$ can be expressed as $T(s) = As + \theta$, where $s = (s_x, s_y)^T$, $\theta = (\theta_x, \theta_y)^T \in \mathbb{R}^2$. Matrix $A$ in the transformation operator $T$ relates to rotation and scaling. In addition, $A$ can be expressed in several ways as following proposition.

**Proposition 2.3.1.** *If* $A = \left(\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}\right)$, *then under the constraint* $A_{11}A_{12} + A_{21}A_{22} = 0$, *matrix* $A$ *can also be written as:*

$$A = \begin{pmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{pmatrix} \begin{pmatrix} \operatorname{sign}(A_{11})\sqrt{A_{11}^2 + A_{21}^2} & 0 \\ 0 & A_{22}/\cos(\omega) \end{pmatrix} \tag{2.14}$$

*where* $\omega = \arctan(A_{21}/A_{11})$.

*Proof.* In general, for the case of 2D images, matrix $A$ can be expressed in the following form that is related to rotation and scaling (Wang et al., 2021)[3].

$$A = \begin{pmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{pmatrix} \times \begin{pmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{pmatrix},$$

where $\omega \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $\sigma_x > 0$, $\sigma_y > 0$. Hence, we first rewrite the transformation matrix $A$ as below:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

$$\begin{cases} a\cos\omega - c\sin\omega = A_{11} & ① \\ b\cos\omega - d\sin\omega = A_{12} & ② \\ a\sin\omega + c\cos\omega = A_{21} & ③ \\ b\sin\omega + d\cos\omega = A_{22} & ④ \end{cases}$$

From equations ① and ③, we could get

$$\begin{cases} c\sin\omega = a\cos\omega - A_{11} \\ c\cos\omega = A_{21} - a\sin\omega \end{cases}$$

$$\begin{cases} ① \times \cos\omega = a\cos^2(\omega) - c\sin\omega\cos\omega = A_{11}\cos\omega \\ ③ \times \sin\omega = a\sin^2(\omega) + c\sin\omega\cos\omega = A_{21}\sin\omega \end{cases}$$

$$\Rightarrow a = A_{11}\cos\omega + A_{21}\sin\omega \tag{2.15}$$

Similarly from equations ② and ④, we could get

$$\begin{cases} d\sin\omega = b\cos\omega - A_{12} \\ \\ d\cos\omega = A_{22} - b\sin\omega \end{cases}$$

$$\begin{cases} ② \times \cos\omega = b\cos^2(\omega) - d\sin\omega\cos\omega = A_{12}\cos\omega \\ \\ ④ \times \sin\omega = b\sin^2(\omega) + d\sin\omega\cos\omega = A_{22}\sin\omega \end{cases}$$

$$\Rightarrow b = A_{12}\cos\omega + A_{22}\sin\omega \tag{2.16}$$

Since $A = \begin{pmatrix} \cos\omega & -\sin\omega \\ \\ \sin\omega & \cos\omega \end{pmatrix} \times \begin{pmatrix} \sigma_x & 0 \\ \\ 0 & \sigma_y \end{pmatrix}, \omega \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, then $b = c = 0$.

When $c = 0$,

$$\begin{cases} a\cos\omega - A_{11} = 0 \Rightarrow a = \frac{A_{11}}{\cos\omega} \\ \\ A_{21} - a\sin\omega = 0 \Rightarrow \sin\omega = \frac{A_{21}}{a} \end{cases}$$

$$\Rightarrow \quad \sin\omega = \frac{A_{21}}{A_{11}}\cos\omega$$

$$\Rightarrow \quad \tan\omega = \frac{A_{21}}{A_{11}}$$

$$\Rightarrow \quad 1 + \tan^2(\omega) = \frac{A_{11}^2 + A_{21}^2}{A_{11}^2} = \frac{1}{\cos^2(\omega)}$$

$$\Rightarrow \quad \cos^2(\omega) = \frac{A_{11}^2}{A_{11}^2 + A_{21}^2}$$

$$\Rightarrow \quad \cos\omega = \frac{\text{sign}\,(A_{11})\,A_{11}}{\sqrt{A_{11}^2 + A_{21}^2}}, \sin\omega = \frac{A_{21}}{A_{11}}\cos\omega = \frac{\text{sign}\,(A_{11})\,A_{21}}{\sqrt{A_{11}^2 + A_{21}^2}} \tag{2.17}$$

When $b = 0$, we could obtain:

$$\begin{aligned} b = A_{12}\cos\omega + A_{22}\sin\omega = \frac{\operatorname{sign}(A_{11})(A_{11}A_{12} + A_{22}A_{21})}{\sqrt{A_{11}^2 + A_{21}^2}} = 0 \\ A_{11}A_{12} + A_{22}A_{21} = 0 \end{aligned} \tag{2.18}$$

For $a$, replace $\cos\omega$ and $\sin\omega$ with (2.17):

$$\begin{aligned} a = A_{11}\cos\omega + A_{21}\sin\omega &= \frac{\operatorname{sign}(A_{11})A_{11}^2 + \operatorname{sign}(A_{11})A_{21}^2}{\sqrt{A_{11}^2 + A_{21}^2}} \\ &= \operatorname{sign}(A_{11})\sqrt{A_{11}^2 + A_{21}^2} \end{aligned} \tag{2.19}$$

For $d$, since $d\cos\omega = A_{22} - b\sin\omega(\cos\omega \neq 0)$, we then replace $b$ with (2.16):

$$\begin{aligned} d &= \frac{A_{22} - (A_{12}\cos\omega + A_{22}\sin\omega)\sin\omega}{\cos\omega} \\ &= \frac{A_{22}\left(1 - \sin^2\omega\right) - A_{12}\cos\omega\sin\omega}{\cos\omega} = A_{22}\cos\omega - A_{12}\sin\omega \end{aligned} \tag{2.20}$$

Then, $b = c = 0$, $a = \operatorname{sign}(A_{11})\sqrt{A_{11}^2 + A_{21}^2}$, $d = A_{22}\cos\omega - A_{12}\sin\omega$,

$$A = \begin{pmatrix} \cos\omega & -\sin\omega \\ \sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} \operatorname{sign}(A_{11})\sqrt{A_{11}^2 + A_{21}^2} & 0 \\ 0 & A_{22}\cos\omega - A_{12}\sin\omega \end{pmatrix} \tag{2.21}$$

If we also replace $\cos\omega$ and $\sin\omega$ with (2.17), under the constraint we obtained from (2.18), we will get:

$$\begin{aligned} A_{22}\cos\omega - A_{12}\sin\omega &= \frac{A_{22}\operatorname{sign}(A_{11})A_{11} - A_{12}\operatorname{sign}(A_{11})A_{21}}{\sqrt{A_{11}^2 + A_{21}^2}} \\ &= \frac{\operatorname{sign}(A_{11})[A_{22}A_{11} - A_{12}A_{21}]}{\sqrt{A_{11}^2 + A_{21}^2}} \\ &= \frac{\operatorname{sign}(A_{11})\left[A_{22}A_{11} + \frac{A_{21}^2 A_{22}}{A_{11}}\right]}{A_{11}\sqrt{A_{11}^2 + A_{21}^2}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\operatorname{sign}(A_{11}) \cdot A_{22} \cdot (A_{11}^2 + A_{21}^2)}{A_{11}\sqrt{A_{11}^2 + A_{21}^2}} \\
&= \frac{A_{22} \cdot \sqrt{(A_{11}^2 + A_{21}^2)}}{\operatorname{sign}(A_{11})\, A_{11}} = A_{22}/\cos\omega
\end{aligned}
\tag{2.22}
$$

Hence,

$$A_{22}\cos\omega - A_{12}\sin\omega = A_{22}/\cos\omega.$$

Then the decomposition of the matrix $A$ can also be simplified as

$$
A = \begin{pmatrix} A_{11} & A_{12} \\ \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} \cos\omega & -\sin\omega \\ \\ \sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} \operatorname{sign}(A_{11})\sqrt{A_{11}^2 + A_{21}^2} & 0 \\ \\ 0 & A_{22}/\cos\omega \end{pmatrix}.
\tag{2.23}
$$

□

We note that the expression above is the result of the QR decomposition (Wang et al., 2021)[3]. The first term is the rotation matrix, followed by the scaling matrix. Hence, under the constraint, we obtained from (2.18): $A_{11}A_{12} + A_{21}A_{22} = 0$, there exists a bijection mapping between $(\sigma_x, \sigma_y, \omega)^T$ and $\mathrm{vec}(A)$. In addition, according to (2.23), we can obtain positive scalings ($\sigma_x > 0, \sigma_y > 0$) by setting $A_{11} > 0$ and $A_{22} > 0$. In the next section, we will use this relationship to formulate the prior for the transformation parameters.

## 2.4 Regularization of Prior

In the general registration problem, optimizing for transformation parameters faces several problems. Firstly, Fischer and Modersitzki (2008)[31] highlighted a typical in-

stance where two different registration procedures might produce the same outcome. In this case, they subtly illustrated how either translation or rotation could produce the same optimum. Secondly, it can be challenging to steer clear of undesirable local optima. However, these issues can be resolved by incorporating a regularization or penalty term. To efficiently address this problem, this idea is modified following the method of Wang et al. (2021)[3]. In practice, the method is based on the regularization of priors of the generalized Bayes framework.

The regularization of $T$ is denoted by $\mathbf{Reg}(T)$. $\mathbf{Reg}(T)$ is used to calculate how far the proposed $T$'s transformed coordinates are from an initial warping map $T_0$, which is determined by the hyperparameters $A_0$ and $\theta_0$. As in Wang et al. (2021)[3], we consider that $\mathbf{Reg}(T)$ has the following representation:

$$\mathbf{Reg}(T) = \sum_{s\in\mathcal{R}} ||T(s) - T_0(s)||^2_{L_2} = \sum_{s\in\mathcal{R}} ||(A - A_0)s + \theta - \theta_0||^2_{L_2}. \tag{2.24}$$

Then, the regularization of $T$ above can also be simplified as:

$$\mathbf{Reg}(T) = \sum_{s\in\mathcal{R}} [(A - A_0)s + \theta - \theta_0][(A - A_0)s + \theta - \theta_0]^T.$$

Since $\theta = (\theta_x, \theta_y)^T, s = (s_x, s_y)^T$, then we have:

$$\mathbf{Reg}(T) = \mathbf{trace}\bigg([(A - A_0)s + \theta - \theta_0][(A - A_0)s + \theta - \theta_0]^T\bigg).$$

Then

$$\mathbf{Reg}(T) = \mathbf{trace}\bigg([s_x, s_y, \mathbf{1}]([A, \theta]^T - [A_0, \theta]^T)([s_x, s_y, \mathbf{1}]([A, \theta]^T - [A_0, \theta_0]^T))^T\bigg),$$

and then

$$\mathbf{Reg}(T) = \mathbf{trace}\bigg([s_x, s_y, \mathbf{1}][M - M_0][M - M_0]^T[s_x, s_y, \mathbf{1}]^T\bigg),$$

this gives

$$\mathbf{Reg}(T) = \mathbf{trace}\bigg([M - M_0]^T \Sigma_s [M - M_0]\bigg),$$

where

$$M = [A, \theta]^T = \begin{pmatrix} A_{11} & A_{12} & \theta_x \\ A_{21} & A_{22} & \theta_y \end{pmatrix}^T, \; M_0 = [A_0, \theta_0]^T, \text{and } \Sigma_s = \begin{pmatrix} s_x^T s_x & s_x^T s_y & s_x^T \mathbf{1} \\ s_x^T s_y & s_y^T s_y & s_y^T \mathbf{1} \\ s_x^T \mathbf{1} & s_y^T \mathbf{1} & \mathbf{1^T 1} \end{pmatrix}.$$

In addition, $M_0$ is the prior transformation parameters of $T$. Similarly, the regularization of $T_{rv}$ can also be written as:

$$\mathbf{Reg}(T_{rv}) = \mathbf{trace}\bigg([M' - M'_0]^T \Sigma_s [M' - M'_0]\bigg),$$

with $M' = [A', \theta']^T, T_{rv} = A's + \theta'$, $M'_0$ is the prior transformation parameters of $T_{rv}$. At the same time, to satisfy the inverse-consistency rules, we have $A'_0(A_0 s + \theta_0) + \theta'_0 = s$, so that $T_{rv0}$ is constrained to its inverse $T_0$. We introduce how to apply $\mathbf{Reg}(T)$ and $\mathbf{Reg}(T_{rv})$ in the generalized Bayes framework in the rest of this section.

Based on the loss function (2.4), and as in Wang et al. (2021)[3], the warped parameters' generalized (Gibbs) posterior distribution has the following form:

$$p(T, T_{rv}, \phi, b, b' | \mathbf{Y}, \mathbf{R}) \propto \mathbf{exp}\{-\ell_{\mathbf{ic}}(\mathbf{Y}, \mathbf{R}, b, b', \phi, w, w')\} \pi(T, T_{rv}, \phi, b, b'), \quad (2.25)$$

where $\pi(T, T_{rv}, \phi, b, b') = \pi(T|\phi)\pi(T_{rv}|\phi)\pi(b'|\phi)\pi(b|\phi)\pi(\phi)$, is the regularised prior. Therefore, the posterior takes the form:

$$\begin{aligned} p(T, T_{rv}, \phi, b, b' | \mathbf{Y}, \mathbf{R}) &\propto \exp\{-\ell_{\mathrm{ic}}(\mathbf{Y}, \mathbf{R}, b, b', \phi, w, w')\} \cdot \exp\{\log[\pi(\mathrm{T}, \mathrm{T_{rv}}, \phi, \mathrm{b}, \mathrm{b}')]\} \\ &= \exp\{-\ell_{\mathrm{ic}}(\mathbf{Y}, \mathbf{R}, b, b', \phi, w, w') + \log[\pi(T, T_{rv}, \phi, b, b')]\}. \end{aligned}$$

In order to apply $\mathbf{Reg}(T)$ and $\mathbf{Reg}(T_{rv})$, we need to select proper priors for $T$ and $T_{rv}$. The proper priors for $T$ and $T_{rv}$ can lead the posterior include terms related to $\mathbf{Reg}(T)$ and $\mathbf{Reg}(T_{rv})$. At the same time, as in Wang et al. (2021)[3], the intensity-correction terms are regularized by the factor $(\log(b) - \log(b_0))^2$ and $(\log(b') - \log(b'_0))^2$ in the regularization part. Without penalizing the values of $b$ away from some $b_0$, the **MCMC** may become caught in loss function valleys centred on a local minimum. For instance, for $b = 0$, the loss function will be flat for all values of the transformation parameters $\mathbf{w}$. Hence, we also need to select proper priors for $b$ and $b'$ so that the posterior includes regularization terms related to $(\log(b) - \log(b_0))^2$ and $(\log(b') - \log(b'_0))^2$. The term $b_0$ and $b'_0$ are two hyper-parameters whose selection will be discussed in the next section.

As in Wang et al. (2021)[3], the prior for $b$ leading to a relation with $(\log(b) - \log(b_0))^2$, is given by $\log(b)|\phi \sim \mathcal{N}(\log(b_0), \frac{\phi^2}{\lambda_b})$. The prior for $b'$ is similar to the prior for $b$. Hence, we can first obtain the priors for $b$ and $b'$ as below:

$$\pi(b|\phi) \propto \exp\left\{ -\frac{\lambda_b}{2\phi^2}(\log(b) - \log(b_0))^2 \right\},$$

$$\pi(b'|\phi) \propto \exp\left\{ -\frac{\lambda'_b}{2\phi^2}(\log(b') - \log(b'_0))^2 \right\}.$$

The prior for $M$ leading to a relation with $\mathbf{Reg}(\mathrm{T})$ is given by $\mathrm{vec(M)}|\phi \sim \mathcal{N}_2($ $\mathrm{vec}(M_0), \frac{\phi^2}{\lambda_T}\mathbf{I} \otimes \Sigma_s^{-1})$, where $\otimes$ denotes the Kronecker product and $\mathbf{I} \in \mathbb{R}^{\mathbf{2}\times\mathbf{2}}$, is the identity matrix. The prior for $M'$ is similar to the prior $M$. As we mentioned above, the priors for $M$ and $M'$ with parameters $M_0$ and $M'_0$ are used to ensure that the Gibbs posterior distribution include $\mathbf{Reg}(T)$ and $\mathbf{Reg}(T_{rv})$. Therefore, we first rewrite

matrix $M$ as $(M_1, M_2)$, where $M_1 = (A_{11}, A_{12}, \theta_x)^T, M_2 = (A_{21}, A_{22}, \theta_y)^T$. Then we rewrite matrix $M_0$ as $(M_{01}, M_{02})$, where $M_{01} = (A^0_{11}, A^0_{12}, \theta^0_x)^T, M_{02} = (A^0_{21}, A^0_{22}, \theta^0_y)^T$. We can then obtain the prior for the transformation parameters of $T$ as below:

$$\begin{aligned}\pi(T|\phi) &\propto \exp\left\{ -\frac{\lambda_T}{2\phi^2}\sum_{r=1}^{2}(M_r - M_{0r})^T\Sigma_s(M_r - M_{0r})\right\} \\ &= \exp\left\{ -\frac{\lambda_T}{2\phi^2}\mathbf{trace}\{(M - M_0)^T\Sigma_s(M - M_0)\}\right\} \\ &= \exp\left\{ -\frac{\lambda_T}{2\phi^2}\mathbf{Reg}(T)\right\}.\end{aligned}$$

Similarly,

$$\begin{aligned}\pi(T_{rv}|\phi) &\propto \exp\left\{ -\frac{\lambda_{Trv}}{2\phi^2}\mathbf{trace}\{(M' - M'_0)^T\Sigma_s(M' - M'_0)\}\right\} \\ &= \exp\left\{ -\frac{\lambda_{T_{rv}}}{2\phi^2}\mathbf{Reg}(T_{rv})\right\}.\end{aligned}$$

Then we combine all the results above and obtain the term $\pi(T|\phi)\pi(T_{rv}\phi)\pi(b'|\phi)\pi(b|\phi)$ of the Gibbs posterior distribution (2.25) as below:

$$\begin{aligned}&\pi(T|\phi)\pi(T_{rv}\phi)\pi(b'|\phi)\pi(b|\phi) \propto \\ &\exp\left\{ -\frac{\lambda_b}{2\phi^2}(\log(b) - \log(b_0))^2 - \frac{\lambda_T}{2\phi^2}\text{trace}\{(M - M_0)^T\Sigma_s(M - M_0)\}\right. \\ &\left. -\frac{\lambda'_b}{2\phi^2}(\log(b') - \log(b'_0))^2 - \frac{\lambda_{Trv}}{2\phi^2}\text{trace}\{(M' - M'_0)^T\Sigma_s(M' - M'_0)\}\right\}.\end{aligned} \tag{2.26}$$

As in Wang et al. (2021)[3], the terms $\lambda_b, \lambda'_b, \lambda_T$ and $\lambda_{T_{rv}}$ are the tuning parameters used to control the trade-off between the fitting of the registration likelihood and the regularization priors.

Finally, as in Wang et al., 2021[3], by using the standard reference prior for the scale parameter $\phi, \pi \propto 1/\phi^2$, we can obtain the full Gibbs posterior distribution of the

warping parameters by combining all the pieces in the generalized Bayes framework as below:

$$p(\theta,\sigma,w) \propto \phi^{-2V-14}\exp\Bigg\{-\frac{1}{2\phi^2}\|\mathbf{R}-b\mathbf{Y}(T)\|_{L_2}^2-\frac{1}{2\phi^2}\|\mathbf{Y}-b'\mathbf{R}(T_{ic})\|_{L_2}^2 \\ -\frac{\lambda_b}{2\phi^2}((\log(b)-\log(b_0))^2)-\frac{\lambda_T}{2\phi^2}\text{trace}\{(M-M_0)^T\Sigma_s(M-M_0)\} \\ -\frac{\lambda_{b'}}{2\phi^2}((\log(b')-\log(b'_0))^2)-\frac{\lambda_{T_{rv}}}{2\phi^2}\text{trace}\{(M'-M'_0)^T\Sigma_s(M'-M'_0)\}\Bigg\}. \tag{2.27}$$

We should note that by using the priors given by Wang et al. (2021)[3], the full Gibbs posterior distribution has the terms related to the regularization of $T, T_{rv}, b, b'$. At the same time, the explicit penalty for inverse consistency has been removed because it is already incorporated within the selection of prior transformations.

The choice of hyperparameters like $A_0$ and $\theta_0$ is important since it relates to the identifiability of the registration. We show the detail about how to select proper hyperparameters in Section 2.5. Note here, $M$ has 6 parameters ($A_{11}, A_{12}, A_{21}, A_{22}, \theta_x$ and $\theta_y$) but our transformation is defined by 5 parameters $w, \sigma_x, \sigma_y, \theta_x, \theta_y$ with correspond to rotation, scaling and translation. This is because the general 2D affine transformation which parametrized by $M$ is under the restriction $A_{11}A_{12} + A_{21}A_{22} = 0$. This restriction has a one-to-one relationship between $(\sigma_x, \sigma_y, \omega)^T$ (3 parameters) and vec($A$) (4 parameters), as stated at the end of Section 2.3. Thus, as in Wang et al. (2021)[3], we only use the multivariate normal prior for the 3-dimensional subset of $A$ along with $\theta$ to correspond to the priors on transformation parameters.

## 2.5 Corresponding Feature-based Prior Estimation

In order to ensure the registration problem is identified, it is essential to choose the regularization hyper-parameters $b_0$ and $M_0$ correctly. This section will discuss how to estimate prior (hyper-parameters) based on corresponding features. Loss functions (2.2) and (2.3) are solely based on the intensity information of the images and do not use landmark information from activation maps in image registration. By considering the landmark data, the prior (hyper-parameters) of the transformation can be estimated. The landmark is defined as a cluster's local maximum (maxima), including eight neighbouring voxels (Wang et al., 2021)[3].

Let $\mathbf{P}^R = \{s_i^R \in \mathbb{R}^2 | i = 1, ..., N_R\}$ and $\mathbf{P}^Y = \{s_i^Y \in \mathbb{R}^2 | i = 1, ..., N_Y\}$ denote the landmarks of reference map and floating map, respectively. $N_R$ and $N_Y$ are the number of landmarks in the reference map and floating map. As the first step, the coarse subregion of interest is manually determined by scientific principles. This subregion contains a subset of landmarks, $\hat{\mathbf{P}}^R \subset \mathbf{P}^R$, of size denoted by $\hat{N}_R$. We aim to find the corresponding subset of the landmarks $\hat{\mathbf{p}}^Y$ from $\mathbf{p}^Y$. For point set matching, many traditional methods are effective, such as Geometric Hashing (Mian et al., 2006)[32] and Iterative Nearest Point (Yang and Medioni, 1992; Besl and McKay, 1992)[33][34]. However, these approaches fail to capture the intensity properties of the maps. Because the size of the set of landmarks is somewhat tiny (i.e., $N_Y < 20$ and $\hat{N}_R < 5$), it allows us to apply the Brute Force Search approach. The Brute Force Search method will traverse all possible subsets of $\mathbf{p}^Y$ with $\hat{N}_R$ elements to find the

matching landmarks.

The Procrustes analysis can be used to estimate the similarity transformation between the landmarks. Here, we use the ABC Procrustean Algorithm, a modified form of Procrustes analysis (Awange et al., 2008)[35], which is presented in Figure 2.1.

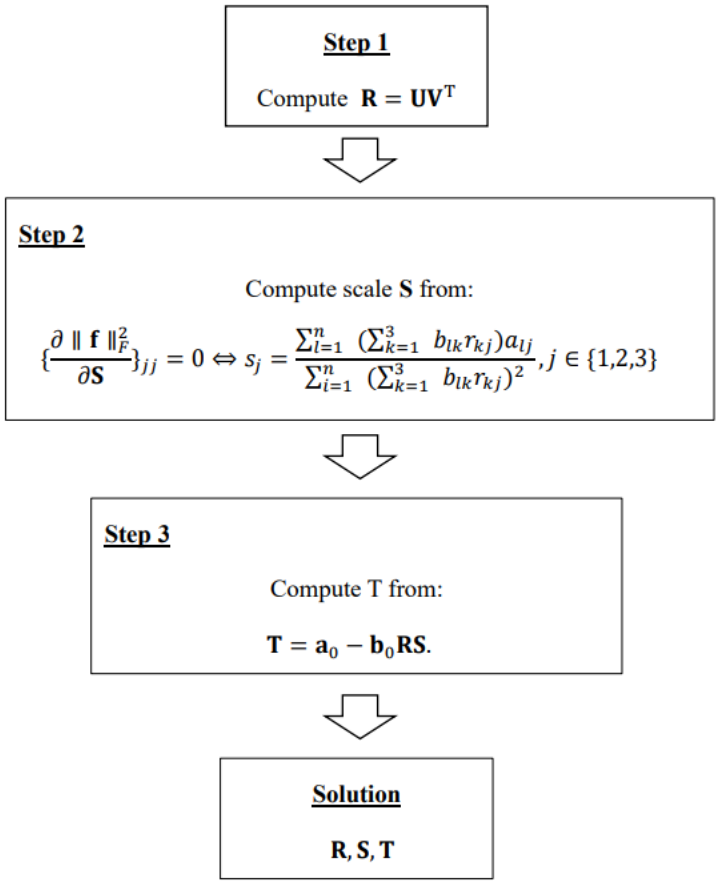



Figure 2.1: Procrustean 9-parameter transformation algorithm

To state this algorithm specifically, let us consider two coordinate configurations $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{B} \in \mathbb{R}^{n \times m}$ consisting of $n$ points in an $m$-dimensional Euclidean space.

In our context, A is the reference map, and B is the floating map. The 9-parameter transformation problem is defined by Awange et al.(2008)[35] as below:

$$\mathbf{A} = \mathbf{BRS} + \mathbf{vT}^T, \tag{2.28}$$

where the scalar parameter $\mathbf{S}$ is a diagonal matrix. The diagonal corresponding elements of $\mathbf{S}$ are $s_x, s_y, s_z$, where $s_i \in \mathbb{R}$ with $i = (x, y, z)$ being the $X, Y, Z$ scales in the directions, respectively. The solution of (2.28) includes nine parameters, i.e., three scales $\mathbf{S} \in \mathbb{R}^{3\times3}$, three rotations elements of $\mathbf{R} \in \mathbb{R}^{3\times3}$ and three translation parameters $\mathbf{T} \in \mathbb{R}^{3\times1}$.

The ABC Procrustean transformation algorithm introduced by Awange et al. (2008)[35] is presented in Figure 2.1. Before this algorithm is implemented, we must first calculate the rotation matrix $\mathbf{R}$ as step 1. Following the work by Awange et al. (2008)[35], the rotation matrix $\mathbf{R}$ can be obtained by using the Singular Value Decomposition(SVD) of $\mathbf{B}^T\mathbf{A}$, which is denoted as $\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{B}^T\mathbf{A}$. Then, we can successfully calculate the rotation matrix $\mathbf{R}$ as step 1 using the matrix $\mathbf{U}$ and $\mathbf{V}$ which are obtained from the SVD of $\mathbf{B}^T\mathbf{A}$.

Secondly, as described in Awange et al. (2008)[35], $\mathbf{a}_0$ and $\mathbf{b}_0$ in step 3 are supposed to be the center of mass (centroid) of the two coordinate configurations.

$$\mathbf{a}_0 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{a}_i, \quad \mathbf{b}_0 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{b}_i. \tag{2.29}$$

Finally, this algorithm will output all the transformation matrix $\mathbf{R}, \mathbf{S}, \mathbf{T}$. We selected this modified ABC Procrustean Algorithm because it can determine the specific scale parameters in each axial direction (Wang et al., 2021)[3].

As in Wang et al. (2021)[3], in order to search the correspondence between the landmarks in reference and floating maps, we use the squared sum of the distance between the query landmarks and the converted candidates. As mentioned at the beginning of this section, we must include a constraint term to deal with the problem of ill-posed registration to satisfy the inverse consistency. The mechanism can be expressed in formal terms as follows:

$$||\hat{\mathbf{p}}^R - T_\alpha(\hat{\mathbf{p}}^Y)||^2 + ||\hat{\mathbf{p}}^Y - T_\alpha^{-1}(\hat{\mathbf{p}}^R)||^2 \leq 2d \tag{2.30}$$

subject to $S(\hat{\mathbf{p}}^Y - T_\alpha(\hat{\mathbf{p}}^Y)) + S(\hat{\mathbf{p}}^R - T_\alpha^{-1}(\hat{\mathbf{p}}^R)) < \alpha$, where $T_\alpha$ is the transformation operator which has been introduced in Section 2.3. We should note that the activation maps are discrete realizations of continuous maps. Hence, it is proposed to choose the distance less than the threshold $d$ rather than minimizing the distance between landmarks. This is because the unit of the activation map is a voxel, not a single point. In addition, because the points within a pixel are not unique, a landmark's actual location may be anywhere inside a pixel. The discreteness may lead to a bias in the coordinates of landmarks. Hence, as in Wang et al. (2021)[3], we relax the landmark within a voxel using $d = \hat{N}_R$. According to the objective function (2.30), we create a set of correspondence candidates, and based on the intensity loss, we select the optimal correspondence.

The constraint $S$ above is motivated by diffusion registration (Fischer and Modersitzki, 2003)[36]. Fischer and Modersitzki chose the following diffusive regularize. This type of regularize has been applied to affine models by Chumchob and Chen (2009)[37]. As in Wang et al. (2021)[3], it can be stated by defining a function

$D : \mathbb{R}^2 \to \mathbb{R}^2$ with $D(\mathbf{p}) = (D_1(\mathbf{p}), D_2(\mathbf{p}))^T = \mathbf{p} - T(\mathbf{p})$, then

$$S(\mathbf{p} - T(\mathbf{p})) = \frac{1}{2}\int (|\nabla D_1(\mathbf{p})|^2 + |\nabla D_2(\mathbf{p})|^2) d\mathbf{p} = \|\mathbf{I} - \mathbf{A}\|_F^2, \tag{2.31}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\mathbf{I}$ is the identity matrix, $\mathbf{A}$ is defined in (2.14). The restriction specified in (2.31) is also used to determine the orientation of $\hat{\mathbf{p}}^Y$. For instance, in the case of $\hat{N}^R = 3$, $\hat{\mathbf{p}}^Y = \{s_1^Y, s_2^Y, s_3^Y\}$, there are six potential correspondences to $\hat{\mathbf{p}}^R$, and the transformation could be calculated for each of them. Mismatched correspondence may lead to excessive transformation, which would cause repetitive rotation and scaling. Hence, the one minimizing $S(\hat{\mathbf{p}}^Y - T_\alpha(\hat{\mathbf{p}}^Y))$ is used to determine the orientation of $\hat{\mathbf{p}}^R$.

The Photometric Error Criterion (PEC) (Brunet et al., 2010)[38] is used to automatically select the threshold $\alpha$ in (2.30). For each $\alpha$, the optimization problem in (2.30) would be implemented to determine the corresponding features and the appropriate transformation. The criterion fully takes advantage of intensity information to evaluate the hyper-parameter $\alpha$, which is stated as follows:

$$\mathcal{C}(\alpha) = \frac{1}{|\mathcal{G}|}\sum_{\mathbf{p}\in\mathcal{G}} ||\mathbf{R}(\mathbf{p}) - b\mathbf{Y}(T_\alpha(\mathbf{p}))||^2 + \frac{1}{|\mathcal{G}'|}\sum_{\mathbf{p}\in\mathcal{G}'} ||\mathbf{Y}(\mathbf{p}) - b'\mathbf{R}(T_\alpha^{-1}(\mathbf{p}))||^2. \tag{2.32}$$

Here, $\mathcal{G}$ and $\mathcal{G}'$ are the regions of interest in $\mathbf{R}$ and $\mathbf{Y}$ respectively, $|\cdot|$ means the sizes of the region. We should note that (2.32) is the summation of the mean squared error of regressing $\mathbf{R}$ against $\mathbf{Y}(T_\alpha)$ and $\mathbf{Y}$ against $\mathbf{R}(T_\alpha^{-1})$. The scaling factor $b$ and $b'$ correct the scale difference between the target and reference maps. In addition, $b$ can be evaluated as the regression coefficients of $\mathbf{R}(\mathbf{p})$ against $\mathbf{Y}(T_\alpha(\mathbf{p}))$, $b'$ can be evaluated as the regression coefficients of $\mathbf{Y}(\mathbf{p})$ against $\mathbf{R}(T_\alpha^{-1})(\mathbf{p})$ without an

intercept respectively. The estimators of regression coefficients are denoted as $b_0$ and $b_0'$, and they can be used as a hyper-parameter in the prior of $b$ and $b'$ in Section 2.5.

By assuming a heavy-tailed distribution on the residuals, such as the Student **T** distribution with a small degree of freedom, we can further construct the robust Photometric Error Criteria (RPEC). Using RPEC, we can successfully select the proper threshold $\alpha$. Finally, the chosen $\alpha$ can minimize the criterion (2.32) and create the feature correspondence. At the same time, the selected $\alpha$ also provides the estimation of transformation as $M_0$ and $M_0'$. The estimators of $M_0$ and $M_0'$ will be used as the hyper-parameters in the prior for $M$ and $M'$, respectively, defined in Section 2.4. We can then summarize the full steps of prior estimation in the algorithm in Figure 2.2.

**Algorithm 1:** Corresponding Feature-based Prior (hyper-parameter) Estimation

Input: reference image $\mathbf{R}$, reference landmarks $\mathbf{p}^R$, floating image $\mathbf{Y}$, floating landmarks $\mathbf{p}^Y$, bound for distance between landmarks $d$, bound for tuning parameter $\alpha^{max}$.

Construct the set $\mathcal{S}^p = \{\hat{\mathbf{p}}^Y \in \mathbf{p}^Y : |\hat{\mathbf{p}}^Y| = |\mathbf{p}^R|\}$,

Construct the corresponding transformation set $\mathcal{S}^T = \{\hat{T} : \hat{T} = \text{argmin}(\|\mathbf{p}^R - T(\hat{\mathbf{p}}^Y)\|), \text{for } \hat{\mathbf{p}}^Y \in \mathcal{S}^p\}$.

**for** *each* $\alpha \in (0, \alpha^{max})$ **do**

  Find the subset $\hat{\mathcal{S}}_\alpha^p \subseteq \mathcal{S}^p$, such that $S(\hat{\mathbf{p}}^Y - T_\alpha(\hat{\mathbf{p}}^Y)) + S(\hat{\mathbf{p}}^R - T_\alpha^{-1}(\hat{\mathbf{p}}^R)) < \alpha$ and $||\hat{\mathbf{p}}^R - T_\alpha(\hat{\mathbf{p}}^Y)||^2 + ||\hat{\mathbf{p}}^Y - T_\alpha^{-1}(\hat{\mathbf{p}}^R)||^2 \leq 2d\}$

  Collect the corresponding transformation subset $\hat{\mathcal{S}}_\alpha^T \subseteq \mathcal{S}^T$,

  Compute the metric of $\alpha$, $\mathcal{C}(\alpha)$ defined in (2.32).

Output: choose $\alpha$ that minimizes $\mathcal{C}(\alpha)$. The corresponding $T_\alpha$ is the proposed prior.

Figure 2.2: Corresponding Feature-based Prior (hyper-parameter) Estimation

## 2.6 Choice of Reference Map.

Selecting a suitable reference map is essential while performing image registration. Using the group mean map as a reference across all floating maps is an option. However, this method usually makes the reference map blurry and the estimated transformation much variant. The reference subject of our application is selected using scientific information. For instance, researchers may select the functional map that best represents the areas of interest as the reference map. We concentrate on estimating the transformation concerning discrete regions of interest rather than aligning the full functional map. This specifically selected reference map prevents other nuisance regions without a signal of interest from overly influencing estimation.

## 2.7 Model Fitting.

The draws from the posterior distribution are used to infer the parameters. Under the default settings, we apply the Markov Chain Monte Carlo (MCMC) algorithm in R-Stan software (Carpenter et al., 2017)[21] to simulate the posterior distributions. R-Stan is implemented with No-U-Turn sampler (NUTS) (Hoffman and Gelman, 2014)[39]. We refer readers interested in the algorithm's details to the original work in Carpenter et al. 2017[21]. In addition, three different chains are fitted and initialized at values close to the prior estimates to speed up the model fitting.

Only the regularization-controlling hyperparameters $\lambda_b, \lambda_T, \lambda_{b'}$ and $\lambda_{T_{rv}}$ are not

determined in Section 2.5. Cross-validation is a kind of method frequently used to select such hyper-parameters. However, using this approach incorporated into MCMC runs is computationally intensive and needs a substantial amount of parallel computation processing to cover all potential splits of the data. Pareto smoothed importance sampling (PSIS-LOO), suggested by Vehtari et al. (2017)[40], is more simple to compute as a method for approximating leave-one-out cross-validation. Furthermore, compared to Watanabe-Akaike information (WAIC; Watanabe (2010))[41], PSIS-LOO is more robust in the finite case with weak priors or influential observations.

The Kriging interpolation, discussed in Section 2.2, is the primary source of computational complexity. The order of computational complexity is $O(V^3)$. Due to the relatively small $V$ in our experiment, it is computationally manageable. Nevertheless, for a large $V$, the computation may be intractable. It is parallelizable across subjects because we register each map independently. Parallel fitting is done on the models using different regularization values. The convergence is assessed using internal diagnostics provided by R-STAN, such as the Gelman-Rubin potential scale reduction factor. The full software implementation of our method is based on https : //github.com/gqwang1001/BayesianFunctionalRegistrationOfFMRIMaps.

# Chapter 3

# Practical Application and Analysis of Real Dataset

## 3.1 Introduction of Data Set

The data are from an fMRI study of the correspondence of categorical and feature-based representations of music in the human brain; see Nakai et al. (2021)[22] for an in-depth discussion. In the MRI and behavioural studies, five healthy participants with normal hearing (ID01-05; age range 23–33; 2 females; music experience, 4–15 years) complete the study. Each participant provided informed consent before participating in the study. This experiment is approved by the National Institute of Information and Communications Technology ethics and safety committee in Osaka, Japan.

There are 18 runs conducted for each experiment, 12 of which are regarded as

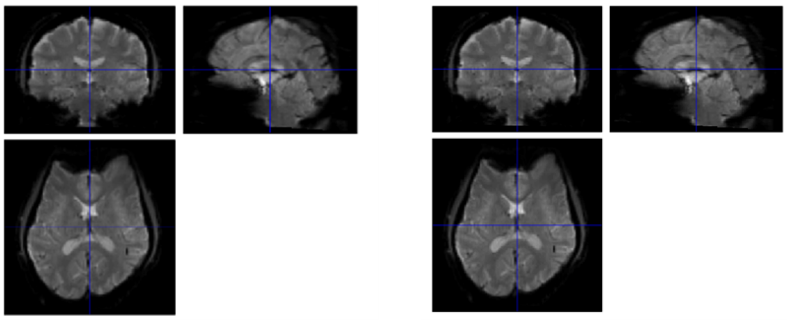
training runs and 6 as test runs. During our analysis, the runs are concatenated and analyzed together. There are 480 music clips used in the training runs, while 60 clips are used in the test runs. A set of 10 music clips is presented four times in the same order during each test run to ensure data repeatability. The order of the clips is randomized throughout the experiment. The participants are instructed to listen to the music clips using MRI-compatible insert headphones while focusing on a fixation cross (Model S14, Sensimetrics) in the center of the screen. This model has been widely applied in earlier MRI research using auditory stimuli and can reduce scanner noise. The duration of all stimuli is 15s. For each clip, 2s of fade-in and fade-out effects are applied, and the overall signal intensity is normalized in terms of the root mean square. For the training runs, the first stimulus (0-15s) is the same as the last stimulus of the previous run (600-615s). For the test runs, the first stimulus (0-15s) is the same as the last stimulus of the same run (600-615s). The original music stimuli (GTZAN dataset) can be found here: http : //marsyas.info/downloads/datasets.html. According to their self-reports, nobody slept during the experiments. The experiment is executed for three days, with six runs performed each day.

Scanning is performed using a 3.0 T MRI scanner (TIM Trio; Siemens, Erlangen, Germany) equipped with a 32-channel head coil. For functional scanning, they scan 68 interleaved axial slices with a thickness of 2.0 mm without a gap using a T2*-weighted gradient echo multi-band echo-planar imaging (MB-EPI) sequence (Moeller et al., 2010) (repetition time (TR) = 1,500 ms, echo time (TE) = 30 ms, flip angle (FA) = 62°, field of view (FOV) = $192 \times 192\ mm^2$ , voxel size = $2 \times 2 \times 2\ mm^3$, multi-

band factor = 4). A total of 410 volumes are obtained for each run. For anatomical reference, we acquire high-resolution T1-weighted images of the whole brain from all participants using a magnetization-prepared rapid acquisition gradient echo sequence (MPRAGE, TR = 2,530 ms, TE = 3.26 ms, FA = 9°, FOV = 256 × 256 $mm^2$ , voxel size = 1 × 1 × 1 $mm^3$ ).

## 3.2 Data preprocessing

We apply **SPM8** to preprocess the data and take subject1 for example to describe all the steps. The functional images are first corrected differences in image acquisition time between slices using slice-timing. Below are the functional maps pre and post slice-timing.



(a) Original Functional Map

(b) Functional Map after Slice-timing

Figure 3.1: Functional Images Pre and Post Slice-Timing.

After slice-timing, we realign (motion correct) the functional images. This routine removes movement artifacts in fMRI and PET time series (or, more generally,

longitudinal studies) (J. Ashburner et al.,1997)[23]. The transformation details are displayed below as plots of translation and rotation.

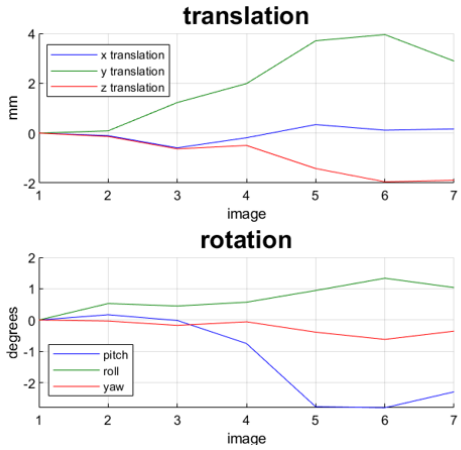



Figure 3.2: Details of Transformation

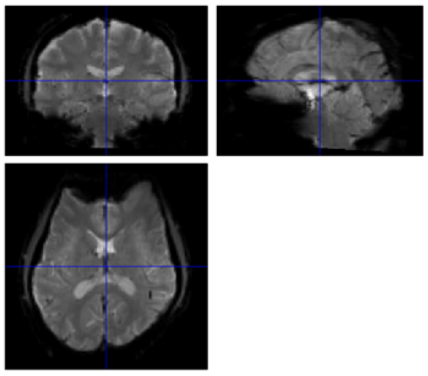


Figure 3.3: Functional Map after Realign

Then the structural images are co-registered to the mean functional image using the iterative mutual information-based algorithm. At the end of co-registration, the voxel-to-voxel affine transformation matrix and histograms for the images in the original and final orientations are displayed. The co-registered images are displayed in Figure 3.4.

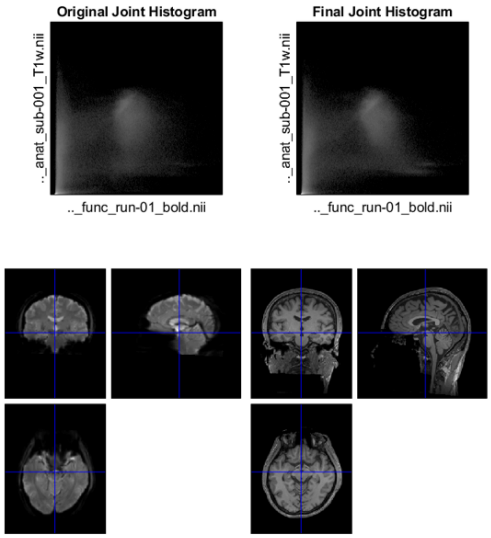



Figure 3.4: Normalised Mutual Information Coregistration

To match the subject's structural image to the functional image, it is necessary to segment the structural image. The structural image is generally segmented into gray matter, white matter and cerebrospinal fluid. Thereafter, images are normalized to Montreal Neurological Institute (MNI) space using SPM8's generative segment-and-normalize algorithm, registered images are displayed below:

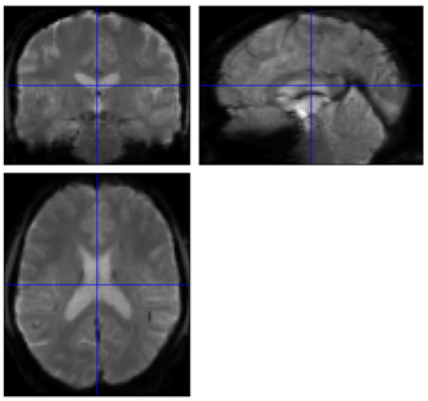


Figure 3.5: Functional Images after Normalization

Finally, these images are warped to SPM's normative atlas using warping parameters estimated from co-registered high-resolution structural images and smoothed with an 8mm full width at half maximum (FWHM) Gaussian kernel. A high-pass filter of 180s is applied to the time series data, and the result is given in Figure 3.6. This step suppresses the noise and effects of residual differences in functional and gyral anatomy during inter-subject averaging.

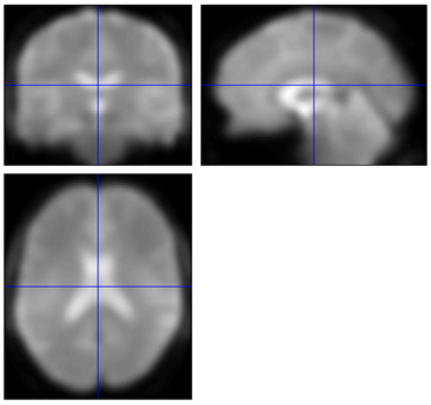


Figure 3.6: Functional Images after Smooth

A voxel-wise general linear model (GLM) analysis is performed for each subject using SPM8. For each music genre, boxcar regressors convolved with the canonical hemodynamic response function (Lindquist et al., 2009)[24] are constructed to model periods corresponding to the 15s music stimulation. Other regressors of non-interest (i.e., nuisance variables) include (a) a run-specific intercept; (b) linear drift across time within each run; (c) the six estimated head movement parameters (x, y, z, roll, pitch, and yaw), their mean-centred squares, their derivatives, and squared derivative for each run; (d) indicator vectors for outlier time points identified based on their multivariate distance from the other images in the sample; (e) indicator vectors for

the first two images in each run; (f) signals from white matter and ventricle (Wang et al., 2021). For each music genre, the estimated regression coefficients corresponding to that genre-specific regressor at each voxel are combined into a single activation map. Hence, for each subject, we have ten different brain maps depicting the functional response across the brain to each of the music genres. The GLM analysis result is shown below:

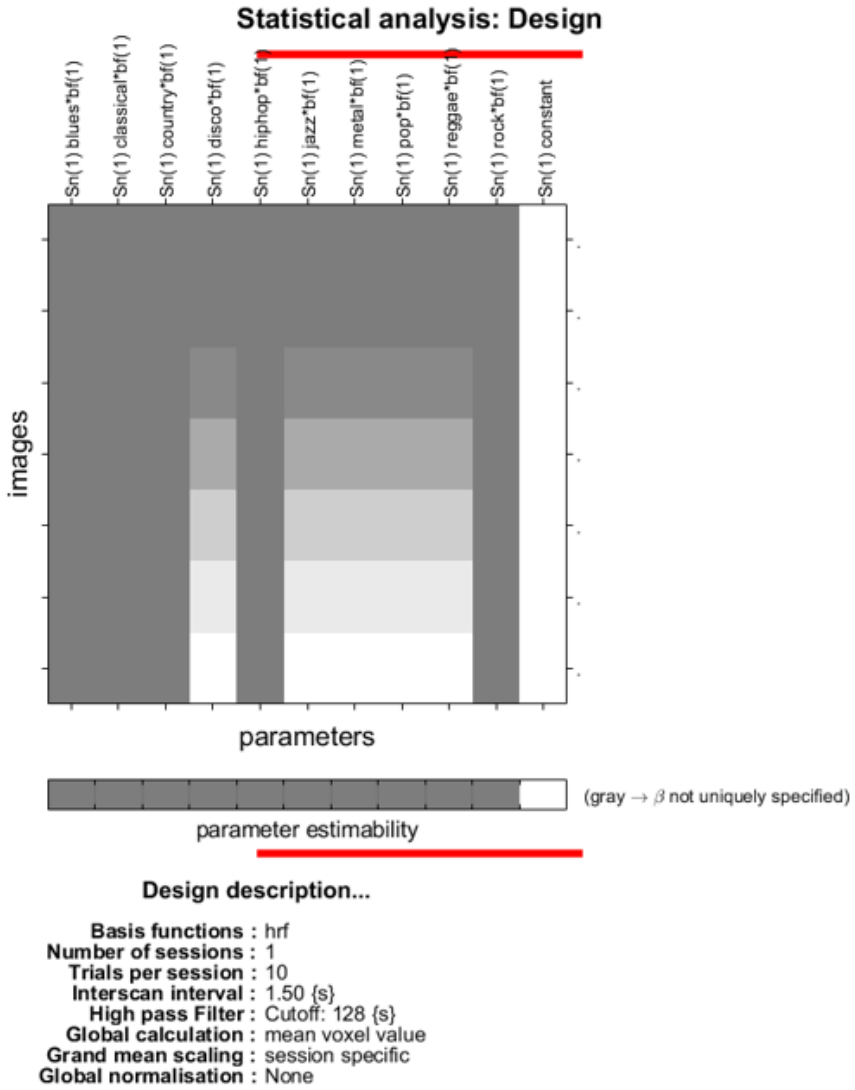



Figure 3.7: GLM Analysis Result

This work primarily focuses on the subject-specific mean across the ten different brain maps to analyze the inter-subject variability. We use this data to perform the functional alignment. In particular, we focus the analysis on the brain's superior temporal gyrus region. According to Alluri et al. (2012)[42], a strong association exists between timbre, harmony, and rhythmic characteristics and bilateral superior temporal gyrus (STG) activity. Additionally, Toiviainen et al. (2014)[43] found that the bilateral STG is involved in decoding timbral characteristics. Figure 3.8 shows a square region with sides of 30 voxels across subjects, and there exists substantial inter-subject variability in their exact location. The superior temporal gyrus region of the brain is explicitly chosen based on the coordinates offered by meta-analyses of neuroimaging music studies (see, for example, NeuroSynth (Yarkoni et al., 2011)[44]).

## 3.3 Analysis of Real Dataset

To perform functional alignment, we use the data prepossessed in Section 3.2. Figure 3.8 below shows the region of interest of functional activation maps for all 5 participants.

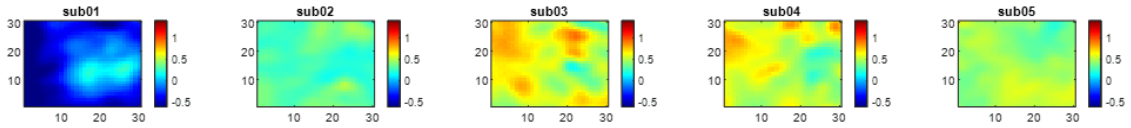



Figure 3.8: Subject-specific activation maps of the region of interest across the 5 subjects. Note the differences in the location of peaks across subjects, indicating significant inter-subject differences.

As in Wang et al. (2021)[3], these functional activation maps are then processed using Bradley and Roth's adaptive thresholding (2007)[45]. The regional maxima of the neighbourhood of eight nearby locations are then determined as landmarks which are shown in Figure 3.9.

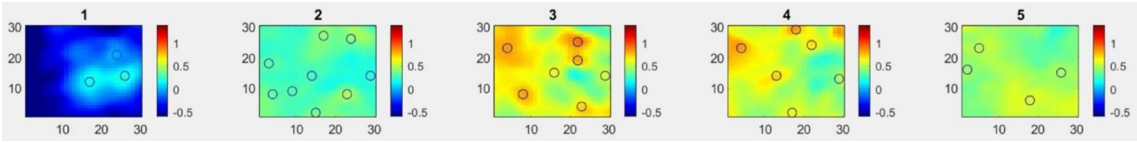
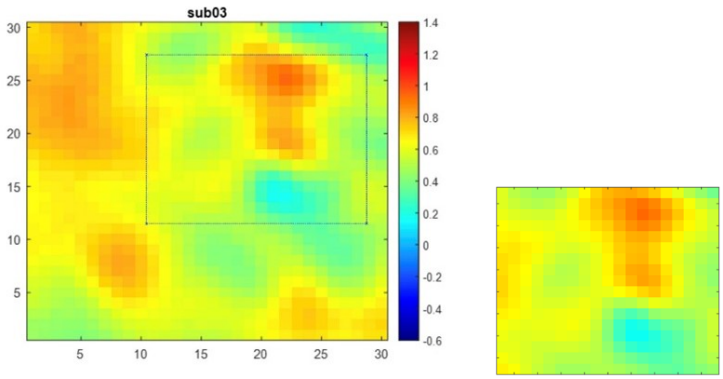



Figure 3.9: The features (local peaks) are circled.

The left panel of Figure 3.10(a) shows the reference functional activation map based on Section 2.6. The query part of the map is manually cropped and shown within the dashed rectangle and in Panel (b).




(a) Activation Map (b) Query Map

Figure 3.10: (a) An example of a functional activation map. A query map indicating landmarks of interest is manually determined and shown within the dashed rectangle when used as a reference map. (b) The query map extracted from the reference map.

The subject-specific maps are shown in Figure 3.11, and the estimated correspond-

ing areas based on prior information obtained from Section 2.5 are indicated using the dashed boxes.

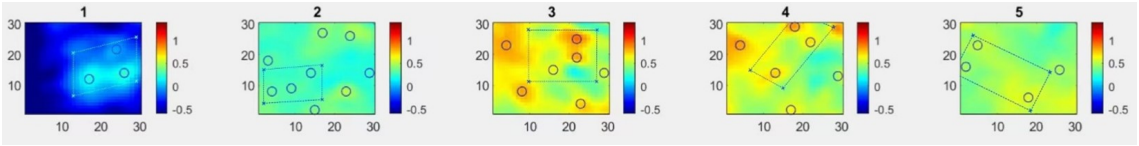



Figure 3.11: The corresponding cropped area (bounding box) based on prior information in each subject-specific map. The features (local peaks) are circled.

Figure 3.12 shows the matching area warped by the posterior mean of transformation, along with the 95% credible region of the transformations superimposed in the gray-shaded region.

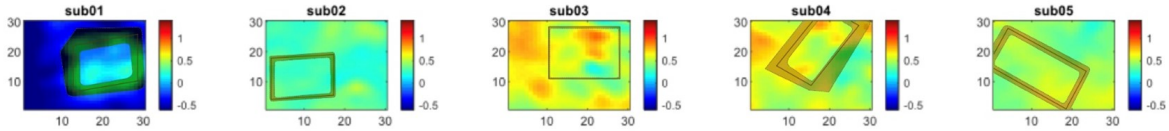



Figure 3.12: The warped bounding box using the transformation computed with the posterior mean is shown in the solid line. The 95% credible interval of the transformation map is illustrated in the shaded area.

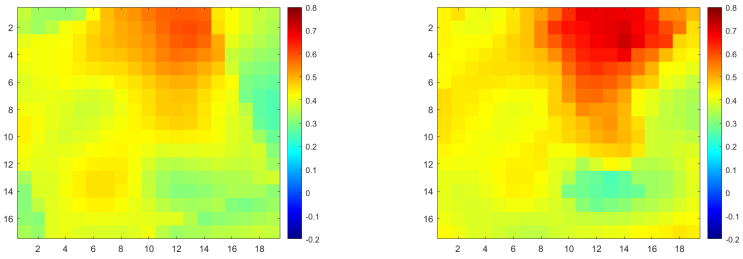
The density-based clustering algorithm (DBSCAN), developed by Ester et al.(1996)[46], is used to identify the credible region of transformation (Wang et al., 2021)[3]. The DBSCAN algorithm is an unsupervised clustering method that divides clusters according to their densities. We specify the epsilon region with at least five samples to apply DBSCAN and use a grid search of the size of the epsilon neighbourhood to ensure that the only cluster captures approximately 95% of the posterior samples.

Each posterior sample of the transformation then warps the searching box and generates the shaded area in Figure 3.12. The estimated credible regions can be used to display the inter-subject variation in registration. For instance, subject 01 exhibits a high degree of uncertainty, while subject 02 exhibits relatively low uncertainty.

We contrast the results to those obtained by the general standard method using Registration Estimator in the image processing toolbox in MATLAB (version R2021a). After that, we conduct a second-level (across-subjects) study to test for significant music-related effects to evaluate the performance of the functional alignment process. Results are displayed in Figure 3.13. Note that all methods should first perform an initial anatomical registration into a standard anatomical space which has been discussed in Section 3.2.

The proposed method shows increased similarity in the activation profiles across subjects by studying the voxel-wise group means, standard deviations and t-statistics from the two methods. In addition, from the standard deviation of (c) and (d) in Figure 3.13, the lighter colour in (d) means by using the proposed method, the variation between subjects is decreased. At the same time, we can find that (d) is more concentrated on the regions of interest compared with (c). It means while the locations of the peaks are now consistent across subjects, the inter-subject differences in intensity remain. This is important for testing for brain-behaviour correlations across subjects because we want to increase the similarity between different subjects but keep the large difference within each subject. The difference within a subject usually relates to different brain functions.

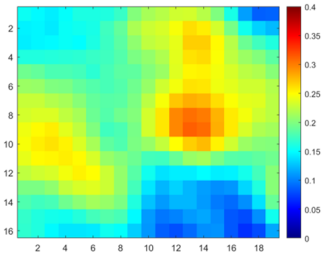
In general, the specific part of the brain shows more sensitivity (higher voxel value) when the stimuli it receives are related to the function of that part. Hence, the deeper colour in (f) compared with (e) represents the proposed method has higher t-statistic values than the standard method, indicating increased sensitivity for group-level inference. The full t-statistic values are shown in Appendix B. The standard registration method results show a similar pattern to the proposed approach. However, the t-statistic is significantly lower, indicating a lower sensitivity for group-level inference. Together, the results show the benefits of the proposed functional registration approach.



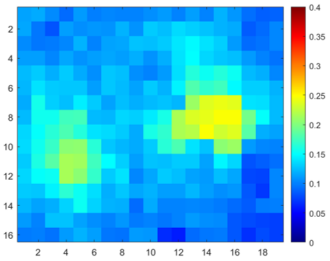
(a) Mean-Standard



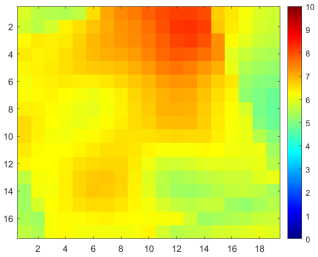
(b) Mean-Proposed



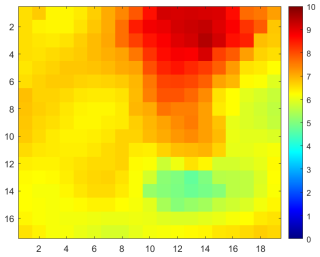
(c) SD-Standard



(d) SD-Proposed



(e) T Statistics-Standard



(f) T Statistics-Proposed

Figure 3.13: A comparison between the standard approach (left column) and our proposed approach (right column). Group-level statistics (t-statistics, group means, group standard deviations) are calculated for both.

# Chapter 4

# Discussion

In this study, we apply a Bayesian framework based on a methodology developed by Wang et al. (2021)[3] to minimize misalignment between people in functional brain systems. The generalized Bayes framework combines feature-based information from the prior estimation with intensity-based information from the loss function to enable inference on the transformation using the posterior samples. When using fMRI data from a study of brain activity under discrete music stimuli, the proposed registration method increases the sensitivity for group-level inference.

However, a major shortcoming of the proposed approach is that we need to pre-specify the template map, which may generate potential bias estimation. In addition, the cropping box selection can impact the performance of registration. For instance, the performance of the proposed methodology may degrade because of the outlier region when the cropping box covers redundant space (Wang et al., 2021)[3]. Finally, in this work, we have implemented the proposed method in the two-dimensional brain

map. We start with the 2D version because it enables us to give a simplified explanation of the methodology. However, a three-dimensional version might be created similarly. Hence, this proposed method has great potential to be applied to three-dimensional brain maps. Furthermore, implementing the proposed approach in 3D brain maps will have more practical meaning than the 2D version.

In future work, we will explore how functional registration can help us to study based on much more subjects rather than single people. Specifically, we will examine how functional brain representation can be used to predict behaviour, performance, clinical status, and prognosis more accurately. One common problem within this procedure is functional misalignment. Hence, the features used in predictive models need to be aligned as precisely as possible to avoid impacting the accuracy of prediction. Comparing the two approaches mentioned in Chapter 3, we find that the proposed method can better correct this shortcoming than the standard method as an initial feature alignment step. Using the proposed approach as a preliminary feature alignment step will potentially rectify the shortcoming mentioned above. At the same time, the proposed approach allows for population-level inference to improve the efficiency of relevant studies and expand the reference base of research subjects. These are all the potential benefits of the proposed Bayesian functional registration.

# Bibliography

[1] Huettel, S. A., Song, A. W., and McCarthy, G. (2009). Functional Magnetic Resonance Imaging, Sunderland, MA: Sinaur Associates. Inc., January, 16.

[2] Rinck, P. A. (2019). Magnetic resonance in medicine: a critical introduction. BoD–Books on Demand.

[3] Wang, G., Datta, A., and Lindquist, M. A. (2022). Bayesian functional registration of fMRI activation maps. The Annals of Applied Statistics, 16(3), 1676-1699.

[4] Lindquist, M. A. et al. (2008). The statistical analysis of fmri data. Statistical science, 23(4):439–464.

[5] Ombao, H., Lindquist, M., Thompson, W., Aston, J. (2016). Handbook of neuroimaging data analysis. Chapman and Hall/CRC.

[6] Rademacher, J., Caviness Jr, V. S., Steinmetz, H., Galaburda, A. M. (1993). Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cerebral Cortex, 3(4), 313-329.

[7] Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., and Zilles, K. (2000).

Brodmann's areas 17 and 18 brought into stereotaxic space—where and how variable?. Neuroimage, 11(1), 66-84.

[8] Thompson, P. M., Schwartz, C., Lin, R. T., Khan, A. A., and Toga, A. W. (1996). Three-dimensional statistical analysis of sulcal variability in the human brain. Journal of Neuroscience, 16(13), 4261-4274.

[9] Vogt, B. A., Nimchinsky, E. A., Vogt, L. J., and Hof, P. R. (1995). Human cingulate cortex: surface features, flat maps, and cytoarchitecture. Journal of Comparative Neurology, 359(3), 490-506.

[10] Duncan, K. J., Pattamadilok, C., Knierim, I., and Devlin, J. T. (2009). Consistency and variability in functional localisers. Neuroimage, 46(4), 1018-1026.

[11] Iordan, M. C., Joulin, A., Beck, D. M., and Fei-Fei, L. (2016). Locally-optimized inter-subject alignment of functional cortical regions. arXiv preprint arXiv:1606.02349.

[12] Allison, T., Puce, A., Spencer, D. D., and McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cerebral cortex, 9(5), 415-430.

[13] McCarthy, G., Puce, A., Belger, A., and Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. Cerebral cortex, 9(5), 431-444.

[14] Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B.

R., Gobbini, M. I., ... and Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron, 72(2), 404-416.

[15] Sabuncu, M. R., Singer, B. D., Conroy, B., Bryan, R. E., Ramadge, P. J., and Haxby, J. V. (2010). Function-based intersubject alignment of human cortical anatomy. Cerebral cortex, 20(1), 130-140.

[16] Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. science, 303(5664), 1634-1640.

[17] Conroy, B. R., Singer, B. D., Guntupalli, J. S., Ramadge, P. J., and Haxby, J. V. (2013). Inter-subject alignment of human cortical anatomy using functional connectivity. NeuroImage, 81, 400-411.

[18] Nenning, K. H., Liu, H., Ghosh, S. S., Sabuncu, M. R., Schwartz, E., and Langs, G. (2017). Diffeomorphic functional brain surface alignment: Functional demons. NeuroImage, 156, 456-465.

[19] Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). Hierarchical modeling and analysis for spatial data. CRC press.

[20] Cressie, N., and Wikle, C. K. (2015). Statistics for spatio-temporal data. John Wiley and Sons.

[21] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt,

M., ... and Riddell, A. (2017). Stan: A probabilistic programming language. Journal of statistical software, 76(1).

[22] Nakai, T., Koide-Majima, N., and Nishimoto, S. (2021). Correspondence of categorical and feature-based representations of music in the human brain. Brain and Behavior, 11(1), e01936.

[23] Ashburner, J., and Friston, K. (1997). The role of registration and spatial normalisation in detecting activations in functional imaging. CLINICAL MRI, 7(1), 26-27.

[24] Lindquist, M. A., Loh, J. M., Atlas, L. Y., and Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. Neuroimage, 45(1), S187-S198.

[25] Chernozhukov, V., and Hong, H. (2003). An MCMC approach to classical estimation. Journal of econometrics, 115(2), 293-346.

[26] Christensen, G. E., and Johnson, H. J. (2001). Consistent image registration. IEEE transactions on medical imaging, 20(7), 568-582.

[27] Johnson, H. J., and Christensen, G. E. (2002). Consistent landmark and intensity-based image registration. IEEE transactions on medical imaging, 21(5), 450-461.

[28] Chung, S. Y., Venkatramanan, S., Elzain, H. E., Selvam, S., and Prasanna, M. V. (2019). Supplement of missing data in groundwater-level variations of peak type

using geostatistical methods. GIS and geostatistical techniques for groundwater science, 33-41.

[29] Rodolphe Le Riche. Introduction to kriging. Doctoral. France. 2014. ffcel-01081304f

[30] Cederberg, J. (2004). A course in modern geometries. Springer Science and Business Media.

[31] Fischer, B., and Modersitzki, J. (2008). Ill-posed medicine—an introduction to image registration. Inverse problems, 24(3), 034008.

[32] Mian, A. S., Bennamoun, M., and Owens, R. (2006). Three-dimensional model-based object recognition and segmentation in cluttered scenes. IEEE transactions on pattern analysis and machine intelligence, 28(10), 1584-1601.

[33] Chen, Y., and Medioni, G. (1992). Object modelling by registration of multiple range images. Image and vision computing, 10(3), 145-155.

[34] Besl, P. J., and McKay, N. D. (1992, April). Method for registration of 3-D shapes. In Sensor fusion IV: control paradigms and data structures (Vol. 1611, pp. 586-606). Spie.

[35] Awange, J. L., Bae, K. H., and Claessens, S. J. (2008). Procrustean solution of the 9-parameter transformation problem. Earth, planets and space, 60, 529-537.

[36] Fischer, B., and Modersitzki, J. (2003). Curvature based image registration. Journal of Mathematical Imaging and Vision, 18, 81-85.

[37] Chumchob, N., and Chen, K. (2009). A ROBUST AFFINE IMAGE REGISTRATION METHOD. International Journal of Numerical Analysis and Modeling, 6(2).

[38] Brunet, F., Bartoli, A., Navab, N., and Malgouyres, R. (2010, November). Pixel-based hyperparameter selection for feature-based image registration. In International Workshop on Vision, Modeling and Visualization (pp. 33-40).

[39] Hoffman, M. D., and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res., 15(1), 1593-1623.

[40] Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing, 27, 1413-1432.

[41] Watanabe, S., and Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of machine learning research, 11(12).

[42] Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. NeuroImage, 59(4), 3677–3689.

[43] Toiviainen, P., Alluri, V., Brattico, E., Wallentin, M., and Vuust, P. (2014).

Capturing the musical brain with Lasso: Dynamic decoding of musical features from fMRI data. NeuroImage, 88, 170–180.

[44] Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. Nature methods, 8(8), 665–670.

[45] Bradley, D., and Roth, G. (2007). Adaptive thresholding using the integral image. Journal of graphics tools, 12(2), 13-21.

[46] Ester, M., Kriegel, H. P., Sander, J., and Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd (Vol. 96, No. 34, pp. 226-231).

# Appendix A: Definition of Stationary

**Definition 1.** The process is said to be strictly stationary (sometimes strong stationarity) if, for any given $n \geq 1$, any set of $n$ sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ and any $\mathbf{h} \in \Re^r$, the distribution of $(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))$ is the same as that of $(Y(\mathbf{s}_1 + \mathbf{h}), \ldots, Y(\mathbf{s}_n + \mathbf{h}))$ (Banerjee et al., 2014)[19].

**Definition 2.** A spatial process is called weakly stationary if $\mu(\mathbf{s}) \equiv \mu$ (i.e., it has a constant mean) and $\mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$ for all $\mathbf{h} \in \Re^r$(Banerjee et al., 2014)[19].

# Appendix B: T-statistic

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5.8774 | 5.7594 | 6.2592 | 6.4379 | 6.3001 | 6.1662 | 6.199 | 6.3699 | 6.5806 | 6.5892 | 6.4104 | 6.3176 | 5.6109 | 5.3059 | 5.2904 | 5.3932 | 5.6033 |
| 5.5613 | 5.5587 | 6.4204 | 6.3592 | 6.3454 | 6.3111 | 6.2568 | 6.344 | 6.4018 | 6.3576 | 6.2796 | 6.2243 | 6.1655 | 6.1239 | 5.8517 | 5.3069 | 5.4521 |
| 5.5012 | 5.7993 | 6.3854 | 6.405 | 6.4115 | 6.3767 | 6.2906 | 6.2623 | 6.2297 | 6.1659 | 6.2149 | 6.2368 | 6.1859 | 6.1681 | 6.2164 | 6.2042 | 6.1955 |
| 5.4621 | 6.3441 | 6.4379 | 6.5268 | 6.5294 | 6.436 | 6.2425 | 6.1485 | 6.1261 | 6.1091 | 6.2411 | 6.3607 | 6.3776 | 6.3893 | 6.3352 | 6.2741 | 6.2794 |
| 5.606 | 6.5426 | 6.6671 | 6.6926 | 6.6447 | 6.4005 | 6.1448 | 6.0166 | 6.047 | 6.1719 | 6.3063 | 6.4915 | 6.6079 | 6.6022 | 6.4902 | 6.2827 | 6.1553 |
| 6.5324 | 6.8189 | 6.9379 | 6.8933 | 6.724 | 6.3695 | 6.0808 | 5.9997 | 6.1218 | 6.2623 | 6.4253 | 6.5971 | 6.7353 | 6.7719 | 6.5927 | 6.3218 | 6.1187 |
| 7.0888 | 7.06 | 7.0937 | 7.0032 | 6.7869 | 6.4565 | 6.1789 | 6.1377 | 6.2791 | 6.4039 | 6.5251 | 6.6182 | 6.7585 | 6.7425 | 6.5884 | 6.329 | 6.1139 |
| 7.3788 | 7.2869 | 7.253 | 7.0836 | 6.8488 | 6.6218 | 6.4016 | 6.3397 | 6.3999 | 6.486 | 6.5572 | 6.6375 | 6.6697 | 6.6122 | 6.4521 | 6.2759 | 6.1011 |
| 7.5294 | 7.4152 | 7.3725 | 7.2221 | 7.026 | 6.7692 | 6.6237 | 6.5658 | 6.5676 | 6.5219 | 6.4464 | 6.4163 | 6.3901 | 6.366 | 6.304 | 6.2479 | 6.1632 |
| 7.793 | 7.6551 | 7.5529 | 7.4336 | 7.2183 | 6.9522 | 6.8042 | 6.7481 | 6.6717 | 6.5142 | 6.3063 | 6.0905 | 5.9099 | 5.942 | 6.1027 | 6.1906 | 6.308 |
| 8.0071 | 7.9594 | 7.8124 | 7.6964 | 7.4928 | 7.1629 | 6.9805 | 6.98 | 6.8428 | 6.5966 | 6.1797 | 5.8085 | 5.5681 | 5.537 | 5.8324 | 6.1718 | 5.8109 |
| 8.093 | 8.2095 | 8.08 | 7.8734 | 7.5903 | 7.2806 | 7.0703 | 7.0066 | 6.9096 | 6.6229 | 6.1962 | 5.7593 | 5.4098 | 5.3155 | 5.639 | 6.1598 | 5.5294 |
| 8.0922 | 8.2497 | 8.1408 | 7.8244 | 7.4699 | 7.2363 | 7.039 | 6.9558 | 6.8833 | 6.5994 | 6.2063 | 5.8377 | 5.4862 | 5.2897 | 5.5644 | 5.7782 | 5.6142 |
| 8.0016 | 8.1246 | 7.9675 | 7.6237 | 7.2451 | 6.977 | 6.8068 | 6.7311 | 6.6804 | 6.5666 | 6.2617 | 5.9405 | 5.5928 | 5.3888 | 5.628 | 5.2803 | 5.6638 |
| 6.7146 | 6.7851 | 7.3179 | 7.3219 | 7.0014 | 6.6855 | 6.5473 | 6.4598 | 6.4145 | 6.3833 | 6.2422 | 5.9952 | 5.6787 | 5.5317 | 5.6977 | 5.3742 | 5.7968 |
| 6.449 | 6.4309 | 6.2578 | 5.9511 | 6.047 | 6.4673 | 6.3277 | 6.2284 | 6.1767 | 6.1691 | 6.1274 | 5.94 | 5.7072 | 5.6602 | 5.2784 | 5.4529 | 5.8319 |
| 6.1438 | 6.1219 | 5.9382 | 5.6905 | 5.4432 | 5.2449 | 5.1413 | 5.8079 | 6.0283 | 6.0323 | 6.0269 | 5.9511 | 5.7312 | 5.6705 | 5.1732 | 5.5423 | 5.8986 |
| 5.8854 | 5.8575 | 5.7196 | 5.5015 | 5.2889 | 5.0739 | 4.9367 | 4.9003 | 5.0074 | 5.5369 | 6.0883 | 6.0245 | 5.9207 | 5.6126 | 5.3613 | 5.6761 | 5.9809 |
| 5.7681 | 5.7147 | 5.5904 | 5.413 | 5.2207 | 5.0041 | 4.7959 | 4.7422 | 4.7935 | 5.0471 | 5.3209 | 5.5455 | 6.0431 | 5.3684 | 5.5158 | 5.7305 | 5.9056 |
| **T-Statistic of Standard Method** | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6.5577 | 6.5594 | 6.5483 | 6.5327 | 6.4686 | 6.5239 | 6.8422 | 6.8117 | 6.7616 | 6.7301 | 6.5636 | 6.4014 | 6.2646 | 6.2445 | 6.2236 | 6.3195 | 6.5106 |
| 6.7371 | 6.3613 | 6.4592 | 6.4865 | 6.5627 | 6.6608 | 6.6713 | 6.6802 | 6.6242 | 6.5677 | 6.4794 | 6.3603 | 6.2278 | 6.1428 | 6.174 | 6.1979 | 6.3284 |
| 6.3071 | 6.2952 | 6.4415 | 6.5734 | 6.6681 | 6.7243 | 6.6663 | 6.6069 | 6.5236 | 6.4305 | 6.4138 | 6.3542 | 6.2506 | 6.1583 | 6.1481 | 6.1332 | 6.165 |
| 6.2995 | 6.3217 | 6.5127 | 6.6734 | 6.7591 | 6.7256 | 6.5759 | 6.4833 | 6.4339 | 6.3476 | 6.3755 | 6.4169 | 6.3693 | 6.3016 | 6.1816 | 6.1245 | 6.1349 |
| 6.4113 | 6.5175 | 6.6889 | 6.7486 | 6.788 | 6.6729 | 6.4795 | 6.3352 | 6.3062 | 6.341 | 6.3951 | 6.5006 | 6.5374 | 6.4484 | 6.2864 | 6.0956 | 6.032 |
| 6.7115 | 6.7729 | 6.8934 | 6.8753 | 6.81 | 6.6369 | 6.4107 | 6.2902 | 6.3262 | 6.3823 | 6.4822 | 6.5985 | 6.6389 | 6.5826 | 6.3627 | 6.082 | 5.9514 |
| 7.0579 | 7.0649 | 7.0788 | 6.9986 | 6.8753 | 6.6607 | 6.4116 | 6.3242 | 6.4019 | 6.4533 | 6.5881 | 6.5896 | 6.642 | 6.5692 | 6.3628 | 6.048 | 5.8584 |
| 7.4066 | 7.6504 | 7.6418 | 7.1434 | 7.0442 | 6.7355 | 6.5108 | 6.4235 | 6.4158 | 6.647 | 6.6684 | 6.6442 | 6.4985 | 6.4148 | 6.2424 | 6.0427 | 5.8566 |
| 7.8692 | 8.4558 | 8.0961 | 7.5675 | 7.1114 | 6.9502 | 6.9075 | 6.9325 | 6.6924 | 6.6604 | 6.4103 | 6.2553 | 6.2046 | 6.1803 | 6.1426 | 6.092 | 6.0001 |
| 8.5584 | 8.8406 | 8.5375 | 8.2047 | 7.9426 | 7.5741 | 7.48 | 7.0315 | 6.9241 | 6.8861 | 6.3937 | 6.1799 | 5.7917 | 5.2109 | 5.9502 | 6.0232 | 6.1018 |
| 9.1474 | 8.9159 | 8.7667 | 8.6177 | 8.3672 | 8.0791 | 7.8951 | 7.5944 | 7.1249 | 7.043 | 6.6595 | 5.7042 | 5.1063 | 5.0579 | 5.354 | 5.9176 | 6.1713 |
| 9.1255 | 9.156 | 9.0436 | 8.7928 | 8.5283 | 8.1209 | 7.9144 | 7.4789 | 7.3598 | 7.2418 | 6.8401 | 6.0878 | 4.9971 | 4.9269 | 5.058 | 5.8949 | 6.1905 |
| 9.1771 | 9.2324 | 9.1412 | 8.7809 | 8.3345 | 8.0327 | 7.709 | 7.4646 | 7.5321 | 7.3644 | 6.913 | 6.509 | 4.7114 | 4.7889 | 5.2279 | 5.8832 | 6.2694 |
| 9.1313 | 9.3317 | 9.48 | 9.1055 | 8.1354 | 7.786 | 7.3498 | 7.1754 | 7.1616 | 7.1802 | 6.9814 | 6.3228 | 4.9758 | 4.9041 | 5.3454 | 5.9413 | 6.3254 |
| 9.0976 | 9.0293 | 9.1532 | 8.0943 | 7.8082 | 7.1249 | 6.5178 | 6.4207 | 6.8042 | 6.8982 | 6.6614 | 5.8027 | 5.5299 | 5.0414 | 5.6507 | 6.0434 | 6.4662 |
| 8.6006 | 8.7609 | 8.5665 | 7.8332 | 7.2076 | 6.9488 | 6.2289 | 6.2193 | 6.0021 | 6.0308 | 6.0119 | 5.7834 | 5.5546 | 5.5585 | 5.7501 | 6.1357 | 6.5135 |
| 7.6602 | 8.2212 | 8.435 | 7.1513 | 6.4907 | 6.2759 | 5.9586 | 5.8917 | 5.8955 | 5.9736 | 5.989 | 5.8662 | 5.6224 | 5.6031 | 5.882 | 6.2478 | 6.5966 |
| 7.5769 | 7.8322 | 6.8735 | 6.6443 | 6.4174 | 6.1853 | 5.8248 | 5.7585 | 5.8171 | 5.9618 | 6.0725 | 6.0101 | 5.8965 | 5.9249 | 6.1087 | 6.4149 | 6.6975 |
| 7.1613 | 6.7778 | 6.7314 | 6.5379 | 6.2361 | 5.9098 | 5.6829 | 5.6082 | 5.6332 | 5.8715 | 6.1409 | 6.2698 | 6.2244 | 6.1762 | 6.3008 | 6.487 | 6.624 |
| **T-Statistic of Proposed Method** | | | | | | | | | | | | | | | | |

Table 1: T-statistic values of Standard Method and Proposed Method

# Vita Auctoris

Xiaoxuan Wang was born in 1998 in Xi'an, China. She graduated from Shaanxi Normal University and obtained a Bachelor's Degree of Science in Mathematics and Applied Mathematics in 2021. She is currently a candidate for the Master's Degree of Science in MSc-Math and Stats-Statistics at the University of Windsor and hopes to graduate in May 2023.