Bayesian Functional Spatial Partitioning Methods for Prostate Cancer Lesion Detection

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Maria Anna Masotti

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Lin Zhang, Joseph S. Koopmeiners

May, 2022

Acknowledgements

I am deeply grateful to my advisor Dr. Lin Zhang, whose ideas inspired the methods in Chapter 2 and beyond. This work would not have been possible without her expertise and commitment to my success. Dr. Joe Koopmeiners recruited me to the University of Minnesota and later became my co-advisor. I am thankful for his support and leadership through the years.

Many thanks to my collaborator Dr. Greg Metzger for his contributions to the data which supports and inspires this work. I am also especially thankful for Dr. Ethan Leng, for answering my questions and responding to my many emails despite being busy pursuing his dual MD-PhD degree. Dr. Jim Hodges has offered insightful ideas and thoughtful edits to my writing.

Thanks to Dr. Rebecca Cogswell, who became a wonderful mentor over the last few years. Her curiosity for uncovering the truth and commitment to good science has been inspiring. Her patience and friendship allowed me to grow into a confident and accomplished collaborator. I am excited for what we will accomplish together.

I would not have survived graduate school without the support of my peers. I thank my PhD cohort for helping me with homework, commiserating with me after tough exams, and always making me feel at home. I am especially grateful to Clara Drew for her unwavering commitment to Monday night episodes of the Bachelor and to Dr. Grace Lyden for her delicious baked goods and expert trivia (and statistical) knowledge. I am thankful for Lily Haine and the students of the SAC for striving to make the division a more inclusive and supportive environment for graduate students. Members of the BCOE continue to inspire me with their commitment to community and service.

The biggest thanks to my family, without whom, I could not have accomplished this.

Dedication

To Drew, who has been a constant source of support over the last 5 years, to Mom, who inspires me to be brave in the pursuit of my goals, and to Dad, who instilled my curiosity and love of learning.

Abstract

Manual protocols to predict the number, size, and location of cancerous lesions in the prostate using imaging data are highly dependent on reader experience and expertise. Existing computer-aided voxel-wise classifiers do not directly provide estimates of lesion boundaries, which are clinically important. Spatial partitioning methods estimate boundaries separating regions of local stationarity in spatially registered data, but existing methods are inadequate for the application of lesion detection because the boundaries are restricted to be linear or piecewise linear. We first introduce a novel Bayesian functional spatial partitioning method (BFSP-1) which estimates the partitioning boundary around an anomalous region of data with a distinct distribution or spatial process. Our algorithm transitions between a fixed Cartesian and a moving polar coordinate system to model the boundary with functional estimation tools. Using adaptive Metropolis-Hastings, the BFSP-1 algorithm simultaneously estimates the partitioning boundaries and the parameters of the spatial distributions within each region. BFSP-1 assumes the data contain one and only one anomalous region. To create a more clinically useful tool, we build upon our original boundary estimation framework and propose BFSP-M for multiple region discovery. This method uses reversible jump Markov chain Monte Carlo to jointly estimate the number of lesions, their boundaries, and their distinct spatial processes. Finally, we discuss BFSP-3D, which extends the BFSP framework to three dimensions. Through simulation, we show that our methods are robust to the shape of the target zone and region-specific spatial processes. Our methods prove to be a clinically useful tool for automatic boundary drawing of cancerous lesions using non-invasive prostate imaging data.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Prostate cancer (PCa) remains a leading cause of cancer death in American men. In fact, in 2022 there will be an estimated 268,490 new cases of PCa and 34,500 deaths [3]. The burden of PCa is not equally distributed by race. Black men in America have a two-fold higher mortality rate as compared to their White counterparts [3]. Screening techniques offer the opportunity to reduce morbidity and mortality from PCa [4]. Currently, the most widely used modality is serum prostate-specific antigen (PSA) testing. However, PSA testing continues to have low positive and negative predictive value. This leads to unnecessary invasive prostate biopsy, while missing PCa in other patients [5].

Multiparametric MRI (mpMRI), a combination of several imaging parameters, has an increasingly important role in early detection of PCa [6]. MpMRI includes T2-weighted (T2w) MRI, diffusion-weighted imaging (DWI), dynamic contrast-enhanced MRI (DCE-MRI), and MR spectroscopy (MRS). Across several studies, sensitivity for detecting PCa was generally high (5/8 studies found sensitivity exceeding 90%). However, negative predictive value (NPV) is more clinically important than sensitivity. The NPV asnwers, "In a patient with a negative MRI, can we be confident that no significant prostate cancer is present?". MRI guided biopsy, where a targetable lesion is identified, has much higher NPV (96.9%) than standard transrectal ultrasound guided biopsy [7].

The diagnostic performance of mpMRI depends on reader skill and experience. Attempts to measure the NPV of mpMRI have come to differing conclusions based on the expertise of the study radiologists [6]. Computer aided diagnosis (CAD) systems aim to automate the process of lesion detection and thereby improve the accuracy and

consistency of diagnosis and detection [8]. CAD systems can make a clinical decision quickly, effectively, and reliably. Further, CAD systems have been shown to match or exceed the ability of experienced observers to detect cancer [9]. Recent developments in CAD systems include accounting for differences in cancer probability between anatomical regions of the prostate and modeling spatial correlations of the imaging parameters [10].

Previously, our group developed a voxel-wise classifier for prostate cancer using mpMRI data [1, 10], but an additional step is needed to translate the voxel-wise classifications into lesions for clinical practice [11]. Our goal is to partition the prostate and flag the cancerous lesion or lesions using the voxel-wise probabilities from Jin et al. [1], which are derived from the non-invasive mpMRI imaging data. Compared to traditional imaging analysis of voxel-wise cancer classification, lesion detection is more interpretable and meaningful in clinical practice.

The voxel-wise probabilities exhibit strong spatial autocorrelation and non-stationarity, complicating possible modeling strategies. Spatial autocorrelation is the phenomenon where measurements taken at nearby locations tend to be more similar than measurements taken far apart. Statistical methods applied to imaging data often ignore this autocorrelation. This type of model misspecification can lead to poor results in terms of false positives and false negatives, a phenomenon that was identified in the landmark paper by Eklund et al. [12]. When the mean or covariance is not constant across a space, we say the data exhibit spatial non-stationarity. In the mpMRI data and in turn the voxel-wise probability maps, the areas of cancer exhibit distinct properties, most noticeably a raised mean surface. Our methods leverage these properties by searching the image for the optimal boundaries separating the two (or more) distinct areas of the prostate. Thus the problem of lesion detection is equivalent to estimating the boundary curves or surfaces that partition the image into contiguous regions of homogeneous or stationary spatial processes.

Techniques for lesion detection generally fall into one of three categories: clustering, partitioning, or boundary/edge detection. Examples of clustering include k-means, and mixed models. In the context of voxel-wise probabilities, these algorithms assign cluster memberships by thresholding the voxel-wise probabilities. Some clustering methods account for the spatial nature of the data and encourage spatial contiguity [13, 14, 15].

However, none of these methods ensure perfect spatial contiguity. Spatial partitioning methods find the optimal set of boundaries that divide the data into spatially contiguous groups of similar observations. The simplest example are Regression and Classification Trees (CART). Spatial partitioning methods have also been widely developed to model spatial correlations [16, 17, 18]. Existing methods in this space generate strictly linear or piecewise linear boundaries which are insufficient for the problem of lesion detection, in which boundaries are smooth. Boundary and edge detection methods detect discontinuities in voxel intensity. Existing methods may not produce a closed boundary [19] or may not properly address the spatial correlations in the data [20]. Our proposed method enjoys the advantages of existing spatial partitioning methods and boundary detection methods.

We aim to partition the prostate with one or more closed boundaries into contiguous and disjoint regions with distinct mean surface and covariance structure in order to identify regions of tissue which are likely cancerous. Although our motivations are specific, our goal is to develop methods that have wide applicability. Our methods can be applied to 2D or 3D non-stationary spatial data to identify anomalous regions or "hot spots". Possible applications include structural parcellation of the brain, prostate, or other tissues based on imaging data and identification of elevated risk regions in disease mapping.

In Chapter 2 we introduce our novel spatial partitioning method for spatial data exhibiting exactly one anomalous region or "hot spot", BFSP-1. We develop a boundary estimation technique utilizing functional estimation tools that can be generalized for estimating boundaries with varying shape and smoothness. We assume distinct spatial distributions in each resulting partition and estimate those distributions simultaneously with the boundary parameters using Metropolis Hasting. Uncertainty of the boundary can be easily visualised with 95% point wise credible bands. A simulation study provides evidence that BFSP-1 can locate and estimate boundaries of varying shape, separating "hot spots" from noisy background data. In application to prostate imaging data containing exactly one cancerous lesion, BFSP-1 outperforms competing clustering, image segmentation, and boundary detection methods.

Our work in Chapter 3 is motivated to overcome a major drawback of BFSP-1. In practice, prostate imaging data may contain more than one cancerous region or no cancer

at all. We develop BFSP-M to fill this gap. BFSP-M uses Reversible Jump MCMC to jointly estimate the number of lesions, their boundaries, and their distinct spatial distributions. In addition to boundary uncertainty, we can quantify lesion uncertainty, i.e., how certain we are that each identified lesion is distinct from the surrounding healthy tissue. Our novel approaches for displaying uncertainty are more clinically relevant than voxel-level uncertainty, as we illustrate in the data analysis. BFSP-M uses the same boundary estimation technique as BFSP-1. Simulation studies show that BFSP-M is accurate in identifying the true number of lesions whether there are multiple or no lesions present.

Most existing lesion detection methods analyze one slice of mpMRI data at a time. We hypothesize that leveraging a patient's entire 3D dataset will improve the accuracy of lesion detection methods. In Chapter 4, we present BFSP-3D, a method to spatially partition 3D imaging data. We expand upon the methods developed in Chapter 2 to model a 2D boundary surface which partitions the image into an inner anomalous volume and its surrounding area. Uncertainty can be visualized with 95% pointwise credible surfaces. In simulation, BFSP-3D can estimate simulated volumes more accurately than competing methods. In application to 3D voxel-wise probabilities, our method provides a clinically useful presentation of cancer lesion location, shape, and volume.

Chapter 5 provides a discussion of previous chapters and avenues for future work.

Chapter 2

Bayesian Functional Spatial Partitioning for Single Lesion Detection

2.1 Introduction

Spatial partitioning is used to separate a spatially registered region of data into non-overlapping regions of local homogeneity or stationarity. Spatial partitioning estimates cluster membership by searching the (x, y) plane for a set of boundaries that maximizes the observed differences between zones. Decision tree methods are the simplest approach to spatial partitioning. The Classification and Regression Tree (CART) algorithm is the most widely used [21]. The CART algorithm is binary recursive; the split which maximizes the reduction in impurity is chosen, the data set is split and the process is repeated. In the 2D variable space, CART partitions the data into strictly rectangular zones.

Current methods can jointly estimate the partitions and spatial processes within each partition via a Bayesian approach. These methods have proven useful in modeling soil permeability [16], NASA simulations of a proposed rocket booster [17], and ozone data [18]. Further developments have utilized spatial partitioning models to reveal anomalies or "hot spots" to identify regions of heightened childhood cancer risk in Florida [22].

The method by Kim et al. [16] creates a Voronoi tesselation of the space by proposing centers and assigning each data point to its nearest center. This method provides added flexibility over treed methods as its partitioning boundaries are not restricted to be perpendicular to the axes. However, this method makes a strong assumption that spatial locations are classified to the nearest center. The method by Gramacy et al. [17] divides the space as in the CART algorithm to create regions of spatial stationarity. Konomi et al. [18] further developed this method using covariance approximations for more efficient computation with large datasets. In these methods, zones created by the partitions are assumed to follow stationary Gaussian processes. The partitioning boundaries estimated by these methods are strictly linear and lack the flexibility to detect arbitrarily shaped regions.

Our goal is to develop a spatial partitioning method that creates contiguous regions of an arbitrary shape that can be applied to any non-stationary spatial data to identify anomalous regions or "hot spots". Possible applications include structural parcellation of the brain, prostate, or other tissues based on imaging data and identification of elevated risk regions in disease mapping. We are specifically motivated by the desire to detect prostate cancer using multi-parametric magnetic resonance imaging (mpMRI) data from Metzger et al. [2].

Previously, our group developed a voxel-wise classifier for prostate cancer using these data [1], but an additional step is needed to translate the voxel-wise classifications into lesions for clinical practice [11]. Our goal is to partition the prostate and identify the cancerous lesion. The cancer lesion can be thought of as an anomalous contiguous region with a distinct spatial process within the healthy tissue separated by a closed boundary curve. Thus the problem of lesion detection is equivalent to estimating the boundary curve that partitions the image into contiguous regions of homogeneous spatial processes. Compared to traditional imaging analysis of voxel-wise cancer classification, lesion detection is more interpretable and meaningful in clinical practice.

Simple thresholding techniques fail to identify the entire lesion, while also identifying spurious lesions. Existing clustering, segmentation, and boundary detection methods are popular in the domain field of MRI but are inadequate for our specific motivating problem. Ensemble tree methods such as Boosted trees [23] and Bootstrapped or Bagged trees [24] are constructed using multiple decision trees. Tree ensembles tend

to have higher classification accuracy but are difficult to interpret Voxel-wise clustering methods such as K-means [25] and K-medoids [26] can create a partition of a 2D space by assigning cluster membership at each location, but they fail to create contiguous zones. Mixture models assign each data point with probabilities of belonging to each cluster and have been adapted to account for spatially registered data. The mixture methods by Shao et al. [13] and Xiong et al. [14] model spatial data via Markov Random Fields. The R-package bayesImageS offers computationally efficient methods to fit gaussian mixtures with spatial contiguity encouraged by a hidden Potts model [15]. Contiguous regions are not guaranteed by these methods. Boundary or edge detection methods search for points at which the data changes sharply. One such method, Bayesian areal wombling, identifies a set of edges that separate areas with high dissimilarity [19]. Bayesian areal wombling may not produce fully connected boundaries. The boundary detection method, "BayesBD", by Li at. al. [27] identifies a smooth and connected boundary but assumes no presence of spatial correlation in the data and assumes that the boundary center is known and fixed. Misspecification of the boundary center can result in poor performance. Convolutional neural networks (CNNs) are widely used in segmenting medical imaging data from MRI but require large sets of training data and may not lead to spatially contiguous clusters [28]. Further, deep learning methods lack interpretability and do not quantify the uncertainty in the results.

In this paper, we propose a novel Bayesian functional spatial partitioning (BFSP) method which estimates the closed curve surrounding a contiguous target zone with a distinct spatial process. Our method utilizes transitions between the fixed Cartesian coordinate system of the image space and a moving polar coordinate system to estimate the boundary using a Fourier-based functional approximation. The BFSP algorithm estimates the boundary and spatial parameters jointly through a Bayesian Markov chain Monte Carlo (MCMC) framework which allows for quantification of uncertainty in the boundary via point-wise credible bands. To our knowledge no method exists which jointly models spatial correlated data and guarantees spatially contiguous clusters via boundary detection. Although our motivating application is segmentation of MRI images, our method offers a general statistical framework for modeling non-stationary spatial data by a novel partitioning method. The functional tools implemented to detect boundaries of spatial correlation structures can be extended to 3D, longitudinal,

and/or multivariate spatial data analysis. In simulations, BFSP performs well, with the flexibility to detect irregularly shaped regions with high sensitivity and specificity compared to existing competing methods. An application to the data from Metzger et al. [2] and Jin et al. [1] illustrates the flexibility to identify cancerous lesions in the prostate using mpMRI.

The remainder of the paper is organized as follows. We introduce our novel method for functional spatial partitioning, discuss Bayesian modeling of the underlying spatial process, and discuss the computational implementation of our method in Section 2.2. We evaluate the statistical properties of our method via simulation and compare its performance to existing methods in Section 2.3. This is followed by an application of our method to the data from Jin et al. [1] in Section 2.4. Finally, we conclude with a discussion and suggest future extensions in Section 2.5.

2.2 Methods

Let $Z(s)$ indicate a scalar variable that is observed at location $s = (x, y) \in D$, where D is a region in $\mathcal{R}^2$. Let $\boldsymbol{s} = (s_1, ...s_n)$ and $\boldsymbol{Z} = (Z(s_1), ..., Z(s_n))$ denote a set of locations within the region D and the corresponding set of observations. We assume that D can be partitioned into a contiguous target region D_1 and an outer region D_0 by a closed boundary curve B, such that $n = n_0 + n_1$. Let $\boldsymbol{s}_0$ and $\boldsymbol{s}_1$ be the vectors of spatial locations within regions D_0 and D_1 respectively, and $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$ the vectors of observed values at those locations. Let $h(s) = 1$ if $s \in D_1$ and $h(s) = 0$ if $s \in D_0$. We further assume $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$ follow different spatial processes. Our goal is to estimate $\boldsymbol{h} = (h(s_1), ..., h(s_n))$ by $\hat{\boldsymbol{h}}$ via finding the boundary B based on the likelihood of $\boldsymbol{Z}$.

The target regions from our motivating data have irregular shapes. Our objective is to estimate the boundary that partitions the data into two regions based on heterogeneity in the underlying spatial processes with minimal restrictions on the shape of the target region. Our partitioning process utilizes functional tools for boundary estimation in a moving polar coordinate system and is accomplished iteratively via MCMC to jointly estimate the boundary and spatial parameters based on their posterior probabilities given the data. Details on approximating the partitioning boundary are provided in Section 2.2.1, details on modeling the spatial process within each partition is provided

in Sections 2.2.2 and 2.2.3, and a computational overview is given in Section 2.2.4.

2.2.1 Modeling the partitioning boundary

Popular tree-based partitioning methods generate boundary curve, B, only in rectangular shapes as they partition space by horizontal or vertical segments iteratively. However, this is a strong restriction that does not apply to general applications of nonstationary spatial data. Here we consider the case of allowing the boundary curve to be a smooth closed curve. Our proposed method can detect shapes in the star domain. A region S is a star domain if there exists an x_0 in S such that for all x in S the line segment from x_0 to x is in S.

A key feature of our method is that any partition of D is uniquely determined by the separating boundary curve B around the contiguous target zone, D_1. Given a point $c = (x_c, y_c)$ located at the center of D_1, the boundary curve can be modeled as a function of an angle in a polar coordinate system centered at c: $B = \{s \in D : d_{c \to s} = f_c(\theta_s), \theta_s \in [0, 2\pi]\}$, where $d_{c \to s}$ is the distance between a point s to the polar center c, and θ_s is the angle corresponding to s. Thus modeling the boundary curve is equivalent to estimating the location c and function $f_c(\theta)$.

Figure 2.1 illustrates the main idea of our method. For a given boundary curve, such a polar function is not unique but determined by the location of the polar center. To avoid identifiability problems in our Bayesian computation, we choose the polar center to always be fixed at the centroid of the current estimate. As the boundary is being updated, the centroid is updated as the mean of the spatial locations within D_1, $c \equiv \bar{s}_1$. In this way, the polar coordinate system is not fixed but moves with the centroid c as it updates. Each spatial location s in the fixed Cartesian coordinates can be uniquely defined in the polar coordinates by its angle $\theta_c(s)$ and distance $d_c(s)$ referenced to current centroid c. At a given iteration, we classify s as within the target region if s is within the boundary i.e. the distance $d_c(s)$ is less than the boundary curve at $\theta_c(s)$

$$\hat{h}(s) = \begin{cases} 1 \text{ if } d_c(s) \leq f(\theta_c(s)) \\ 0 \text{ if } d_c(s) > f(\theta_c(s)). \end{cases}$$

We model f using the Fourier approximation.
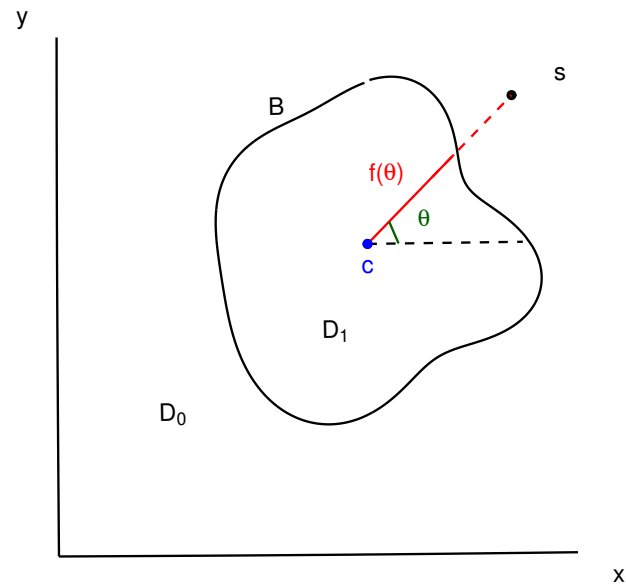
Figure 2.1: An example of how spatial locations are assigned given a boundary. A boundary B is shown in black. $s = (x, y)$ is a spatial location, $c = (x_c, y_c)$ is the centroid at the current iteration, θ is the angle between s and c, and $f(\theta)$ is the magnitude of the boundary at θ. In this example, s is categorized as outside the target region: the distance from s to c is greater than $f(\theta)$.

$$f\left(\theta_c(s)\right) \approx b_0 + \sum_{k=1}^{K} \left\{b_{1,k}\sin(k\theta_c(s)) + b_{2,k}\cos(k\theta_c(s))\right\} \ \text{for } 0 \leq \theta_c(s) \leq 2\pi,$$

where $b_0, \boldsymbol{b_1} = \{b_{1,1}, \cdots, b_{1,K}\}, \boldsymbol{b_2} = \{b_{2,1}, \cdots, b_{2,K}\}$ are coefficients for the Fourier basis series in the sine-cosine form to be estimated and K is a fixed positive integer. The Fourier approximation ensures a smooth and closed curve, i.e. $f(0) \equiv f(2\pi)$. This allows the target region D_1 to be flexible in size and shape. The use of a Fourier series to develop periodic regression models has been proposed previously [29, 30], but using this approach to develop partitioning boundaries is novel. The function $f()$ may also be modeled by any periodic basis function satisfying $f(0) = f(2\pi)$ to produce smooth, jagged, or straight edged boundary curves. In Section 2.3 we display the results of BFSP with the fourier basis specified above alongside results produced with periodic B-splines. To our knowledge, our method is the first to use functional tools in modeling a boundary to partition a region of non-stationary spatial data. This novel implementation expands the scope of applicability of spatial partitioning to general non-stationary spatial process data that is composed of homogeneous regions of arbitrary shapes.

A higher value of K will produce a boundary that captures fine detail whereas a lower value of K will produce a smoother boundary. We choose K to balance the desired level of smoothness, detail, and computational efficiency. Through simulation we determined that $K = 5$ can approximate shapes well with reasonable computational speed. In other applications, the user may specify the value of K based on desired smoothness of the boundary. We found that increasing K did not result in better DIC values for our data, see Figure A.1.

In the Markov Chain Monte Carlo (MCMC) computation of our Bayesian model, we update $\boldsymbol{b} = [b_0, \boldsymbol{b_1}, \boldsymbol{b_2}]$ and corresponding boundary curve B using the Metropolis Hastings (MH) algorithm. The probability for accepting a proposed boundary is proportional to the ratio of the likelihood of the data given the proposed boundary to the likelihood of the data given the previous boundary. Computation details are provided in Section 2.2.4.

2.2.2 Spatial Modeling

The estimation of the spatial distributions within each region happens jointly with the estimation of the boundary. Given a current estimate of c and $f_c(\cdot)$, we can partition the space into $D_1 = \{s \in D : d_c(s) \leq f_c(\theta_c(s)), \theta_c(s) \in [0, 2\pi)\}$ and $D_0 = \{s \in D : d_c(s) > f_c(\theta_c(s)), \theta_c(s) \in [0, 2\pi)\}$. In each region we assume a distinct stationary distribution.

$$
\begin{aligned}
\boldsymbol{Z_j}|c, f_c(\cdot) &\sim N(\mu_j + \xi_j, \tau^2 I_{n_j}) \\
\xi_j &\sim N(0, \sigma_j^2 H_j),
\end{aligned}
$$

where μ_j is the mean specific to region j, and ξ_j is the vector of spatial random effects that follows a region-specific Gaussian distribution with covariance $\sigma^2 H_j$. τ^2 is the variance for non-spatial random errors common to both regions.

In a Bayesian framework, we encourage spatial smoothness within each region using the proper conditional autoregression (CAR) prior for ξ_j's distribution. This leads to a valid marginal likelihood of data and can be used to efficiently approximate many spatial processes [31]. Our method can be implemented with various assumed covariance structures (see Figure A.2). However, we choose CAR as it is computationally efficient and approximates the Gaussian process well enough as demonstrated by our simulation and imaging data analysis. With specification of a CAR prior on the spatial random effects ξ_j, we have

$$
H_j = (I_{n_j} - \rho_j W_j)^{-1},
$$

where $W_{ii'}$ is the inverse Euclidian distance between locations s_i and $s_{i'}$ and I is the $n \times n$ identity matrix. We set the value of ρ_j to be 0.99 encourage a strong spatial correlation for each region [32] and avoid large matrix inversions in estimation. We show via simulation that our method is robust to any sensible value of ρ (see Figure A.3).

2.2.3 Prior specification

Let $\boldsymbol{\Phi} = \{(\mu_0, \mu_1), \boldsymbol{\sigma^2} = (\sigma_0^2, \sigma_1^2), \tau^2, \boldsymbol{b}, c\}$ be all of the parameters of the Gaussian distributions and boundary curve. The polar center c is the centroid of the target

region, the location of which is assumed to be uniform over the space domain D. We assign the following prior for $\boldsymbol{\Phi}$ assuming independence of the parameters:

$$
\begin{aligned}
\pi(\boldsymbol{\Phi}) &= \pi(\tau^2)\pi(\boldsymbol{\mu})\pi(\boldsymbol{\sigma^2})\pi(\boldsymbol{b}) \\
&= \pi(\tau^2)\pi(\mu_0)\pi(\mu_1 - \mu_0)\pi(\boldsymbol{\sigma^2})\pi(b_0)\pi(\boldsymbol{b_1})\pi(\boldsymbol{b_2})
\end{aligned}
$$

Without prior information, we assign vague priors for all the parameters in the boundary function and spatial parameters. Specifically, we assign vague normal priors $N(0, 100)$ to all basis coefficients in $\boldsymbol{b}$ except for b_0, which we assign a uniform prior covering zero to the approximate radius of the entire image space D as b_0 controls the approximate radius of the boundary. This encourages $f(\theta)$ to be positive and ensures that the area within the closed boundary does not greatly exceed the entire space of interest. Since the approximate radius of D is guaranteed to be larger than the approximate radius of the target region, results are not sensitive to choice of upper bound of the uniform prior. We assign the mean of the outer region a vague normal prior. The difference in means between the target region and the outer region is assumed to follow a vague half-normal prior because we have prior information that the mean of the anomalous region is elevated as compared to its complement. This can be modified for other applications. All variance parameters are assigned vague inverse Gamma priors $IG(.1, .1)$. To prevent the boundary from extending beyond the entire space of interest (specified by the user) we set a prior on B, $\pi(B)$, which equals 0 if any points in B lie outside the space of interest and 1 if not.

2.2.4   Bayesian Computation

MCMC algorithm

Model fitting is completed iteratively via MCMC. At each iteration a new set of boundary parameters and distribution parameters are proposed. They are accepted with probability determined by the ratio of the posterior probability of the proposed parameters to the posterior probability of the current parameters. For a partition of the data $\hat{\boldsymbol{h}}$, we can calculate the joint posterior probability of the boundary parameters and the

distribution parameters by the marginal likelihood times the prior.

$$\pi(\boldsymbol{\Phi}|\boldsymbol{Z}) \quad \propto \quad L(\boldsymbol{Z}|\boldsymbol{\Phi})\pi(\boldsymbol{\Phi}),$$

where $\pi(\boldsymbol{\Phi})$ is given in Section 2.2.3 and the form of $L(\boldsymbol{Z}|\boldsymbol{\Phi})$ will be discussed in Section 2.2.4.

Due to the complex nature of our partitioning method, we use adaptive MCMC to aid in convergence as we sample from the posterior [33]. Adaptive measures allow for adjustment of the proposal density based on the acceptance ratio. Let $\boldsymbol{\Psi} = \boldsymbol{\Phi} \setminus \boldsymbol{c}$ be the set of parameters to update via MCMC. $\boldsymbol{\Psi}$ is blocked into 3 blocks: $\boldsymbol{\Psi}_1 = \{\mu_0, \sigma_0^2, \tau^2\}$, $\boldsymbol{\Psi}_2 = \{\mu_1, \sigma_1^2\}$, $\boldsymbol{\Psi}_3 = \{b_0, \boldsymbol{b}_1\}$, and $\boldsymbol{\Psi}_4 = \{\boldsymbol{b}_2\}$. After initialization the generating algorithm is as follows: For $m = 1, 2, 3$:

1. Draw $\boldsymbol{\Psi}_m^*$ at iteration t from the proposal density $Q_t(\boldsymbol{\Psi}_m^*|\boldsymbol{\Psi}_m)$

2. Accept $\boldsymbol{\Psi}_m^*$ with probability $\alpha = \min\left(1, \frac{\pi(\boldsymbol{\Psi}^*|\boldsymbol{Z})}{\pi(\boldsymbol{\Psi}|\boldsymbol{Z})}\right)$

3. Update c^* to be the centroid of the shape defined by $\boldsymbol{\Psi}$ and c.

4. Update $\boldsymbol{b}$ based on new centroid c^* by solving $\boldsymbol{d}(B_t, c^*) = X_{c^*}\boldsymbol{b}$ for $\boldsymbol{b}$ where $\boldsymbol{d}(B_t, c^*)$ is $n \times 1$ vector of distances between the boundary points B_t and new centroid c^* and X_{c^*} is the $n \times (2 + 1K)$ matrix of basis functions referenced to c^*.

The proposal density $Q_i(\boldsymbol{\Psi}_m^*|\boldsymbol{\Psi}_m)$ is determined by the iteration number t and is based on an example from Roberts et al. [33].

$$Q_t(\boldsymbol{\Psi}_m^*|\boldsymbol{\Psi}_m) \sim N_d(\boldsymbol{\Psi}_m, (.1)^2 I_d/d) \quad \text{for} \quad t < 50$$
$$Q_t(\boldsymbol{\Psi}_m^*|\boldsymbol{\Psi}_m) \sim (1 - \gamma)N_d(\boldsymbol{\Psi}_m, \exp(l_{mt})\boldsymbol{\Sigma}_{mt}) + \gamma N_d(\boldsymbol{\Psi}_m, (.1)^2 I_d/d) \quad \text{for} \quad t \geq 50,$$

where d is the dimension of $\boldsymbol{\Psi}_m$, γ is a small positive constant (we set $\gamma = .01$), and $\boldsymbol{\Sigma}_{mt}$ is the empirical estimate of the covariance structure of the target distribution at iteration t based on the run so far. l_{mt} controls the step-size of the proposal distribution and is adjusted to achieve the desired acceptance rate of about 40% every 20 iterations by either adding or subtracting 0.1 to the previous value for l_n. The final partition is defined by the posterior mean of the all parameters after discarding samples from the burn-in period.

Marginal Likelihood Approximation

For a given partition of the data, we can calculate the marginal likelihood as

$$L(\boldsymbol{Z}|\boldsymbol{\Phi}) = L(\boldsymbol{Z_0}|\boldsymbol{\Phi})L(\boldsymbol{Z_1}|\boldsymbol{\Phi}), \tag{2.1}$$

where the distribution of $\boldsymbol{Z_j}$ is assumed to follow a multivariate Normal distribution $N(\mu_j, \tau^2 I_{n_j} + \sigma_j^2 H_j), j = 0, 1$ as in Section 2.2.2. Fitting the likelihood for the above model with large n is infeasible due to the time requirements of calculating the determinant and inverse of the covariance matrices. To further increase the computational efficiency of our method we assume a sparse covariance matrix by multiplying the assumed covariance matrix by a tapering kernel $C_\nu(s_i, s_{i'})$. Tapered covariances are commonly used to analyze large spatial datasets as they allow for sparse matrix algorithms. The tapering kernel must be a positive definite matrix such that the covariance between two spatial locations becomes zero after a given distance. The Wendland family of tapered covariance functions [34] are widely used. We assume the form of

$$C_\nu(s_i, s_{i'}) = \left(1 - \frac{h}{\nu}\right)_+^4 \left(1 + 4\frac{h}{\nu}\right),$$

where $h = ||s_i - s_{i'}||$ and we fix $\nu = .1$. This yields our approximation of the covariance:

$$\Sigma_j = \sigma_j^2 H_j \odot T_j + \tau^2 I_{n_j},$$

where H is a CAR correlation matrix, T_j is the $n_j \times n_j$ matrix with the (i, i')-th entry $C_\nu(s_i, s_{i'})$.

Posterior Boundary Uncertainty

A 95% point-wise credible band can be computed to visually display the posterior uncertainty of the boundary estimate. The MCMC samples require post processing due to the centroid not being fixed post burn-in. First, the average centroid, c_T, is computed using the post burn-in samples. Then, for each $\boldsymbol{b}_t, t = 1, ..., T$ where T is the total number of post burn-in samples, we compute the boundary parameters corrected

to centroid c_T, $\boldsymbol{b}_{t,c_T}$, by solving:

$$\boldsymbol{d}(B_t, c_T) = X_{c_T}\boldsymbol{b}_{t,c_T}$$

for $\boldsymbol{b}_{c_T}$ where $\boldsymbol{d}(B_t, c_T)$ is an $n \times 1$ vector of distances between boundary points B_t and c_T, and X_{c_T} is the $n \times 2K + 1$ matrix of basis functions where θ is measured to c_T. Then we can compute the vector $\boldsymbol{f}_{t,c_T}$ via:

$$\boldsymbol{f}_{t,c_T} = X\boldsymbol{b}_{t,c_T}$$

where X is a $200 \times 2K + 1$ vector of basis functions based on 200 equally spaced angles between 0 and 2π.

The 95% credible bands can be estimated with the corrected boundary functions $\boldsymbol{f}_{t,c_T}$ referenced to c_T as in Syring at. al. [20]. For all, $t = 1, ..., T$ MCMC samples post burn-in, we compute.

$$u_t = \sup_\theta \{|f_{t,c_T}(\theta) - \hat{f}_{c_T}(\theta)|/\hat{s}(\theta)\},$$

where $\hat{f}_{c_T}(\theta)$ and $\hat{s}(\theta)$ are the posterior mean and standard deviations of $\{f_{t,c_T}(\theta)\}_{t=1}^T$. The 95% credible interval for $f(\theta)$ is given by:

$$[\hat{f}_{c_T}(\theta) - L\hat{s}(\theta), \hat{f}_{c_T}(\theta) + L\hat{s}(\theta)],$$

where L is the 95th percentile of the u_t's. See Figure A.4 for an example 95% point-wise credible band for a simulated dataset

2.3   Simulation Study

We evaluate the classification accuracy and statistical properties of BFSP in comparison to competing partitioning methods through simulations. On a 50x50 regular grid we divide the area into a target region and an outer region. The target region is either square, triangular, or heart shaped. We model each of the two distinct areas with independent Gaussian processes with Matérn covariance structures. To test the robustness

of our method across different situations we vary the smoothness of the target region, ν_1, and the spatial nugget parameter, τ^2. For 3 different shapes we randomly generate 50 spatial data sets for each of the following settings.

1. Strong spatial correlation with low noise: $\nu_1 = 1.5, \tau^2 = .1$

2. Weak spatial correlation with low noise: $\nu_1 = .5, \tau^2 = .1$

3. Strong spatial correlation with moderate noise: $\nu_1 = 1.5, \tau^2 = .5$

4. Weak spatial correlation with moderate noise: $\nu_1 = .5, \tau^2 = .5$

The performance of BFSP and the competing partitioning methods are evaluated in terms of sensitivity, specificity, and the Dice coefficient which measures overall similarity of two partitions. Further details about the simulation settings and evaluation methods are provided in Section A.1.

We present the results of BFSP on simulated data as specified in Section 2.2. The MCMC chain is initialized by setting the centroid at the center of the entire space D, the partitioning boundary to be a large circle, the means of the Gaussian processes at -1 and 1, and the variance parameters at 1. We collect 50,000 iterations and discard the first 10,000 based on convergence of the boundary points B. We also present the results of BFSP using a spline basis for specifying the boundary function. [35].

We compare the results of BFSP to four competing methods. The first two, K-means (KM) with 2 clusters using the function kmeans() from the R Stats package [36] and classification trees (CART) using R package "tree" [37], are basic clustering methods that do not account for spatial correlations in the data. We also compare BFSP to two image segmentation methods, BayesBD [27] using the R-package "BayesBD" [20] and BayesImageS [15] using R-package "bayesImageS" [38] function mcmcPotts(). BayesBD partitions the space using a single connected boundary while BayesImageS is a voxel-wise classifier that encourages contiguous clusters. Details on the implementation of the competing methods are provided in Section A.1.

Figure 2.2 shows the distributions of the sensitivity, specificity, and Dice coefficient for detecting the target zone across 600 simulated datasets (50 for each of the 12 settings). Figure 2.2 shows that while spatial partitioning methods CART and BayesBD suffer from low sensitivity and voxel-wise method KM tends to have low specificity,
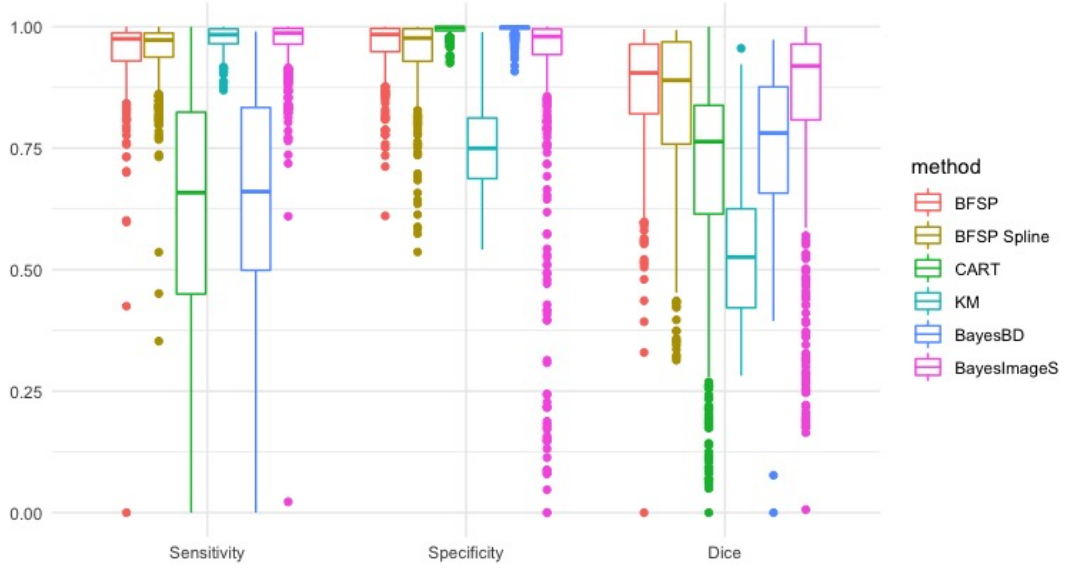
Figure 2.2: Distributions of sensitivity, specificity, and Dice coefficient of BFSP, BFSP Spline, CART, KM, BayesBD, and BayesImageS over all 600 simulated data sets.

BFSP maintains both high sensitivity and specificity in our simulated data. The performance of BFSP and BFSP Spline are similar to the performace of BayesImages in all measures of accuracy. On average, BFSP and BFSP Spline achieve sensitivity of 0.951 and 0.952, respectively, and specificity of 0.965 and 0.947 respectively. Further, BFSP and BFSP Spline achieve a strong average Dice coefficient of 0.876 and 0.840, respectively.

The mean and standard deviation of sensitivity, specificity, and Dice score for BFSP and competing methods are provided for each of the 12 unique settings for a heart shaped region in Table 2.1. Tables providing full results for square and triangular shaped regions are provided in Tables A.1, and A.2. Our method is robust to varying spatial smoothness and shape of the target region. Average sensitivity never falls below 89% and average specificity is at least 96% for all simulation settings. BFSP performs best on estimating square and heart-shaped target regions. Our algorithm has more difficulty capturing a triangular shaped region due to its sharp angles, but maintains average Dice coefficient of at least 0.81 over all settings.
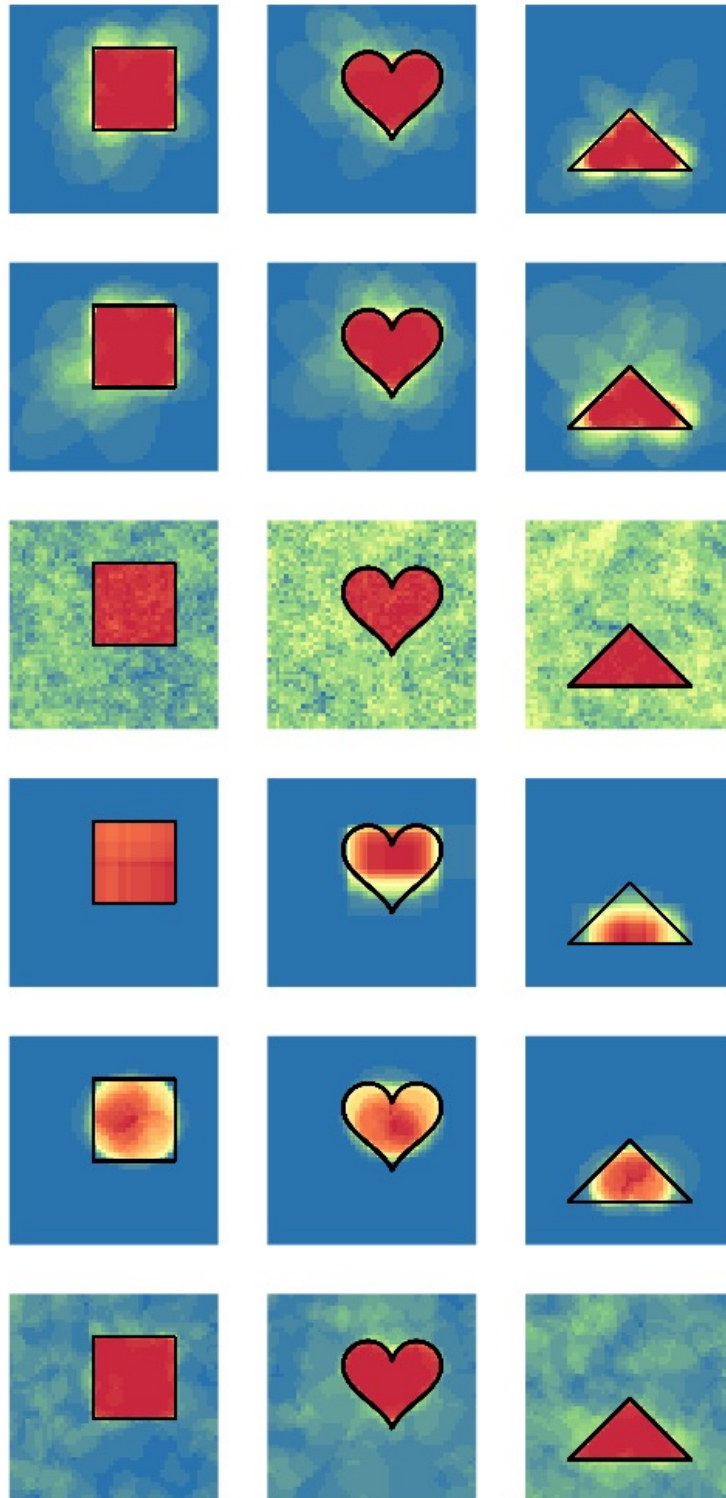
Figure 2.3: From top to bottom: average partitioning results of the BFSP, BFSP Spline, KM, CART, BayesBD, and BayesImageS methods in the simulation study. Each image represents 50 simulations of the same setting. The color represents the proportion of time the method classified each spatial location as within the target region. The black outline shows the true boundary.

To visualize the average classification performance we display results of 50 simulations for each of the 12 settings by averaging the estimated cluster memberships $\hat{\boldsymbol{h}}$ based on the estimated partitions. Figure 2.3 presents the probability that each voxel was included in the target region for 3 of the 12 settings for BFSP, BFSP Spline, CART, KM, BayesBD, and BayesImageS. The BFSP methods results in a high contrast in color between the two regions for the three pictured scenarios, indicating both a high sensitivity in the target zone detection and high specificity of the outer region. KM consistently identifies the target region, but does not guarantee spatial contiguity of the regions and thus random locations in the outer region are incorrectly assigned to the target region resulting in low specificity as seen in Figure 2.2 and Table 2.1. CART and BayesBD tend to detect a region that is smaller than the target region, explaining the low sensitivity seen in Figure 2.2 and Table 2.1. Because these methods do not account for spatial smoothness in the data, they fail to properly identify the entire target region of interest. BayesImageS performs well at identifying the entirety of the target region but often identifies spurious outer groups of voxels.

2.4   Data Analysis

Metzger et al. [2] present a unique set of data of men who received an MRI study prior to prostatectomy as definitive treatment for prostate cancer. After surgery, the removed prostates were sectioned following strict protocols allowing the prostate to be sectioned in planes that matched in vivo imaging. The sectioned prostates were made into slides and digitized allowing the study pathologist to annotate and label the regions of cancer thus identifying the extent and Gleason score of each region. These labeled regions were then co-registered to the in vivo imaging data thus creating the ground truth for disease through which the models were trained and validated. The process of pathologic correlation and registration are detailed in Kalavagunta et al. [39]. The mpMRI data are composed of 34 prostate slices, obtained from 34 patients, with 2098 to 5756 voxels per slice.

Previously our group developed a voxel-wise prediction model for prostate cancer using the mpMRI data detailed above. The goal of this analysis is to further develop a pipeline from imaging to lesion detection using non-invasive methods. Using the

predictive model from Jin et al. [1], we aim to identify the cancerous lesion within the prostate using our method. We specifically apply our partitioning method to the results of the "mcoord" classifier which models the mpMRI parameters and cancer risk by voxel coordinates and outputs a probability heatmap with values between 0 and 1.

Our aim is to translate a heatmap of cancer probabilities into a contiguous region of cancer. First we filtered out slices which contained multiple cancerous lesions or little to no healthy tissue. Then we selected the 6 slices for which the voxel-wise classifier of Jin et al. [1] was most successful at predicting the true voxel-wise cancer status. This allows us to focus on the performance of our segmentation relative to the ground truth, whereas an application for which the underlying voxel-wise classifier is misaligned with the ground truth will result in misaligned segmentation regardless of the performance of our method. We calculate true and false positive and negative rates using the true cancer status of each voxel to measure accuracy.

We compare BFSP to competing clustering methods CART and KM and competing image segmentation methods BayesBD and BayesImageS. We initialize BFSP with $\mu_0 = -1, \mu_1 = 1$, all variance parameters set to 1, and initial boundary set to be a circle centered at the spatial location with the highest probability. We collect 30,000 iterations and discard the first 10,000 based on convergence of the boundary points B. For BayesBD we set the center to be the spatial location with the highest probability. We also found the optimal threshold based on ROC analysis and present the resulting cluster memberships of the optimal threshold.

The results of lesion detection for three images by BFSP, CART, KM, BayesBD, and BayesImages are displayed in Figure 2.4. The drawbacks of tree-based partitions are clear. The linear boundaries created by CART are not reasonable for these data, which contain non-rectangular shaped regions of heightened risk. KM identifies spurious extra regions because it is not constrained to identify contiguous areas. Both methods suffer from their inability to account for spatial smoothness in the data. BayesBD can fail to identify the entire cancerous region for the same reason. BayesImageS and Thresholding also find spurious regions of interest but tend to identify the entire cancerous region quite well.

Over all 6 slices, BFSP achieves average sensitivity of 0.840 with standard deviation of 0.108 and average specificity of 0.946 with standard deviation 0.026. The average
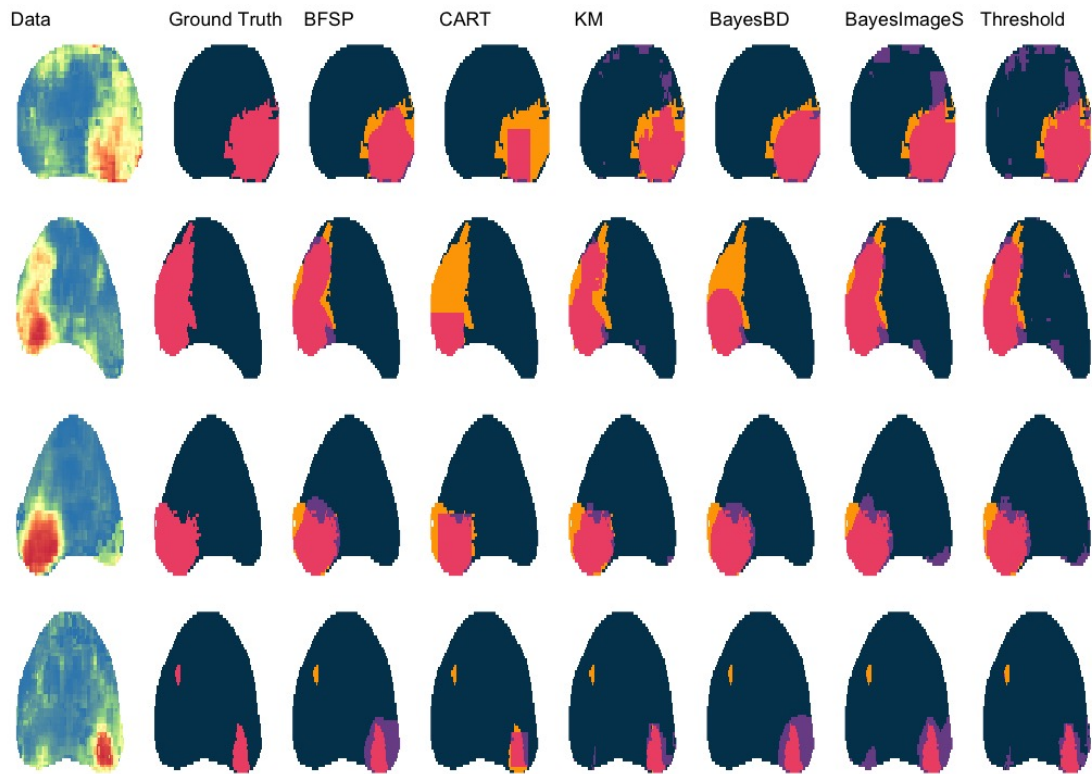
Figure 2.4: From left to right: Map of voxelwise predicted cancer probabilities from four sample slices [1], true cancer status, partitioning results from BFSP, CART, KM, BayesBD, BayesImageS, and Thresholding.

dice coefficient was 0.726 indicating good overlap with the true cancerous lesions. The standard deviation of the dice coefficient was 0.153.

2.5 Discussion

We propose a flexible Bayesian functional spatial partitioning method. A key advantage of our method is the ability to detect arbitrarily shaped regions, as compared to existing methods, which are either limited in their ability to detect non-standard shapes or that do not detect contiguous regions. Our method outperforms competing alternatives in our simulation study, maintaining high levels of sensitivity and specificity in detecting anomalous regions across varying levels of spatial smoothness and noise. We have shown that our method closely approximates the true cancerous lesion boundary in prostate cancer imaging data.

This research was motivated by the desire to identify prostate cancer lesions non-invasively using MRI data. There are previously developed voxel-wise classifiers for prostate cancer using MRI, but an additional step is needed to translate these into lesions for clinical practice. Prostate cancer lesions are contiguous and irregularly shaped and display spatial smoothness, which prevented the direct application of existing spatial partitioning methods. Our proposed method addresses all limitations and is a general approach to spatial partitioning. It results in flexible boundaries, which are more realistic than polygons with straight line boundaries in most cases.

The modeling of spatial heterogeneity and correlation structure in imaging statistics is critically important. BFSP uses novel functional tools to detect boundaries which separate regions of spatial homogeneity. In general, misspecification of correlation structure can lead to very poor results in terms of false positives and false negatives. This phenomenon was identified in the landmark paper by Eklund et al. [40]. In our own simulation study and data analysis, competing methods which ignore the heterogeneous spatial correlation structures exhibited poor performance. BFSP and BayesImageS consistently achieved better results due to correct specification of the spatial correlation structure of the data.

Although our method shows promise, several limitations still exist. First, our method is not as successful at estimating boundaries for regions that contain sharp

angles. To improve this, we could increase the number of Fourier bases at the expense of computation time or use a different set of basis functions. However, in our application cancerous lesions tend to be smooth so this is not needed. Second, due to the complex nature of the partitioning boundary estimation our method may be inefficient, especially on very large data sets. To address this inefficiency we use covariance approximation methods that significantly speed up computation.

As with any Bayesian model with a complex posterior distribution, achieving convergence within a reasonable amount of time is challenging. To address this, we chose to assume a simple and sparse form for the covariance structure within each estimated partition. This allows computational efficiency while assuming some degree of spatial smoothness. In simulation and data analysis we have shown our method's utility despite these assumptions. To further increase the efficiency of our algorithm we use an adaptive sampling scheme to aid in the convergence of the MCMC chain. Our method takes an average of 5.5 hours to run the full 50,000 iterations in the simulated data. Convergence on the target region typically happens quickly regardless of initial values.

Table 2.1: Results of BFSP and competing methods in simulated data. Mean and SD of sensitivity, specificity, and Dice coefficient for each simulation setting referenced in Section 2.3 for a heart shaped target region where ν_1 and τ^2 are parameters of the simulated Gaussian processes.

ν_1	τ^2	Method	Sensitivity Mean (SD)	Specificity Mean (SD)	Dice Mean (SD)
0.5	0.1	BFSP	0.966 (0.072)	0.962 (0.048)	0.877 (0.11)
		BFSP Spline	0.975 (0.039)	0.938 (0.084)	0.84 (0.152)
		CART	0.592 (0.269)	0.989 (0.008)	0.664 (0.249)
		KM	0.98 (0.016)	0.768 (0.067)	0.548 (0.077)
		BayesBD	0.691 (0.181)	0.996 (0.006)	0.791 (0.123)
		BayesImageS	0.985 (0.025)	0.922 (0.123)	0.824 (0.159)
0.5	0.5	BFSP	0.976 (0.029)	0.97 (0.042)	0.903 (0.097)
		BFSP Spline	0.975 (0.025)	0.958 (0.062)	0.88 (0.127)
		CART	0.686 (0.191)	0.985 (0.015)	0.747 (0.154)
		KM	0.974 (0.014)	0.734 (0.055)	0.507 (0.054)
		BayesBD	0.749 (0.2)	0.992 (0.011)	0.814 (0.129)
		BayesImageS	0.98 (0.024)	0.938 (0.14)	0.875 (0.169)
1.5	0.1	BFSP	0.971 (0.044)	0.963 (0.048)	0.883 (0.107)
		BFSP Spline	0.972 (0.037)	0.958 (0.061)	0.877 (0.128)
		CART	0.521 (0.249)	0.992 (0.006)	0.619 (0.242)
		KM	0.989 (0.022)	0.78 (0.085)	0.571 (0.102)
		BayesBD	0.621 (0.224)	0.997 (0.005)	0.732 (0.164)
		BayesImageS	0.979 (0.035)	0.926 (0.157)	0.852 (0.175)
1.5	0.5	BFSP	0.946 (0.075)	0.974 (0.028)	0.893 (0.087)
		BFSP Spline	0.941 (0.116)	0.966 (0.052)	0.877 (0.141)
		CART	0.587 (0.244)	0.986 (0.015)	0.665 (0.217)
		KM	0.973 (0.028)	0.729 (0.078)	0.508 (0.082)
		BayesBD	0.666 (0.174)	0.996 (0.006)	0.773 (0.123)
		BayesImageS	0.966 (0.045)	0.912 (0.207)	0.852 (0.202)

Chapter 3

Bayesian Functional Spatial Partitioning for Multiple Lesion Discovery

3.1   Introduction

Multiparametric magnetic resonance imaging (mpMRI), a combination of MRI images, has improved diagnostic accuracy for detecting prostate cancer compared to previous methods [41]. Computer-aided diagnostic (CAD) systems to estimate voxel-wise cancer probabilities based on mpMRI features continue to be developed and evaluated [9]. Previously, our group developed fully automated voxel-wise classifiers that account for regional heterogeneity in the prostate and spatial dependence in the data without the need for manual segmentation [1, 10]. While these models can guide clinicians in locating approximate areas with high probability of cancer, an additional step to translate probabilities into lesions is needed for use in clinical practice. Besides estimating location and size, correctly identifying the number of cancerous lesions is important, clinically [11].

Previously we developed Bayesian Functional Spatial Partitioning, re-named as BFSP-1, a boundary detection method for spatially correlated data exhibiting a single anomalous region [42]. This method proved to be successful when applied to data

26

derived from mpMRI containing a single cancerous lesion. However, BFSP-1 is not applicable when the data have zero more than one anomalous regions, which is common in practice. As we show in our simulations, BFSP-1 may detect one or more separate anomalous regions with a single partitioning boundary. This precludes applying BFSP-1 repeatedly to one imaging slice containing multiple anomalous regions. Further, we aim to estimate the uncertainty in both the number and location of lesion boundaries, which BFSP-1 cannot do.

Estimating the true number of distinct groups, clusters, or lesions in a spatially referenced dataset is a challenging and largely unsolved problem in frequentist statistics. Existing methods for estimating the true number of clusters tend to be ad-hoc or to require unrealistic assumptions and complicated calculations [43, 44, 45]. Determining the optimal number of terminal nodes or partitions in a decision tree is similarly challenging. Classification and regression trees (CART) [21] recursively split the data until no additional split improves the homogeneity within the nodes. CART tends to overestimate the number of nodes and often requires ad-hoc pruning methods to prevent over-fitting [46]. Clustering methods using likelihood maximization do not provide estimates of uncertainty in the number of clusters or cluster membership.

Bayesian modeling provides a straightforward framework for estimating the number of clusters and the uncertainty of the estimate. This is typically accomplished via Bayesian mixture models that estimate voxel-wise cluster membership [47, 48, 49]. Product partition models, first introduced by Hartigan [50], create clusters using a prior that takes the form of a product of cohesion functions. The cohesion function measures how likely data points are to be clustered together. Page and Quintana [51] developed a spatial version of the product partition model by making the cohesion a function of spatial location. These methods may not lead to spatially disjoint clusters which is a major drawback for our motivating problem. Kang et al.[52] developed a soft-thresholded Gaussian process model to partition the brain in order to detect regions that are highly predictive of alcoholism using EEG data. This regression method detects spatial regions that are important on average across several patients and time points. We aim to detect distinct regions from each patient's imaging data. None of the above methods provide estimates of boundaries or boundary uncertainty, which are of critical importance to our clinicians.

Bayesian random partitioning models can quantify uncertainty in the number and locations of boundaries but existing methods are not well suited for our application. Denison et al. [46] developed a Bayesian CART algorithm which provides insight into a range of good trees of varying depth rather than returning a point estimate. They further developed a Bayesian partition model for count data based on a Voronoi tessellation in which centroids are proposed and data points are assigned to the nearest center [53]. Chipman et al. [54] extended the Bayesian CART method by fitting hierarchical linear models in each terminal node. The spatial partitioning methods by Gramacy and Lee [17] and Konomi et al. [18] fit Gaussian processes (GPs) in the terminal nodes. The method by Kim et al. [16] also uses Voronoi tessellation in which the tiles follow distinct GPs. However, all of the above methods assume linear or piece-wise linear partitioning boundaries. Prostate cancer lesions can be of any shape, rendering these methods inadequate for our application. Our goal is to develop a fully automated lesion detection pipeline that jointly estimates the number of lesions and their boundaries from the mpMRI data using a Bayesian approach. This allows quantification of uncertainty in both the number of lesions and their boundaries.

We propose BFSP-M, a general framework to identify and describe an unknown number of arbitrarily shaped anomalous regions using boundary detection for spatially registered data. BFSP-M utilizes separate moving polar systems for each lesion, within which the boundary curves are estimated with functional approximation, and reversible jump Markov chain Monte Carlo (RJ-MCMC) to explore parameter spaces of different dimensions which vary with the number of lesions. In the same vein as Richardson and Green [55], we develop four jump steps to add, subtract, split, and merge the lesion-defining boundaries. They use jump proposals to add or subtract densities in a mixture model, while our method defines proposals to add or subtract boundaries and updates the corresponding boundary parameters. These moves efficiently avoid the local-trap problem in our Bayesian spatial partitioning and ensure faster mixing and convergence in estimating the number of lesions. Compared to machine learning methods for imaging segmentation, our method allows estimation of uncertainty about the number of lesions and each lesion boundary. Further, our method provides a statistical modeling framework that allows extensions to multi-subject and/or longitudinal analysis of lesion status. We show through simulations that BFSP-M can detect multiple irregularly

shaped regions with higher sensitivity and specificity compared to competing image segmentation methods. An application to the data from Metzger et al. [2] and Jin et al. [1] illustrates the flexibility to identify an unknown number of cancerous lesions of arbitrary shape in the prostate using mpMRI. While our application is specific, this method of partitioning is general and can be applied to any type of spatially related data that contains anomalous areas within a larger homogeneous space.

The remainder of the paper is organized as follows. We introduce our novel Bayesian model for functional spatial partitioning and discuss Bayesian modeling of the underlying spatial process in Section 3.2. We discuss the computational implementation of our method via RJ-MCMC in Section 3.3. We evaluate the statistical properties of our method via simulation and compare its performance to existing methods in Section 3.4. This is followed by an application to the data from Jin et al. [1] in Section 3.5. Finally, we conclude with a discussion and suggest future extensions in Section 3.6.

3.2 Model Specification

Let $Z(s)$ indicate a scalar variable that is observed at location $s = (x, y) \in D$, where D is a region in $\mathcal{R}^2$. In our application, we assume D to be a unit square. Let $\boldsymbol{s} = (s_1, ... s_n)$ and $\boldsymbol{Z} = (Z(s_1), ..., Z(s_n))$ denote a set of locations within the region D and the corresponding set of observations. Assume some partition of the space, P_m, with $m \geq 0$ mutually exclusive contiguous regions, $D_1, .., D_m$, all within D, with boundaries $B_1, ..., B_m$ and $D_0 = D \setminus D_1 \setminus \cdots \setminus D_m$. Let $\boldsymbol{s}_j$ be the vector of spatial locations within region D_j for $j = 0, ..., m$, and $\boldsymbol{Z}_j$ the vector of observed values at those locations. We assume each vector $\boldsymbol{Z}_j$ follows a distinct distribution.

Our objective is to estimate the set of boundaries $\{B_1, ..., B_m\}$ that partitions the data into multiple regions based on heterogeneity in the underlying spatial processes with minimal restrictions on the shapes of the target regions. Estimation is accomplished iteratively via RJ-MCMC to jointly estimate the number of target regions, the boundary parameters, and the spatial parameters based on their posterior probabilities given the data.

We aim to jointly estimate the partition with m regions and the parameters defining that partition and the distinct distributions of each generated region, $\boldsymbol{\Phi}_m = \{\phi_0, ..., \phi_m\}$,

given the data $\boldsymbol{Z} = \{\boldsymbol{Z}_0, ..., \boldsymbol{Z}_m\}$. We assume the following hierarchical model:

$$
\begin{aligned}
\pi(\boldsymbol{\Phi}_m, m | \boldsymbol{Z}) &\propto \pi(\boldsymbol{Z} | \boldsymbol{\Phi}_m, m) \pi(\boldsymbol{\Phi}_m | m) \pi(m) & (3.1) \\
\pi(\boldsymbol{Z} | \boldsymbol{\Phi}_m, m) &= \prod_{j=0}^{m} \pi(\boldsymbol{Z}_j | \boldsymbol{\Phi}_m) \sim \prod_{j=0}^{m} N(\mu_j, \Sigma_j) \\
\pi(\boldsymbol{\Phi}_m | m) &= \prod_{j=0}^{m} \pi(\phi_j) \\
\pi(m) &= \frac{1}{M_{\max} - M_{\min} + 1}
\end{aligned}
$$

Section 3.2.1 discusses how a partition is defined by a set of boundaries, $B_1, ..., B_m$. In Section 3.2.2, we provide details about the likelihood $\pi(\boldsymbol{Z} | \boldsymbol{\Phi}_m, m)$. Details on the form of $\pi(\boldsymbol{\Phi}_m | m)$ and $\pi(m)$ are provided in Section 3.2.3. Finally, details of Bayesian computations are given in Section 3.3 including approximations to the likelihood in 3.3.3.

3.2.1 Defining a Partition

A partition of D with m boundaries $B_1, ..., B_m$ creates $m + 1$ regions $D_0, D_1, ..., D_m$. In our motivating problem, D_0 represents the healthy tissue and regions $D_1, ..., D_m$ represent m non-overlapping cancerous regions. Given a point $c_j = (x_{c_j}, y_{c_j})$ located at the centroid of D_j, the boundary curve of D_j can be uniquely modeled as a function of an angle in a polar coordinate system centered at c_j. We define the boundary $B_j = \{s \in D : d_{c_j \to s} = f_j(\theta_s | c_j), \theta_s \in [0, 2\pi]\}$, where $d_{c \to s}$ is the distance between a point s and the polar center c, and θ_s is the angle corresponding to s. Thus modeling the boundary curves is equivalent to estimating the locations of the c_j and the functions $f_j(\theta | c_j)$ for $j = 1, ..., m$. Each spatial location s can be assigned to a region by its distances $d_{c_j}(s)$ referenced to the centroids c_j for $j = 1, ..., m$. See Figure 3.1 for an illustration of our partitioning method. Boundaries are prevented from overlapping during the estimation process by prohibiting acceptance of overlapping boundaries in the MCMC. See Section 3.3.2 for details.

The boundary functions are estimated by linear combination of K basis functions.

For $j = 1, ..., m$:

$$f_j \left(\theta_{c_j}(s) \right) \approx \sum_{k=0}^{K} \beta_{jk} h_k \left(\theta_{c_j}(s) \right) \text{ for } 0 \leq \theta_{c_j}(s) \leq 2\pi$$

where $\boldsymbol{\beta}_j$ are coefficients to be estimated and $h(\cdot)$ are basis functions. The only constraint on the choice of basis function is that $h(\cdot)$ must be periodic ensuring $f(0) = f(2\pi)$. In our previous work we showed that Fourier and spline basis functions work well for our application [42]. While we use Fourier and spline basis functions for lesion detection, wavelets can also be used to model bumpy boundary functions as long as they satisfy the constraint. For the simulation and data analysis, we implement our method with Fourier basis functions. This restricts the boundaries to be in the star domain. A region S is in the star domain if there exists an x_0 in S such that for all x in S the line segment from x_0 to x is in S. The parameter K corresponds to the level of detail in the boundary. A lower value will produce a smoother boundary. In our previous work we found that $K = 5$ was adequate for lesion detection in our data. However for other applications, the value of K can be treated as an additional parameter and chosen by cross-validation by the user.

3.2.2 Spatial Modeling

Estimation of the spatial process within each region occurs simultaneously with estimation of the boundary. At each iteration, by transitioning back into Cartesian coordinates we can classify s as within region D_j if s is within the boundary, i.e., the distance $d_{c_j}(s)$ is less than the boundary curve at $\theta_{c_j}(s)$. Given a current estimate of m regions with centroids c_j and boundary functions $f_{c_j}(\cdot)$ for $j = 1, ..., m$, we can partition the space into $D_j = \{s \in D : d_{c_j}(s) \leq f_{c_j}(\theta_{c_j}(s)), \theta_{c_j}(s) \in [0, 2\pi)\}$ and $D_0 = \{s \in D : s \notin D_j \forall j = 1, ..., m\}$. For a given partition of the data with m boundaries, we can calculate the likelihood as the product of $m + 1$ independent likelihoods:

$$L(\boldsymbol{Z}|\boldsymbol{\Phi}_m, m) \quad = \prod_{j=0}^{m} L(\boldsymbol{Z_j}|\boldsymbol{\Phi}_m)$$

In each region we assume a distinct stationary Gaussian spatial process. For $j =$
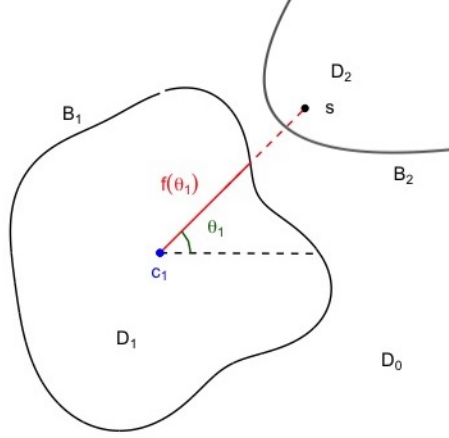
Figure 3.1: An example of how spatial locations are assigned given multiple boundaries. Boundaries B_1 and B_2 are shown in black. $s = (x, y)$ is a spatial location, $c_1 = (x_c, y_c)$ is the centroid of B_1 at the current iteration, θ_1 is the angle between the horizontal and the line from s to c_1, and $f(\theta_1)$ is the magnitude of the boundary at θ_1. In this example, s is categorized as within region D_2 but not D_1: the distance from s to c_1 is greater than $f(\theta_1)$.

$0, ..., m$:

$$\begin{aligned} \boldsymbol{Z_j} | \boldsymbol{\Phi}_m &\sim N(\mu_j + \boldsymbol{\xi}_j, \tau^2 I_{n_j}) \\ \boldsymbol{\xi}_j &\sim N(0, \sigma_j^2 H_j) \end{aligned}$$

where μ_j is the mean specific to region j, and $\boldsymbol{\xi}_j$ is the vector of spatial random effects that follow a region-specific Gaussian distribution with covariance $\sigma_j^2 H_j$. The variance for non-spatial random errors, τ^2, is assumed to be common to all regions.

We encourage spatial smoothness within each region by assuming an exponential kernel.

$$H_j(s_i, s_i') = \exp(-\gamma r_{i,i'})$$

where $r_{i,i'}$ is the Euclidean distance and γ is the spatial decay parameter which is assumed to be common to all regions. Other kernels can be specified. We chose the exponential kernel due to its simplicity and its ability to approximate other types of

spatial correlation structures well.

We marginalize over the spatial random effects and model the data in each region by:

$$
\begin{aligned}
\boldsymbol{Z_j}|\boldsymbol{\Phi}_m &\sim N(\mu_j, \Sigma_j) \\
\Sigma_j &= \sigma_j^2 H_j + \tau^2 I_{n_j}
\end{aligned}
$$

In Section 3.3.3, we provide an approximation to this likelihood in order to increase computational efficiency.

3.2.3 Prior Specification

Let $\boldsymbol{\Phi}_m = \{\phi_0, \phi_1, ..., \phi_m\}$ be all of the parameters describing a given partition P_m where $\phi_0 = \{\mu_0, \sigma_0^2, \tau^2\}$ and $\phi_j = \{\mu_j, \sigma_j^2, \beta_j, c_j\}$ for $j = 1, ..., m$. The joint prior $\pi(\boldsymbol{\Phi}_m, m)$ is equal to the product of the conditional $\pi(\boldsymbol{\Phi}_m|m)$ and $\pi(m)$. We specify a discrete uniform prior between $M_{\min}$ and $M_{\max}$ for m. We allow $M_{\min}$ to be 0 corresponding to no lesions detected in the image slices.

$$
\pi(m) = \frac{1}{M_{\max} - M_{\min} + 1}
$$

We assign the following prior for $\boldsymbol{\Phi}_m$ given m assuming independence of the parameters:

$$
\begin{aligned}
\pi(\boldsymbol{\Phi}_m|m) &= \prod_{j=0}^{m} \pi(\phi_j) \\
&= \pi(\tau^2)\pi(\gamma)\pi(\mu_0)\pi(\sigma_0^2) \prod_{j=1}^{m} \pi(\mu_j - \mu_0)\pi(\sigma_j^2)\pi(\beta_j)
\end{aligned}
$$

Without prior information, we assign vague priors for all parameters in the boundary functions and spatial processes. Specifically, we assign normal priors $N(0, 1)$ to all basis coefficients in β_j except for β_{j0}. This is a vague prior since $\beta_{jk} << \beta_{j0}$ for $k = 1, ..., K$ when the area of the entire region is standardized to be a 1×1 unit square. For β_{j0} which we assign a uniform prior covering zero to the approximate radius of D to ensure that the area within the closed boundary does not greatly exceed the entire space of

interest. We assign the mean of the outer region a vague normal prior $N(0, 100)$. The difference in means between a target region and the outer region is assumed to follow a vague half-normal prior $HN(0, 10)$ because we have prior information that the mean of the anomalous region is elevated as compared to the rest of the space. This can be modified for other applications. Following the advice of Banerjee et al. [32], we assign vague priors for the variance parameters σ^2 and τ^2 and an informative prior for spatial range parameter γ. Thus $IG(.1, .1)$ is the prior used for σ^2 and τ^2 parameters and the spatial range parameter γ follows $G(3, .5)$. We specify a discrete uniform prior between $M_{\min}$ and $M_{\max}$ for P_m. We allow $M_{\min}$ to be 0 corresponding to no lesions detected in the image slices.

3.3 Bayesian Computation via RJ-MCMC

Our proposed method uses the Bayesian paradigm to jointly model the number of lesions and their boundaries. A varying number of target regions with correspondingly varying number of parameters necessitates jumping between parameter spaces of different dimensions. The RJ-MCMC algorithm by Green [56] is an extension of the Metropolis Hastings (MH) algorithm that allows movement between different dimensional spaces. In the RJ-MCMC, we propose addition, deletion, splitting, and merging of boundaries and corresponding lesions to allow the chain to move between different partitions. We refer to the addition of a new boundary as a "birth", the deletion of an existing boundary as a "death". We also introduce the merging of two existing boundaries as a "merge", and the splitting of one existing boundary into two as a "split". The "split" and "merge" types of jump efficiently prevent the MCMC chain from being trapped in local minima and thus greatly speed up mixing and convergence of spatial partitions. With a specified probability, a birth, death, split, or merge is proposed and the current boundary and distribution parameters are updated. Moves are automatically rejected if they lead to overlapping boundaries. Adaptive MH is used to update the boundary and GP parameters given the current number of regions.

Let m be the current number of regions in MCMC, which is between $M_{\min}$ and $M_{\max}$. We allow movements that increase or decrease the number of regions by one at each iteration. The RJ-MCMC algorithm is composed of two steps: 1. propose a

new partition with corresponding parameters, and 2. calculate the acceptance ratio to determine whether to accept or reject the proposed move. Let $\boldsymbol{\Phi}_m$ be all of the parameters of a partition with m boundaries. RJ-MCMC uses two types of proposal distributions.

- $j(m'|m)$ is the model proposal. It defines the probability of switching from P_m to $P_{m'}$ and must be reversible.

- $q(\mathbf{v})$ is the auxiliary variable proposal distribution where $\mathbf{v}$ is used to match dimension between m and m'. In our case, for example, when a new boundary is proposed, $\mathbf{v}$ contains the parameters defining that new boundary and GP.

RJ-MCMC also requires a mapping function $h_{m,m'}(\boldsymbol{\Phi}_m, \mathbf{v})$ which maps $(\boldsymbol{\Phi}_m, \mathbf{v})$ to $(\boldsymbol{\Phi}_{m'}, \mathbf{v}')$. The function $h_{m,m'}$ is a deterministic function and must be bijective so that its inverse is well-defined. The acceptance ratio of a move from m to m' boundaries is given by

$$\alpha\{(P_m, \boldsymbol{\Phi}_m), (P_{m'}, \boldsymbol{\Phi}_{m'})\} = \min\{1, A_{m,m'}(\boldsymbol{\Phi}_m, \boldsymbol{\Phi}_{m'})\}$$

where the ratio $A_{m,m'}(\boldsymbol{\Phi}_m, \boldsymbol{\Phi}_{m'})$ is

$$A_{m,m'}(\boldsymbol{\Phi}_m, \boldsymbol{\Phi}_{m'}) = \frac{\pi(m', \boldsymbol{\Phi}_{m'}|\mathbf{Z})j(m'|m)q(\mathbf{v}'|\boldsymbol{\Phi}_{m'}, m')}{\pi(m, \boldsymbol{\Phi}_m|\mathbf{Z})j(m|m')q(\mathbf{v}|\boldsymbol{\Phi}_m, m)} \left| \frac{\partial h_{m,m'}(\boldsymbol{\Phi}_m, \mathbf{v})}{\partial(\boldsymbol{\Phi}_m, \mathbf{v})} \right|$$

The joint posterior distribution is given by Equation 3.1. We set $j(m+1|m) = j(m|m) = j(m-1|m) = 1/3$ which ensures that at each iteration the chain is equally likely to propose a jump to a higher dimension, a lower dimension, or stay at the same dimension. All other moves have probability zero. When $m = M_{\max}$, $j(m|m) = j(m-1|m) = 1/2$ and other moves have probability zero. When $m = M_{\min}$, $j(m+1|m) = j(m|m) = 1/2$ and other moves have probability zero.

If all proposed model parameters are generated via the proposal density $q()$, the Jacobian is equal to 1. This will be the case for all proposed parameters except for the σ^2 terms. For the birth, split, and merge steps we propose $\log(\sigma^2)$ because σ^2 is constrained to $[0, +\infty)$. In these cases, the Jacobian equals the product of proposed variance parameters. See Section B.2 for a more detailed derivation of the Jacobian for each type of jump.
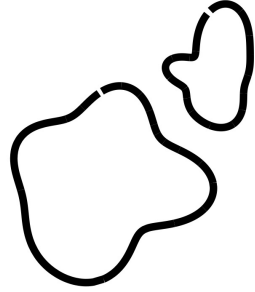
(a) Two boundaries before merge
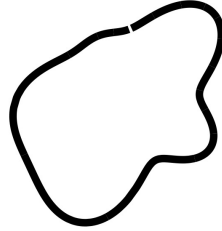
(b) Merged boundaries
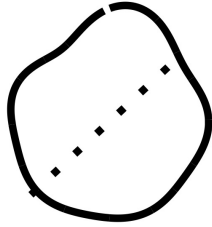
(c) One boundary with random bisecting line (dotted)
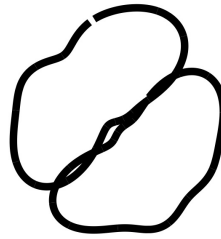
(d) Split boundaries

Figure 3.2: An illustration of the "split" and "merge" steps outlined in Section 3.3.1

Next, we define the proposal density for each type of move: birth, death, merge, and split.

3.3.1 Jump Steps

Birth: The proposal of a new boundary includes the generation of a new boundary function by proposing c_{m+1}, β_{m+1} and defining the GP of the resulting new region by $\mu_{m+1}, \sigma^2_{m+1}$. The parameters of the outer region remain the same. The new parameters are proposed via the proposal distribution $q_b(\phi_{m+1}|\mathbf{\Phi}_m, m)$. The acceptance rate for a

proposed birth is given by:

$$A_{m,m+1}(\mathbf{\Phi}_m, \mathbf{\Phi}_{m+1}) \quad = \quad \frac{\pi(m+1, \mathbf{\Phi}_{m+1}|\mathbf{Z})j(m+1|m)\sigma_{m+1}^2}{\pi(m, \mathbf{\Phi}_m|\mathbf{Z})j(m|m+1)q_b(\phi_{m+1}|\mathbf{\Phi}_m, m)}$$

where $q_b()$ is the proposal density for a birth.

$$
\begin{aligned}
q_b(\phi_{m+1}|\mathbf{\Phi}_m, m) \quad &= \quad q_b(c_{m+1}, b_{0,m+1}, \mu_{m+1}, \log(\sigma_{m+1}^2)|\mathbf{\Phi}_m, m) \\
&= \quad \frac{1}{n_0} \times \text{Unif}(b_{0,m+1}; 0, .5) \times N(\mu_{m+1}; \bar{\mu}, 1) \times N(\log(\sigma_{m+1}^2); \log(\bar{\sigma}^2), 1).
\end{aligned}
$$

where n_0 is the number of voxels in the outer region. The birth proposal distribution generates a circle at a random location within the outer region, of random size, with GP parameters based on existing regions where $\bar{\mu} = (1/m)\sum_{i=1}^m \mu_i$ and $\bar{\sigma}^2 = (1/m)\sum_{i=1}^m \sigma_i^2$. The choices for the proposal densities are dependent on our specific motivating data.

Death: When a death is proposed a boundary is randomly selected with equal probability from m existing boundaries to be removed. The voxels that were within selected boundary m are absorbed by the outer region D_0. The parameters of the outer region remain the same. The proposal density for $\mathbf{v}'$ is $1/m$ since each boundary is equally likely to be chosen for deletion. The acceptance rate for a proposed death is given by:

$$A_{m,m-1}(\mathbf{\Phi}_m, \mathbf{\Phi}_{m-1}) = \frac{\pi(m-1, \mathbf{\Phi}_{m-1}|\mathbf{Z})j(m-1|m)1/m}{\pi(m, \mathbf{\Phi}_m|\mathbf{Z})j(m|m-1)}$$

Split: A split step involves splitting one target region into two with distinct GPs by bisecting one boundary. A split is proposed by selecting one boundary at random with equal probability from the existing boundaries. Without loss of generality, let boundary m be the chosen proposed boundary to split. Then, two points on the boundary p_1, p_2 are sampled with a uniform distribution on $[0, 2\pi]$. The two points define the line, l, that will split the region. The two new centers c_{m^*}, c_{m+1} are proposed as the centroids of the two new areas. The resulting sets of boundary parameters are computed by solving

the following equations for β_{m*} and β_{m+1}:

$$r(\{B_m < l \cup B_l\}, c_{m*}) = X_{c_{m*}}\beta_{m*}$$
$$r(\{B_m \geq l \cup B_l\}, c_{m+1}) = X_{c_{m+1}}\beta_{m+1}$$

where $r(\{B_m < l \cup B_l\}, c_{m*})$ is the vector of distances between the new centroid c_{m*} and the new set of boundary points for the m^* boundary that consists of 100 equally spaced points along l and the boundary points B_m that lie below l. The proposal distribution q_s generates the parameters of the GPs for the resulting boundaries and generates a boundary based on the solution above. We use a proposal density given by:

$$
\begin{aligned}
q_s(\phi_{m*}, \phi_{m+1}|\boldsymbol{\Phi}_m, m) &= q_s(\mu_{m*}, \mu_{m+1}, \log(\sigma_{m*}^2), \log(\sigma_{m+1}^2), p_1, p_2, \beta_{m*}, \beta_{m+1}|\boldsymbol{\Phi}_m, m) \\
&= N(\mu_{m*}; \mu_m, 1) \times N(\log(\sigma_{m*}^2); \log(\sigma_m^2), 1) \times N(\mu_{m+1}; \mu_m, 1) \\
&\quad \times \ N(\log(\sigma_{m+1}^2); \log(\sigma_m^2), 1) \times \text{Unif}(p_1, p_2; 0, 2\pi) \times q_{s,\beta}(\beta_{m*}, \beta_{m+1}; \beta_m, l)
\end{aligned}
$$

where $q_{s,\beta()}$ is a point mass distribution concentrated at the solution of the above equations. The proposal density for $\mathbf{v}'$ is $1/m$ since each boundary is equally likely to be chosen for splitting. The acceptance rate for a proposed split is given by:

$$A_{m,m+1}(\boldsymbol{\Phi}_m, \boldsymbol{\Phi}_{m+1}) = \frac{\pi(m+1, \boldsymbol{\Phi}_{m+1}|\mathbf{Z})j(m+1|m)(1/m)\sigma_{m*}^2\sigma_{m+1}^2}{\pi(m, \boldsymbol{\Phi}_m|\mathbf{Z})j(m|m+1)q_s(\phi_{m*}, \phi_{m+1}|\boldsymbol{\Phi}_m, m)}$$

where ϕ_{m*} and ϕ_{m+1} are the sets of parameters generated when an existing boundary is split into two boundaries and q_s is the proposal distribution for a split given above.

Merge: When a merge is proposed, two boundaries are selected at random to become one region. WLOG, let boundaries m and $m-1$ be proposed to merge. The centroid of the combined voxels is c_{m*}. The boundary parameters of the new merged boundary m^* are computed by solving the following equation for β_{m*}

$$r(B_{m*}, c_{m*}) = X_{c_{m*}}\beta_{m*}$$

where B_{m*} is the set of n boundary points in $\{B_m \cup B_{m-1}\}$. The proposal distribution q_m generates the parameters of the GPs of the resulting new boundary and generates a

boundary based on the solution above. We use a proposal density given by:

$$
\begin{aligned}
q_m(\phi_{m^*}|\mathbf{\Phi}_m, m) &= q_m(\mu_{m^*}, \log(\sigma_{m^*}^2), \beta_{m^*}|\mathbf{\Phi}_m, m) \\
&= N(\mu_{m^*}; (\mu_m + \mu_{m-1})/2, 1) \times N(\log(\sigma_{m^*}^2); (\sigma_m^2 + \sigma_{m-1}^2)/2, 1) \\
&\times q_{m,\beta}(\beta_{m^*}; \beta_m, \beta_{m-1})
\end{aligned}
$$

where $q_{m,\beta()}$ is a point mass distribution concentrated at the solution of the above equation. The proposal density for $\mathbf{v}'$ is $2/m(m-1)$ with is the probability of selecting two boundaries at random from m. The acceptance rate for a proposed merge is given by:

$$
A_{m,m-1}(\mathbf{\Phi}_m, \mathbf{\Phi}_{m-1}) = \frac{\pi(m-1, \mathbf{\Phi}_{m-1}|\mathbf{Z})j(m-1|m)(2/m(m-1))\sigma_{m^*}^2}{\pi(m, \mathbf{\Phi}_m|\mathbf{Z})j(m|m-1)q_m(\phi_{m^*}|\mathbf{\Phi}_m, m)}
$$

where $\mathbf{\Phi}_{m^*}$ is the set of parameters generated when two existing boundaries are merged into one and q_m is the proposal distribution for a merge given above.

3.3.2 MCMC algorithm

Below we summarize our MCMC algorihtm. To aid in convergence, we sample from the posterior via adaptive MCMC [33]. Adaptive sampling allows adjustment of the proposal density of the boundary and GP parameters based on the acceptance ratio. At iteration i with m regions the algorithm proceeds as follows:

1. Propose a jump by sampling m' from the probability distribution $j(m'|m)$

    - If $m' = m$ proceed to step 4

    - If $m' = m + 1$ with equal probability propose a birth or split step. If $m = 0$ propose a birth step.

    - If $m' = m - 1$ with equal probability propose a death or merge step. If $m = 1$ propose a death step.

2. Accept or reject the jump proposed in Step 2 according to acceptance probabilities in Section 3.3.1.

3. Update ϕ_0 given m:

(a) Draw ϕ_0^* at iteration i from the proposal density $Q_{0,i}(\phi_0^*|\phi_0)$

(b) Accept ϕ_0^* with probability $\alpha = \min\left(1, \frac{\pi(\boldsymbol{\Phi}^*|\boldsymbol{Z})}{\pi(\boldsymbol{\Phi}|\boldsymbol{Z})}\right)$

4. Update ϕ_j, for $j = 1,..,m$ given m:

(a) Draw ϕ_j^* at iteration i from the proposal density $Q_{j,i}(\phi_j^*|\phi_j)$

(b) If ϕ_j^* creates a partition where $\exists s_k \in D_m \cap D_n$ for $n \neq m, k \in 1,...,n$ (i.e. overlapping boundaries), reject ϕ_j^* else

(c) Accept ϕ_j^* with probability $\alpha = \min\left(1, \frac{\pi(\boldsymbol{\Phi}^*|\boldsymbol{Z})}{\pi(\boldsymbol{\Phi}|\boldsymbol{Z})}\right)$

(d) Update c_j^* to be the centroid of the shape defined by Φ_j and c_j.

(e) Update β_j based on new centroid c_j^* by solving $\boldsymbol{r}(B_j, c_j^*) = X_{c_j^*}\boldsymbol{b}_j$ for $\boldsymbol{b}_j$ where $\boldsymbol{r}(B_t, c^*)$ is $n \times 1$ vector of distances between the boundary points B_j and new centroid c_j^* and $X_{c_j^*}$ is the $n \times (2K+1)$ matrix of basis functions referenced to c^*.

The proposal density $Q_{j,i}(\phi_j^*|\phi_j)$ is adaptive to the current posterior samples at iteration i and based on an example from Roberts and Rosenthal [33].

$$Q_{j,i}(\phi_j^*|\phi_j) \sim N_{d_j}(\phi_j, (.1)^2 I_{d_j}/d_j) \quad \text{for} \quad i < 50$$
$$Q_{j,i}(\phi_j^*|\phi_j) \sim (1-\omega)N_d(\phi_j, \exp(l_{j,i})\boldsymbol{\Sigma}_i) + \omega N_d(\phi_j, (.1)^2 I_{d_j}/d_j) \quad \text{for} \quad i \geq 50$$

where d_j is the dimension of Φ_j, ω is a small positive constant (we set $\omega = .01$), and $\boldsymbol{\Sigma}_i$ is the empirical estimate of the covariance structure of the target distribution at iteration i based on the run so far. $l_{j,i}$ controls the step-size of the proposal distribution and is adjusted to achieve the desired acceptance rate of about 40% every 20 iterations by either adding or subtracting 0.1 to the previous value for $l_{j,i}$.

3.3.3 Likelihood Approximation

Fitting the likelihood with the spatial model specified in 3.2.2 with large n is infeasible due to the time requirements of calculating the determinant and inverse of the covariance matrices. To further increase the computational efficiency of our method we assume a sparse covariance matrix by multiplying the assumed covariance matrix by a tapering kernel $C_\nu(s_i, s_{i'})$. Tapered covariances are commonly used to analyze large

spatial datasets as they allow for sparse matrix algorithms. The tapering kernel must be a positive definite matrix such that the covariance between two spatial locations becomes zero after a given distance. The Wendland family of tapered covariance functions [34] are widely used. We assume the form of

$$C_\nu(s_i, s_{i'}) = \left(1 - \frac{h}{\nu}\right)_+^4 \left(1 + 4\frac{h}{\nu}\right),$$

where $h = ||s_i - s_{i'}||$ and we fix $\nu = .1$. This yields our approximation of the covariance:

$$\Sigma_j = \sigma_j^2 H_j \odot T_j + \tau^2 I_{n_j}$$

where H is based on the exponential kernel, T_j is the $n_j \times n_j$ matrix with the (i, i')-th entry $C_\nu(s_i, s_{i'})$.

If the sample size is quite large, additional approximations may be neccessary to increase efficiency. For the simulation study and data analysis we make a further simplification to the likelihood by dividing D_0 into four roughly equally sized and independent zones. The mean x and y coordinates of $s \in D_0$ at a given iteration define the boundaries that split the area into $\boldsymbol{Z_{0,1}}, ..., \boldsymbol{Z_{0,4}}$. Thus, the likelihood becomes:

$$L(\boldsymbol{Z}|\boldsymbol{\Phi}_m, m) \quad = \prod_{l=1}^4 L(\boldsymbol{Z_{0,l}}|\boldsymbol{\Phi}_m) \prod_{j=1}^m L(\boldsymbol{Z_j}|\boldsymbol{\Phi}_m)$$

3.3.4   Posterior Boundary Estimation & Uncertainty

The MCMC samples require post processing due to the centroids not being fixed post burn-in. Using all sampled boundaries after burn-in, we group together boundaries that have similar centroids into boundary groups based on the distance between centroids. In our simulations and data analysis, we assume that boundaries are in the same boundary group if their centroids are within a distance of 1/10 of the entire area. Then we keep all boundary groups with group centroids that are present in $\geq 50\%$ of the samples. The number of boundary groups is the estimated number of anomalous regions. For each boundary group, the average centroid, c_T, is computed using the post burn-in samples. Then, for each $\boldsymbol{\beta}_t, t = 1, ..., T$ where T is the total number of post burn-in samples, we

compute the boundary parameters corrected to centroid c_T, $\boldsymbol{\beta}_{t,c_T}$, by solving,

$$\boldsymbol{d}(B_t, c_T) = X_{c_T}\boldsymbol{\beta}_{t,c_T}$$

for $\boldsymbol{\beta}_{c_T}$ where $\boldsymbol{d}(B_t, c_T)$ is an $n \times 1$ vector of distances between boundary points B_t and c_T, and X_{c_T} is the $n \times (2K + 1)$ matrix of basis functions where θ is measured to c_T. Then we can compute the vector $\boldsymbol{f}_{t,c_T}$ by:

$$\boldsymbol{f}_{t,c_T} = X\boldsymbol{\beta}_{t,c_T}$$

where X is a $200 \times (2K+1)$ vector of basis functions based on 200 equally spaced angles between 0 and 2π. The final estimate of a boundary is given by the mean of $\boldsymbol{f}_{t,c_T}$ over all t. This is repeated for all boundary groups.

The 95% credible bands can be estimated with the corrected boundary functions $\boldsymbol{f}_{t,c_T}$ referenced to c_T as in Li and Ghosal [27]. For all, $t = 1, ..., T$ MCMC samples post burn-in, we compute.

$$u_t = \sup_{\theta}\{|f_{t,c_T}(\theta) - \hat{f}_{c_T}(\theta)|/\hat{s}(\theta)\},$$

where $\hat{f}_{c_T}(\theta)$ and $\hat{s}(\theta)$ are the posterior mean and standard deviations of $\{f_{t,c_T}(\theta)\}_{t=1}^T$. The 95% credible interval for $f(\theta)$ is given by:

$$[\hat{f}_{c_T}(\theta) - L\hat{s}(\theta), \hat{f}_{c_T}(\theta) + L\hat{s}(\theta)],$$

where L is the 95th percentile of the u_t's. See Figure 3.3 for an illustration of the estimated partitioning boundaries and 95% credible bands for a simulated dataset.

3.4 Simulation Study

We evaluate the classification accuracy and statistical properties of BFSP-M in comparison to competing spatial partitioning and image segmentation methods through simulation. We consider a unit square image space D of 40 by 40 resolution and generate data for each region from independent Gaussian processes with Matérn covariance
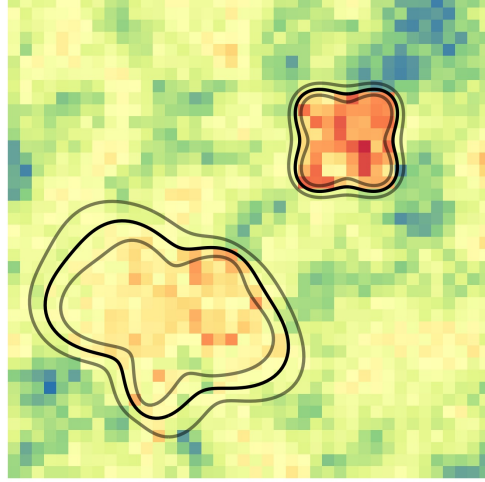
Figure 3.3: Estimated partitioning boundary (black) and 95% credible bands (grey) for one simulated dataset . Color represents voxel intensity. The procedure for simulating this data is outlined in Section 3.4.

structures. For $j = 0, 1, ..., M$

$$\boldsymbol{Z_j}|\nu_j, \rho_j \sim N(\mu_j, \Sigma_j), \quad \Sigma = \sigma_j^2 H(\nu_j, \rho_j) + \tau^2 I_{n_j}$$

$$(H(\nu_j, \rho_j))_{i,i'} = H(\nu_j, \rho_j; d_{i,i'}) = \frac{2^{1-\nu_j}}{\Gamma(\nu_j)} \left(\sqrt{2\nu_j} \frac{d_{i,i'}}{\rho_j} \right)^\nu \mathcal{K}_{\nu_j} \left(\sqrt{2\nu_j} \frac{d_{i,i'}}{\rho_j} \right)$$

where $d_{i,i'}$ is the Euclidean distance between spatial locations s_i and $s_{i'}$, Γ is the gamma function, $\mathcal{K}_\nu$ is the modified Bessel function, ν is a smoothness parameter, and ρ is a spatial range parameter.

To test the robustness of our method, we consider two different scenarios with either zero or two targeted regions in the space. Here we present the results for the simulations with two targeted regions. Refer to Section B.1 for simulation settings and results for data containing zero target regions. In the simulated data with $m = 2$ regions, we include a large heart and a small square shaped anomalous region with means $\mu_1 = 2$ and $\mu_2 = 4$, and vary the smoothness of the target region, $\nu_j, j > 0$, and the spatial nugget parameter, τ^2. We fix the following settings during simulation: $\nu_0 = 1$, $\rho_0 = \cdots = \rho_M = .05$, $\mu_0 = 0$, $\sigma_0^2 = \sigma_1^2 = \cdots = \sigma_M^2 = 1$. We randomly generate 100 spatial

data sets for each of the following settings.

1. $\nu_1 = \nu_2 = 1.5, \tau^2 = .1$

2. $\nu_1 = \nu_2 = .5, \tau^2 = .1$

3. $\nu_1 = \nu_2 = 1.5, \tau^2 = .5$

4. $\nu_1 = \nu_2 = .5, \tau^2 = .5$

We set the maximum number of boundaries that can be proposed $M_{max} = 10$ and the minimum number of boundaries to be $M_{min} = 0$. We collect 50000 MCMC samples and remove the first 20000 as burn-in samples. To increase computational efficiency we divide the outer region into four zones and approximate the likelihood of the entire outer region by aggregating the likelihoods of the four zones. In simulations this did not impact the accuracy of our method but significantly increased computational efficiency.

The performance of BFSP-M is evaluated in terms of the frequency of correctly estimating the number of regions, and the sensitivity, specificity, and the Dice coefficient of voxel-level classification. For the settings with two anomalous regions, we compare the classification results of BFSP-M to four competing methods, two of which are basic clustering methods, K-means (KM) with 2 clusters using the function kmeans() from the R Stats package [36] and a two-stage CART method. The two-stage method involves first estimating the partitions via CART using [37]. This results in a tiling of the space. The second stage is to group the partitions into two clusters via KM. The cluster of tiles with the higher mean represents the target regions. We also compare to an image segmentation method BayesImageS (BIS) [15] using R-package "bayesImageS" function mcmcPotts() for which we specified neighbors via KNN with k=8 and 4 blocks and set the number of clusters at 2. We also ran these three competing methods with the number of clusters equal to 3 and all resulted in worse average Dice and Sensitivity scores. We further include BFSP-1 from our previous paper [42] which assumes the presence of one and only one partitioning boundary. We initialized the MCMC chain with a large circular boundary centered at the center of the space and run for 30,000 iterations discarding the first 10,000 as burn-in samples.

In the 400 simulated data sets, using the number of boundary groups present in at least half of post burn-in simulations, our algorithm correctly identified the number of
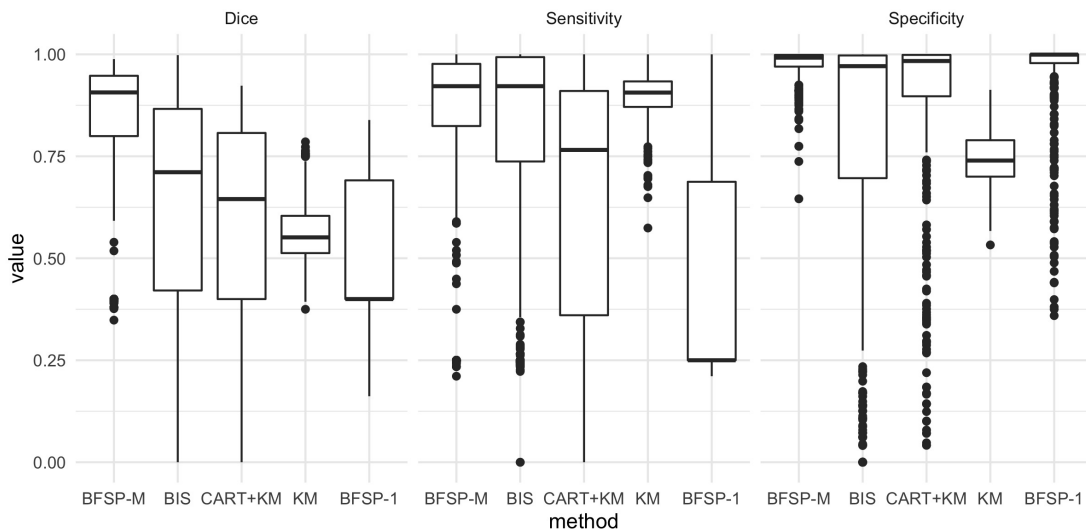
Figure 3.4: Distributions of sensitivity, specificity, and Dice coefficient of BFSP-M, KM, Two-Stage CART, BayesImageS, and BFSP-1 over 400 simulated data sets.

regions in 84% of the cases, identified a single boundary in 9% of cases, and estimated 3 boundaries in 7% of cases. Estimation of the number of distinct clusters is not possible with other competing methods in which the number of clusters must be pre-specified. Figure 3.4 presents the distributions of the sensitivity, specificity, and Dice coefficient in detecting the target zone across the 400 simulated datasets (100 for each of the first 4 settings). Varying the parameters of the spatial processes did not have a large effect on the accuracy of BFSP-M. Figure 3.4 shows in contrast, the two-stage method suffers from low sensitivity and KM tends to have low specificity. BFSP-M maintains both high sensitivity and specificity in all simulation settings. Due to the large difference in means of the two target regions, BayesImages often fails to identify both regions. On average, BFSP-M achieves a sensitivity of 0.841, specificity of 0.975, and Dice coefficient of 0.842.

The average sensitivity, specificity, and Dice coefficient are provided for each of the first 4 settings in Table B.1 along with results for competing methods. Our method is robust to varying spatial smoothness. Average sensitivity is greater than 75%, in all scenarios, and average specificity is at least 97% for all simulation settings. BFSP-M maintains average Dice coefficient of at least 0.77 over all settings.
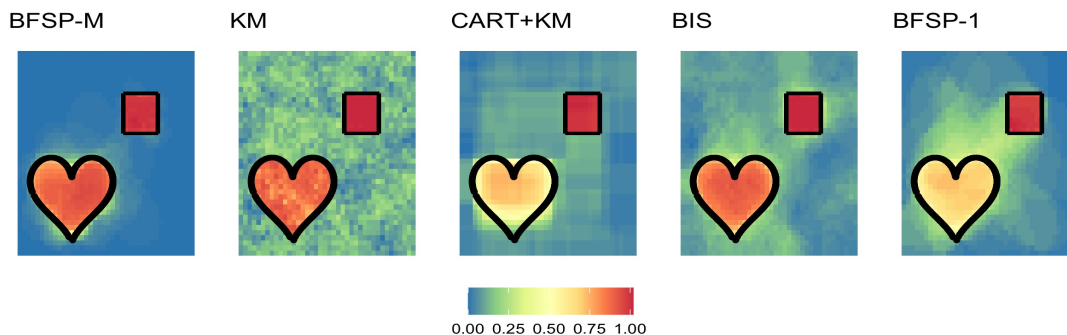
Figure 3.5: From left to right: average partitioning results of the BFSP-M, KM, Two-Stage CART, BayesImageS, BFSP-1 methods in the simulation study. Each image represents 100 simulations of the first simulation setting. The color represents the proportion of time the method classified each spatial location as within a target region with red close to 1 and blue close to 0. The black outline shows the true boundaries.

To visualize the average classification performance we summarize the results of 100 simulations for the first setting by averaging the estimated cluster memberships based on the estimated partitions. Figure 3.5 presents the probability that each voxel was included in a target region for BFSP-M, KM, the two-stage CART method, BIS and BFSP-1. The BFSP-M method leads to the best contrast in color between the two target regions and the background, indicating both a high sensitivity in the target zone detection and high specificity of the outer region. The heart shaped region was more difficult to detect for all methods because its mean is relatively close to that of the outer region. The two-stage method often misses the heart shaped region altogether, explaining the low sensitivity seen in Figure 3.4. KM consistently identifies the target region, but does not guarantee spatial contiguity of the regions and thus random locations in the outer region are incorrectly assigned to the target region resulting in low specificity as seen in Figure 3.4. On average, BayesImageS identifies both regions but is more likely to missclassify voxels in the outer region than BFSP-M. BFSP-1 performs poorly due the assumption that only one target region is present in the data. BFSP-M is the only method that simultaneously estimates the number of separate regions and the voxel-wise cluster membership.

Additional simulation results evaluating the performance of BFSP-M in the presence of zero lesions are presented in Section B.1. Our results indicate that BFSP-M was

highly likely to correctly identify no anomalous regions when the covariance structure is correctly specified, but performance is somewhat degraded when the covariance is misspecified.

3.5   Data Analysis

Metzger et al. [2] present a set of data of men who received an MRI study before prostatectomy as treatment for prostate cancer. The prostates were sectioned in planes to match in vivo imaging. The sectioned prostates were digitized, allowing the pathologist to annotate and label the regions of cancer. These labeled regions were then co-registered to the imaging data, thus creating the ground truth for disease through which the models were trained and validated. This process is detailed in Kalavagunta et al. [39]. The mpMRI data are composed of 34 prostate slices, obtained from 34 patients, with 2098 to 5756 voxels per slice.

Previously our group developed a voxel-wise prediction model for prostate cancer using the mpMRI data detailed above. The goal of this analysis is to further develop a pipeline from imaging to lesion detection using non-invasive methods. Using the results from the predictive model of Jin et al. [1], we aim to identify the cancerous lesions within the prostate using BFSP-M. We specifically apply our partitioning method to the results of the "mregion" classifier, which models the mpMRI parameters and cancer risk by voxel coordinates and outputs a probability heatmap with values between 0 and 1.

Our aim is to translate a heatmap of cancer probabilities into one or more contiguous regions of cancer. We selected the 2 slices for which there was more than one cancerous lesion. Then we selected another 9 slices containing one lesion for which the voxel-wise classifier of Jin et al. [1] was most successful at predicting the true voxel-wise cancer status. This allows us to focus on the performance of our segmentation relative to the ground truth, whereas an application for which the underlying voxel-wise classifier is misaligned with the ground truth will result in misaligned segmentation regardless of the performance of our method. The BFSP-M model is applied to each slice separately. We calculate true and false positive and negative rates using the true cancer status of each voxel to measure accuracy.

We compare BFSP-M to competing spatial clustering and image segmentation methods CART, KM, BayesImageS, and our previous method BFSP-1. First we scale each slice such that all voxels fall within the unit square. We initialize BFSP-M with $\mu_0 = 0$, $\mu_1 = 1$, all variance parameters set to 1, and one boundary set to be a small circle centered at the voxel with the highest probability. We collect 50,000 iterations and discard the first 10,000 based on convergence of the boundary points B. Samples of posterior boundary coefficients corresponding to centroid groups that were not present in at least half of the MCMC samples are discarded as in Section 3.4. The minimum number of possible boundaries is set to be 1 because we have prior knowledge that all slices contain cancer. The maximum number of boundaries is set to be 10. We fixed the maximum number of boundaries at 10 as a conservative upper bound. The true number of lesions is less than 5 in all tested slices. Also, the MCMC sampler never visited the maximum of 10 in any of the slices. The user may consider alternate values for $M_{\max}$ for other applications. To increase computational efficiency we divide the outer region into four zones and treat them as independent to approximate the likelihood. For competing methods we use the same initial settings as discussed in Section 3.4. For BFSP-1 we use the same initial settings as BFSP-M.

The results of lesion detection for three images by BFSP-M, KM, Two-Stage CART, BayesImageS, and BFSP-1 are displayed in Figure 3.6. The drawbacks of linear boundaries are clear. The estimates from CART are not reasonable for these data, which contain non-rectangular shaped regions of heightened risk. KM identifies spurious extra regions because it is not constrained to identify contiguous areas. Both methods often fail to capture the entire lesion due to their inability to account for spatial smoothness in the data. BFSP-1 will only identify one region, therefore, when multiple are present in the data as in row 1 of Figure 3.6, the estimated boundary covers several lesions and severly overestimates the extent of the cancer. BayesImageS and BFSP-M find spurious regions of interest but tend to identify the entire cancerous region quite well. Table B.2 shows sensitivities, specificities, and Dice scores for BFSP-M and all competing methods for each of the 11 slices.

Credible bands can be computed for each discovered lesion as in Section 3.3.4. See Figure B.2 for 95% credible bands for one prostate imaging slice. In addition to the
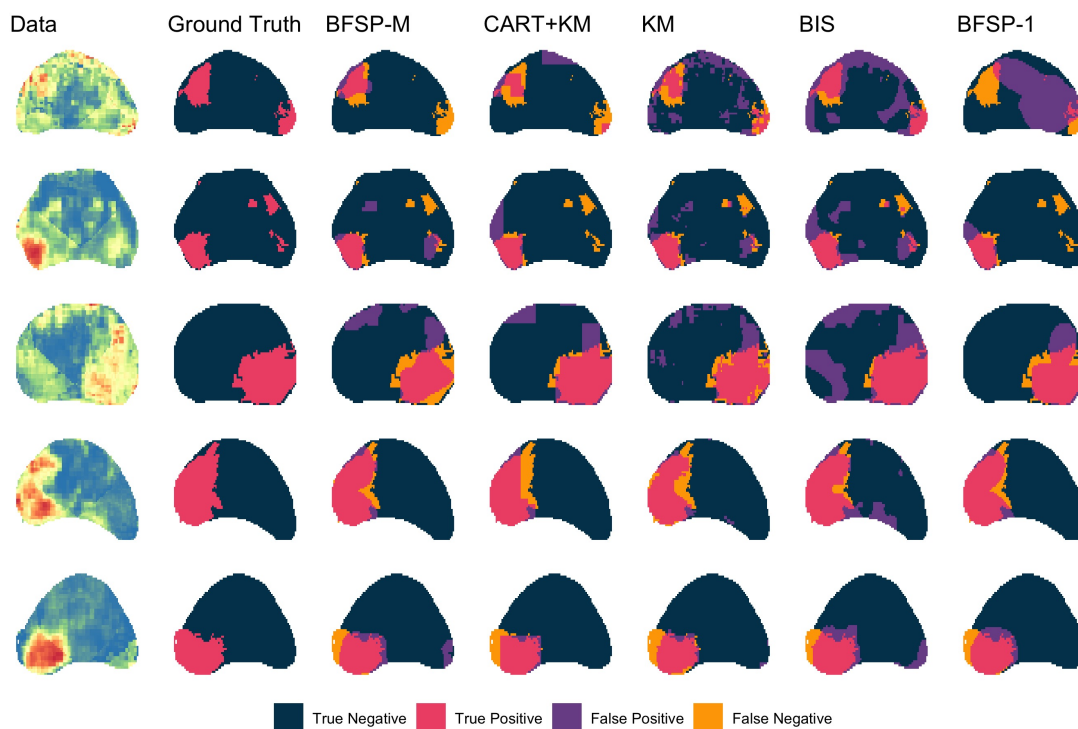
Figure 3.6: From left to right: Map of voxelwise predicted cancer probabilities from five slices where color indicates probability of cancer with red close to 1 and blue close to 0 [1], true cancer status, partitioning results from BFSP-M, Two-Stage CART, KM, BayesImageS, and BFSP-1.

boundary uncertainty, we can also compute the uncertainty of each detected lesion. Lesion uncertainty is determined by calculating the proportion of times that each centroid group appears in an MCMC posterior sample. Figure 3.7 displays the lesion uncertainty for two slices of data that contain multiple distinct lesions. The color of each lesion corresponds to the proportion of MCMC samples in which that lesion is present. Lesions colored red appear in all or almost all of the post burn-in MCMC samples whereas, lesions colored pale green appear less often in MCMC sampling. This result is consistent with the data which displays a high degree of noise and weak signal for several of the true cancerous lesions.
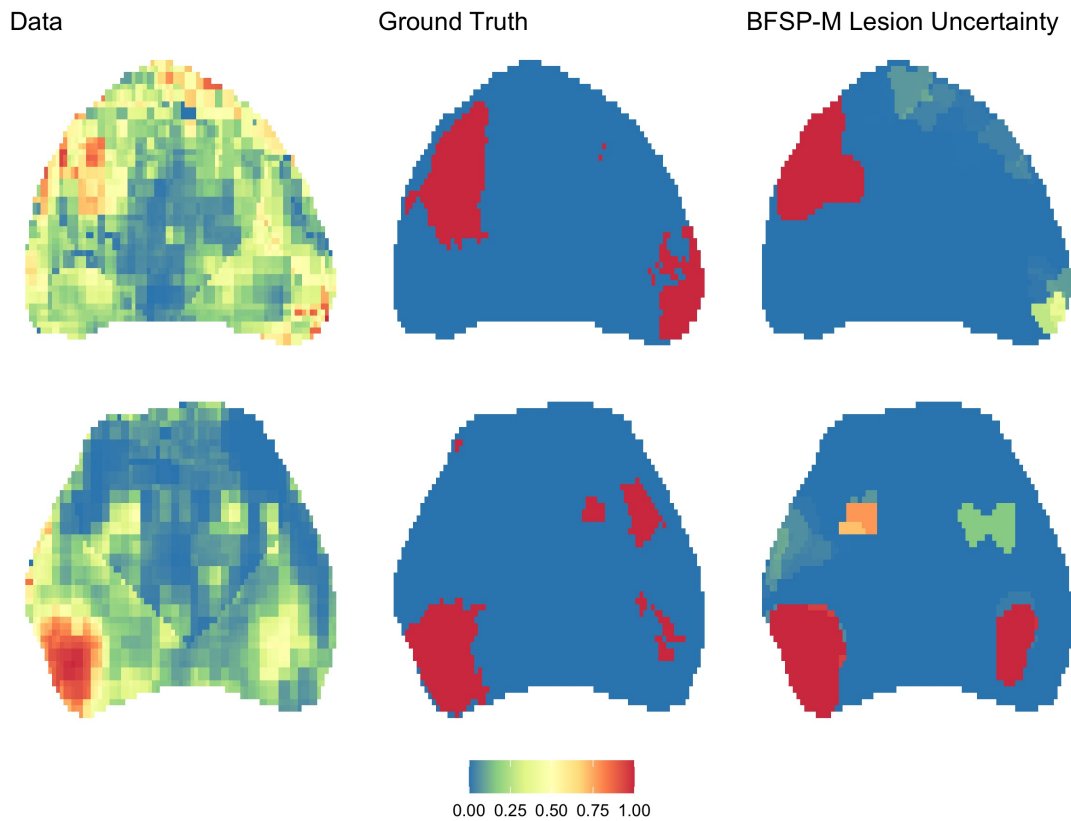
Figure 3.7: From left to right: Map of voxelwise predicted cancer probabilities derived from mpMRI data from 2 slices containing multiple lesions, where color indicates probability of cancer with red close to 1 and blue close to 0 [1], true cancer status, lesion uncertainty of BFSP-M where color indicates the proportion of time that a lesion was included in the MCMC posterior draws post burn-in.

3.6 Discussion

We propose a new method, BFSP-M, to accurately define spatial partitions in spatially registered imaging data containing multiple anomalous zones. The method uses functional estimation tools within multiple separate moving polar systems for boundary estimation. By assuming a minimum of zero boundaries, BFSP-M can also identify data that is generated by a stationary process. We model spatial processes within each region to capture the spatial correlations present in the mpMRI data. Using novel boundary-defined jump steps in RJ-MCMC and likelihood approximations we have developed a computationally efficient and novel method to detect an unknown number of anomalous regions or "hot spots" in imaging data. Unlike competing methods, BFSP-M automatically detects the number of lesions, is flexible to enough detect lesions of arbitrary shape, and is able to evaluate uncertainty in boundary estimation. Further, our novel statistical framework allows for multi-subject and longitudinal analysis.

Extending the methods to detect of an unknown number of regions adds a layer of computational complexity to our previous method, BFSP-1. First, the algorithm requires a longer burn-in period as the birth/death steps slow convergence. Furthermore, adding partitions slows computation of the posterior due to the higher dimensionality of the parameter space. We have implemented two key strategies to ensure BFSP-M is computationally efficient. First, we employ likelihood approximation to speed computation of the posterior. Second, we use adaptive MH sampling to encourage acceptance of proposals in MCMC. Lastly, we sample parameters in blocks to encourage mixing and decrease the number of times the likelihood must be computed. BFSP-M completed 50,000 iterations in an average time of 6.9 hours for the simulated data containing 1600 voxels.

With a birth and death process, we found that our method tended to overestimate the number of boundaries needed to partition the space into regions of local homogeneity. Without the ability to merge two boundaries, the MCMC could converge on a local maximum of the posterior where one target region is being estimated to be two separate GPs with very similar estimates of the spatial distributions. The merge step allows the MCMC to jump to a solution in which one boundary encloses both regions with one GP. With the implementation of the split and merge steps, BFSP-M estimates the true

number of boundaries in 84% of datasets.

One of the advantages of the functional tools used to estimate the boundaries is that they lend themselves well to theoretical extensions. There are several avenues to further extend our methods. The boundary function can be extended to be a function of two angles in the spherical coordinate system allowing estimation of boundary surfaces in three-dimensional images. Leveraging a person's entire set of imaging data rather than one slice at a time would improve localization of prostate cancer and allow estimates of cancer volume. The framework of BFSP-1 could also be extended to model lesion boundaries over time to track disease progression. We leave these for future study.

Chapter 4

Bayesian Functional Spatial Partitioning for Boundary Surface Lesion Detection using 3D MRI

4.1 Introduction

A patient's set of mpMRI imaging data is made up of several slices which must be evaluated together to determine the location, size, and severity of the cancer. The one-slice-at-a-time approach of our previous methods BFSP-1 and BFSP-M is not ideal. Applying boundary detection to each slice of imaging data does not account for spatial alignment of slices and cannot guarantee contiguity of the resulting boundary surface. We hypothesize that leveraging a patient's entire set of imaging data will improve accuracy and interpretablity compared to applying boundary detection methods to each slice individually.

Existing methods to detect anomalous regions in 3D data fall short. Voxel-wise clustering methods do not ensure contiguous regions and may not properly account for the spatial correlations in the data. BayesImageS is a voxel-wise clustering method that encourages spatial contiguity by a hidden Potts model [38] and can be applied to 3D

datasets. Although contiguity is encouraged it is not guaranteed. Further, uncertainty is measured on the voxel level which is not clinically interpretable in the setting of medical imaging. Tree based partitioning methods, the simplest being classification trees, use linear boundaries to divide an area into regions of local homogeneity. Methods such as treed Gaussian processes take into account the spatial relatedness of the data and can be applied to 3D data, but are not scalable in large n settings [17]. Furthermore, linear partitioning boundaries are not adequately flexible for the application of lesion detection. Boundary or edge detection methods like Wombling [19] may not provide a closed surface. BayesBD does provide a closed boundary but does not extend to 3D data and assumes iid data [20]. To our knowledge there is no existing method which estimates a closed boundary surface containing a region of anomalous data and properly accounts for spatial correlation.

We propose BFSP-3D, a general method to estimate the closed surface surrounding a contiguous, arbitrarily-shaped inner region with a distinct spatial process from the outer area. When applied to our motivating data, BFSP-3D estimates the cancerous lesion surface and volume of the estimated region. By utilizing a Bayesian approach, we can produce estimates of boundary uncertainty using 95% credible surfaces. We model the voxel-wise probabilities with region specific Gaussian processes and estimate their parameters jointly with the boundary surface using MCMC. The boundary surface is estimated using a linear combination of periodic basis functions of two angles in the spherical coordinates. Our novel approach performs better than existing methods in simulation and offers more interpretable estimates of uncertainty. In application to 3D data derived from mpMRI, our boundary estimation approach is flexible to the shape of the true lesion boundaries from pathologist annotations.

The remainder of the paper is organized as follows. First, we describe the methods for boundary surface function estimation and spatial modeling in Section 4.2. Section 4.3 covers our Bayesian computational algorithm. We demonstrate the flexibility of our boundary estimation technique to estimate various volumes in Section 4.4. Section 4.5 provides results of BFSP-3D in an application to 3D prostate voxelwise probability data derived from mpMRI. We conclude with a discussion in Section 4.6.

4.2 Methods

Let $Z(s)$ indicate a scalar variable that is observed at location $s = (x, y, z) \in D$, where D is a region in $\mathcal{R}^3$. Let $\boldsymbol{s} = (s_1, ... s_n)$ and $\boldsymbol{Z} = (Z(s_1), ..., Z(s_n))$ denote a set of locations within the region D and the corresponding set of observations. We assume that D can be partitioned into a contiguous target region D_1 and an outer region D_0 by a closed boundary surface B. Let $\boldsymbol{s}_0$ and $\boldsymbol{s}_1$ be the vectors of spatial locations within regions D_0 and D_1 respectively, and $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$ the vectors of observed values at those locations. We assume $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$ follow different spatial processes. Our goal is to estimate the boundary surface B that partitions the data into two regions based on heterogeneity in the underlying spatial processes with minimal restrictions on the shape of the target region. Estimation is accomplished with MCMC to estimate the boundary parameters and the spatial parameters based on their posterior probabilities given the data. Bayesian estimation allows quantification of the uncertainty of the boundary surface with 95% point-wise credible surfaces. Details on approximating the boundary surface by functional estimation are provided in Section 4.2.1, details on modeling the spatial distributions within each region is provided in Sections 4.2.2 with the priors of the parameters specified in 4.2.3, the details of Bayesian computations and boundary uncertainty estimation are given in Section 4.3.

4.2.1 Defining a Partition

Given a point $c = (x_c, y_c, y_c)$ located at the centroid of D_1, the boundary surface B can be uniquely modeled as a function of two angles in a spherical coordinate system centered at c: $B = \{s \in D : d_c(s) = f(\theta_s, \phi_s | c), \theta_s \in [0, \pi], \phi_s \in [0, 2\pi]\}$, where $d_c(s)$ is the distance from the point s to the polar center c, θ_s is the polar angle measured between c and s, and ϕ_s is the azimuthal angle measured between c and s. Modeling the boundary curves is equivalent to estimating the location of c and function $f(\theta, \phi | c)$. Each spatial location s can then be assigned to either D_0 or D_1 by its distance $d_c(s)$ referenced to centroid c.

The boundary functions are estimated by the linear combination of basis functions in spherical coordinates. In the 2D setting we used the following linear combination of

Fourier basis functions in polar coordinates for f [42]:

$$f(\phi_s|c) \approx \beta_0 + \sum_{k=1}^{K} \{\beta_{-k}\sin(k\phi_s) + \beta_k\cos(k\phi_s)\} \text{ for } 0 \le \phi_s \le 2\pi,$$

To span a 3D space, we make the following expansion of the β's:

$$f(\phi_s, \theta_s|c) \approx \beta_0(\theta_s) + \sum_{k=1}^{K} \{\beta_{-k}(\theta_s)\sin(k\phi_s) + \beta_k(\theta_s)\cos(k\phi_s)\} \text{ for } 0 \le \theta_s \le \pi, 0 \le \phi_s \le 2\pi$$

where $\beta_0(\theta)$, $\beta_k(\theta)$, and $\beta_{-k}(\theta)$ are expanded by another linear combination of basis functions. To ensure continuity at the poles we use the following expansion from Cheong [57]:

$$\beta_k(\theta) = \begin{cases} \sum_{l=0}^{K} \beta_{l,k}\cos(l\theta) & \text{for } k = 0 \\ \sum_{l=1}^{K} \beta_{l,k}\sin(l\theta) & \text{for } k \text{ odd} \\ \sum_{l=1}^{K} \beta_{l,k}\sin(\theta)\sin(l\theta) & \text{for } k \text{ even} \end{cases}$$

The parameter K corresponds to the level of detail in the boundary surface. A lower value will produce a smoother surface. In our previous work on 2D partitioning, we found that $K = 5$ was adequate for lesion detection in our data. This corresponded to $2K + 1 = 11$ total boundary parameters. In the 3D context, there are $2K^2 + K + 1$ total boundary surface parameters, which becomes unmanageable for increasing values of K. We continue to use $K = 5$ in simulation and data analysis. However, the value of K can be treated as an additional parameter and chosen by cross-validation by the user for other applications.

4.2.2 Spatial Modeling

We estimate the parameters defining the distribution within each region simultaneously with the estimation of the boundary surface parameters. Each iteration involves transitions between the spherical coordinate boundary surface estimation space and the Cartesian image space. In the Cartesian coordinates we classify s as within region D_1 if s is within the boundary surface, i.e., the distance $d_c(s)$ is less than $f(\phi_s, \theta_s|c)$. We

can calculate the likelihood by the product of 2 independent likelihoods due to the conditional independence of the regions given the spatial partition:

$$L(\boldsymbol{Z}|\boldsymbol{\Phi}) \ = L(\boldsymbol{Z_0}|\boldsymbol{\Phi})L(\boldsymbol{Z_1}|\boldsymbol{\Phi}).$$

When the data exhibit spatial autocorrelation, we assume a distinct stationary Gaussian spatial process in each region. For $j = 0, 1$:

$$\begin{aligned} \boldsymbol{Z_j}|\boldsymbol{\Phi} &\sim & N(\mu_j + \boldsymbol{\xi}_j, \tau^2 I_{n_j}) \\ \xi_j &\sim & N(0, \sigma_j^2 H_j) \end{aligned}$$

where μ_j is the mean specific to region j, and $\boldsymbol{\xi}_j$ is the vector of spatial random effects that follow a region-specific Gaussian distribution with covariance $\sigma_j^2 H_j$. The variance for non-spatial random errors, τ^2, is assumed to be common to all regions.

We model the spatial correlation within each region with an exponential kernel.

$$H_j(s_i, s_i') = \exp(-\gamma r_{i,i'}) \tag{4.1}$$

where $r_{i,i'}$ is the Euclidean distance and γ is the spatial decay parameter which is assumed to be common to all regions. We chose the exponential kernel due to its simplicity and its ability to approximate many other types of spatial correlation structure well. However, other kernels can be specified.

For computational stability, we marginalize over the spatial random effects and model the data in each region by:

$$\begin{aligned} \boldsymbol{Z_j}|\boldsymbol{\Phi} &\sim & N(\mu_j, \Sigma_j) \\ \Sigma_j &= & \sigma_j^2 H_j + \tau^2 I_{n_j} \end{aligned} \tag{4.2}$$

When the data do not exhibit strong spatial autocorrelation or the data becomes very large and computation becomes burdensome, one may opt to drop the spatial random effects from the model and assume distinct independent normal distributions

in each region.

$$\boldsymbol{Z_j}|\boldsymbol{\Phi} \quad \sim \quad N(\mu_j, \sigma_j I) \tag{4.3}$$

We use this strategy in the data analysis, see Section 4.5 for details.

4.2.3 Prior Specification

Let $\boldsymbol{\Phi} = \{c, \boldsymbol{\mu} = (\mu_0, \mu_1), \boldsymbol{\sigma^2} = (\sigma_0^2, \sigma_1^2), \tau^2, \gamma, \boldsymbol{\beta}\}$ be all of the parameters of the Gaussian distributions and boundary surface. The spherical origin c is fixed at the centroid of the target region, the location of which is assumed to be uniform over the space domain D a priori. We set c to be the geometric centroid of D_1 at each iteration and do not estimate c with MCMC. We assign the following prior for $\boldsymbol{\Phi}$ assuming independence of the parameters:

$$
\begin{aligned}
\pi(\boldsymbol{\Phi}) &= \pi(\tau^2)\pi(\gamma)\pi(\boldsymbol{\mu})\pi(\boldsymbol{\sigma^2})\pi(\boldsymbol{\beta}) \\
\text{where } \pi(\boldsymbol{\beta}) &= \pi(\beta_{0,0})\prod_{l=1}^{K}\prod_{k=0}^{K}\pi(\beta_{l,k}) \\
\pi(\boldsymbol{\mu}) &= \pi(\mu_0)\pi(\delta_\mu)
\end{aligned}
$$

For the basis coefficients $\boldsymbol{\beta}$, we assign vague normal priors $N(0,1)$ except for $\beta_{0,0}$, which we assign a uniform prior with support from zero to the approximate radius of the entire image space D. This encourages the surface to have positive volume and not exceed the size of the entire area D. We assign the mean of the outer region μ_0 a vague normal prior $N(0,100)$. The difference in means between the target region and the outer region, $\delta_\mu = \mu_1 - \mu_0$ is assumed to follow a vague half-normal prior. We use prior information that the mean of the anomalous region is elevated as compared to its complement. This can be modified for other applications. Variance parameters, σ_j^2 and τ^2, are assigned vague inverse Gamma priors $IG(.1, .1)$. The spatial decay parameter, γ, is assumed to follow an informative prior $IG(3, .5)$. This follows the advice of Banerjee et al. [32], who recommend vague priors for the variance parameters σ_j^2 and τ^2 and an informative prior for spatial range parameter γ.

4.3 Bayesian Computation

The parameters of the boundary surface and spatial distributions are estimated using Metropolis-Hastings (MH). At each iteration a new set of parameters is proposed and accepted with probability proportional to the ratio of the posterior probability at the proposed and current parameter values. We can calculate the posterior probability of the parameters given the current centroid by:

$$\pi(\boldsymbol{\Phi}|\boldsymbol{Z}) \quad \propto \quad L(\boldsymbol{Z}|\boldsymbol{\Phi})\pi(\boldsymbol{\Phi})$$

where $\pi(\boldsymbol{\Phi})$ is given in Equations 4.4-4.4. The form of $L(\boldsymbol{Z}|\boldsymbol{\Phi})$ is provided in Section 4.2.2. Approximations to the likelihood for computational efficiency are discussed in Section 4.3.2. Due to the high dimensional parameter space, we adopt an adaptive proposal density and use blocking to aid in mixing and maintain a high acceptance rate. These strategies are further discussed in Section 4.3.1. We estimate the final boundary surface and its uncertainty in Section 4.3.3.

4.3.1 MCMC Algorithm

Let $\boldsymbol{\Phi}$ be the parameters of the boundary surface and spatial GPs. We divide the parameter space into $K + 2$ blocks: $\Phi_1 = \{\tau^2, \gamma, \boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$, $\Phi_2 = \{\beta_{0,0}, \cdots, \beta_{K,0}\}$, $\Phi_3 = \{\beta_{1,1}, \cdots, \beta_{K,1}\}$,...,$\Phi_{K+2} = \{\beta_{1,K}, \cdots, \beta_{K,K}\}$. After initialization, the generating algorithm is as follows. For $m = 1, ..., K + 2$:

1. Draw $\boldsymbol{\Phi}_m^*$ at iteration t from proposal density $Q_t(\boldsymbol{\Phi}_m^*|\boldsymbol{\Phi}_m)$

2. Accept $\boldsymbol{\Phi}_m^*$ with probability $\alpha = \min\left(1, \frac{\pi(\boldsymbol{\Phi}_m^*|\boldsymbol{Z})}{\pi(\boldsymbol{\Phi}_m|\boldsymbol{Z})}\right)$

3. Update c^* to be the centroid of the shape defined by $\boldsymbol{\Phi}$ and c.

4. Update $\boldsymbol{\beta}$ based on new centroid c^* by solving $\boldsymbol{d}(B_t, c^*) = X_{c^*}\boldsymbol{\beta}$ for $\boldsymbol{\beta}$ where $\boldsymbol{d}(B_t, c^*)$ is $n \times 1$ vector of distances between the boundary points B_t and new centroid c^* and X_{c^*} is the $n \times (2K^2 + K + 1)$ matrix of basis functions referenced to c^*.

The proposal density $Q_t(\boldsymbol{\Phi}_m^*|\boldsymbol{\Phi}_m)$, based on an example from Roberts et al. [33], depends on the iteration number t:

$$\begin{aligned}
Q_t(\boldsymbol{\Phi}_m^*|\boldsymbol{\Phi}_m) &\sim& N_d(\boldsymbol{\Phi}_m, (.1)^2 I_d/d) \text{ for } t < 50 \\
Q_t(\boldsymbol{\Phi}_m^*|\boldsymbol{\Phi}_m) &\sim& (1-\delta)N_d(\boldsymbol{\Phi}_m, \exp(l_{mt})\boldsymbol{\Sigma}_{mt}) + \delta N_d(\boldsymbol{\Phi}_m, (.1)^2 I_d/d) \text{ for } t \geq 50,
\end{aligned}$$

where d is the dimension of $\boldsymbol{\Psi}_m$, δ is a small positive constant (we set $\delta = .01$), and $\boldsymbol{\Sigma}_{mt}$ is the empirical estimate of the covariance structure of the block of parameters at iteration t based on the run so far. The step size of the proposal distribution is controlled by l_{mt} which is adjusted to achieve the desired acceptance rate of about 40% every 20 iterations by either adding or subtracting 0.1 to the previous value.

4.3.2 Marginal Likelihood Approximation Methods

3D imaging data has a very large number of voxels. Calculating the likelihood of the assumed model is prohibitively computationally intensive with increasing sample size. Therefore, we adopt two strategies to increase the computational efficiency of our method in the simulation study.

First we assume a sparse covariance matrix with tapering. This is accomplished by multiplying the covariance kernel $H(s_i, s_i')$ by a tapering kernel $C_\nu(s_i, s_i')$. The Wendland family of tapered covariance functions [34] are widely used. We assume the form of

$$C_\nu(s_i, s_{i'}) = \left(1 - \frac{h}{\nu}\right)_+^4 \left(1 + 4\frac{h}{\nu}\right), \tag{4.4}$$

where $h = ||s_i - s_{i'}||$ and we assume the range parameter $\nu = .1$. This yields our approximation of the covariance:

$$\Sigma_j = \sigma_j^2 H_j \odot T_j + \tau^2 I_{n_j},$$

where T_j is the $n_j \times n_j$ matrix with the (i, i')-th entry given by Equation 4.4.

We make a further simplification to the likelihood by dividing D_0 into eight roughly equally sized and independent zones. The mean x, y, and z coordinates of $s \in D_0$ at a

given iteration define the boundaries which split the area into $\boldsymbol{Z_{0,1}}, ..., \boldsymbol{Z_{0,8}}$. Thus, we assume:

$$L(\boldsymbol{Z}|\boldsymbol{\Phi}) \quad \approx L(\boldsymbol{Z_1}|\boldsymbol{\Phi}) \prod_{l=1}^{8} L(\boldsymbol{Z_{0,l}}|\boldsymbol{\Phi})$$

By breaking up the large covariance matrix into smaller components, this step significantly decreases computational burden. In simulation we found this assumed likelihood approximates the full likelihood sufficiently well.

4.3.3   Posterior Boundary Estimation & Uncertainty

The posterior samples require post processing due to the center not being fixed over MCMC iterations. Therefore, we must compute the mode of the centroid draws after burn-in and for all samples with centroid not equal to the mode, recompute the boundary surface with respect to the mode centroid.

The modal centroid, c_T, is computed using the post burn-in samples. Then, for samples with centroid not equal to the mode, $\boldsymbol{\beta}_t$ we compute the boundary parameters corrected to centroid c_T, $\boldsymbol{\beta}_{t,c_T}$, by solving,

$$\boldsymbol{d}(B_t, c_T) = X_{c_T} \boldsymbol{\beta}_{t,c_T}$$

for $\boldsymbol{\beta}_{c_T}$ where $\boldsymbol{d}(B_t, c_T)$ is an $n \times 1$ vector of distances between boundary points B_t and c_T, and X_{c_T} is the $n \times (2K^2 + K + 1)$ matrix of basis functions where θ is measured to c_T. Then we can compute the boundary surface vector $\boldsymbol{f}_{t,c_T}$ by:

$$\boldsymbol{f}_{t,c_T} = X \boldsymbol{\beta}_{t,c_T}$$

where X is a $2500 \times (2K^2 + K + 1)$ vector of basis functions based on 50 equally spaced angles between 0 and 2π and 50 equally spaced angles between 0 and π. The final estimate of a boundary is given by the mean of $\boldsymbol{f}_{t,c_T}$ over all t.

We compute the 95% credible surfaces using the corrected boundary surfaces $\boldsymbol{f}_{t,c_T}$ referenced to c_T as in Li et al. [27]. For all, $t = 1, ..., T$ MCMC samples post burn-in,
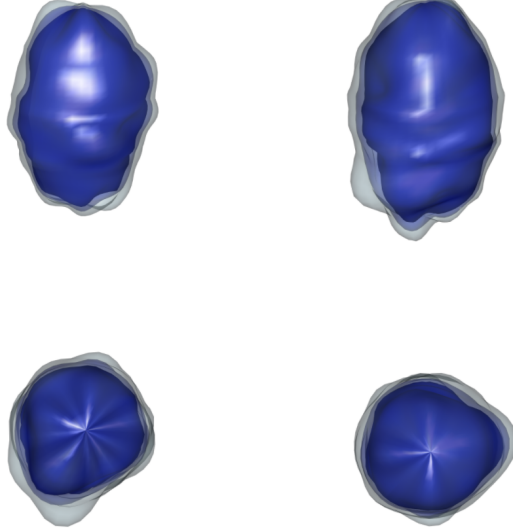
Figure 4.1: Credible surface of BFSP-3D for one simulated dataset containing an ellipsoid target region. The solid blue surface represents the lower 95% credible surface. The transparent blue surface is the point estimate of the surface. The transparent grey surface is the upper 95% credible surface.

we compute.

$$u_t = \sup_{\phi,\theta}\{|f_{t,c_T}(\phi,\theta) - \hat{f}_{c_T}(\phi,\theta)|/\hat{s}(\phi,\theta)\},$$

where $\hat{f}_{c_T}(\phi,\theta)$ and $\hat{s}(\phi,\theta)$ are the posterior mean and standard deviations of $\{f_{t,c_T}(\phi,\theta)\}_{t=1}^{T}$. A 95% credible interval for $f(\phi,\theta)$ is given by:

$$[\hat{f}_{c_T}(\phi,\theta) - L\hat{s}(\phi,\theta), \hat{f}_{c_T}(\phi,\theta) + L\hat{s}(\phi,\theta)],$$

where L is the 95th percentile of the u_t's.

Figure 4.1 displays a boundary surface estimated with BFSP-3D with 95% credible surfaces. The data contained an ellipsoidal inner target region. See Section 4.4 for details on how the data were generated.

4.4   Simulation Study

We evaluate the classification accuracy and statistical properties of BFSP-3D by sim-
ulation. We compare the performance of BFSP-3D with existing clustering, spatial
partitioning, and image segmentation methods. For the simulated data we consider a
25 x 25 x 25 resolution unit cube image space D. We generate data for the inner and
outer regions from independent Gaussian processes with Matérn covariance structures.
For $j = 0, 1$,

$$
\begin{aligned}
\boldsymbol{Z_j}|\boldsymbol{\Phi} &\sim & N(\mu_j, \Sigma_j), &\quad \Sigma = \sigma_j^2 H(\phi_j) + \tau^2 I_{n_j} \\
(H(\phi_j))_{i,i'} &= & H(\phi_j; d_{i,i'}) = \frac{2^{1-\nu_j}}{\Gamma(\nu_j)} \left(\sqrt{2\nu_j}\frac{d_{i,i'}}{\rho_j}\right)^{\nu} \mathcal{K}_{\nu_j}\left(\sqrt{2\nu_j}\frac{d_{i,i'}}{\rho_j}\right)
\end{aligned}
$$

where $d_{i,i'}$ is the Euclidean distance between spatial locations s_i and $s_{i'}$, Γ is the gamma
function, $\mathcal{K}_\nu$ is the modified Bessel function, ν is a smoothness parameter, and ρ is a
spatial range parameter.

Each simulated dataset contains an outer region generated from the above GP with
Matérn covariance. The inner region is one of six shapes of approximately equal volume:
rectangle, hyperboloid, ellipse, "X", cone, or cross. The parameters defining the outer
region are consistent for all simulated datasets, $\mu_0 = 0, \nu_0 = 1, \sigma_0^2 = 1$. For the inner
region, $\mu_1 = 2$ and $\sigma_1^2 = 1$ for all simulated datasets. We randomly generate 20 datasets
for each of the six shapes, and consider four scenarios, which vary ν_1 and τ^2 as follows:

1. $\nu_1 = 1.5, \tau^2 = .1$

2. $\nu_1 = .5, \tau^2 = .1$

3. $\nu_1 = 1.5, \tau^2 = .5$

4. $\nu_1 = .5, \tau^2 = .5$

For the simulation study, we fit the spatial model given by Equation 4.2 with approxima-
tions given in Section 4.3.2. We initialize the MCMC chain with a spherical boundary
surface centered at the voxel with highest intensity. We initialize the parameters of the
GP at $\mu_0 = 0, \mu_1 = 1, \gamma = .1$ and all variance parameters set to $\mathrm{Var}(\boldsymbol{Z})/2$. We collect

12000 MCMC samples and remove the first 4000 as burn-in samples. To increase computational efficiency we divide the outer region into eight zones and approximate the likelihood of the entire outer region by aggregating the likelihoods of the eight zones. In simulations this did not impact the accuracy of our method but significantly increased computational efficiency.

The performance of BFSP-M is evaluated in terms of the sensitivity, specificity, and the Dice coefficient of voxel-level classification. We compare these metrics with those of four competitive methods, two of which are basic clustering methods, K-means (KM) with 2 clusters using the function kmeans from the R Stats package [36] and a two-stage CART method (CART+KM). The two-stage method involves first estimating the partitions using CART with [37]. This results in a tiling of the space. The second stage is to group the partitions into two clusters by KM. The cluster of tiles with the higher mean represents the target regions. We also compare to an image segmentation method BayesImageS (BIS) [15] using R-package "bayesImageS" function mcmcPotts based on a 6 neighbours structure where the neighbours of a vertex comprise its available N,S,E,W, upper and lower adjacencies. We again set the number of clusters at 2. The final competing method is a Gaussian mixture model (GMM) with 2 regions using R function Mclust() from the mclust package [58].

We present summaries of the results by shape and method in Figure 4.2. The boxplots show distributions of the sensitivity, specificity, and Dice coefficient in detecting the inner target region across 80 simulated datasets from the four scenarios for a given shape. Varying the variance parameters of the GP did not have a large effect on the accuracy of BFSP-3D. BFSP-3D maintains reasonably high Dice scores accross the different shaped target regions. We perform best when the inner region is defined by a smooth and concave shape like the ellipse and worse for shapes that contain concavities and/or straight edges like the "X". Voxel-wise clustering methods KM and GMM tend to have high specificities because they do not account for the spatial noise in the outer region. The two stage method divides the space into rectangular volumes and is not adequate for complex shapes. Image segmentation method BIS maintains a similarly good Dice coefficient on average and achieves a higher level of sensitivity due to no restrictions on the shape of the partitioning boundary. However, the method occasionally suffers from extremely low specificity and does not guarantee contiguous
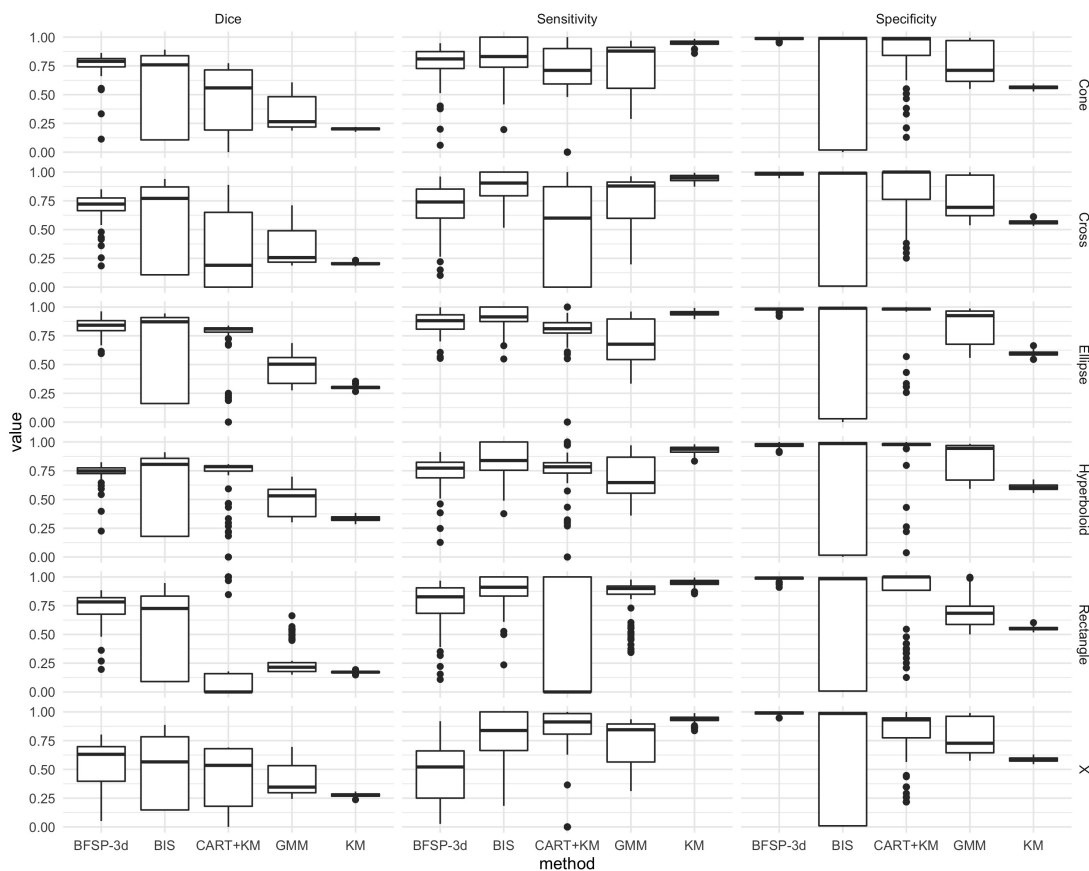
Figure 4.2: Distributions of sensitivity, specificity, and Dice coefficient of BFSP-3d, BayesImageS, Two-Stage CART, Gaussian Mixture Model, and K-means over 80 simulated data sets for each shape.

clusters. Overall, BFSP-3D achieved a mean Dice score of 0.71, a mean sensitivity of 0.72, and a mean specificiy 0.98. The best competitor, BIS, achieved averages of 0.57, 0.86, and 0.65 respectively.

4.5 Data Analysis

The motivating data are comprised of 34 patients who received an MRI prior to undergoing prostatectomy as treatment for prostate cancer, presented by Metzger et al. [2]. Each set consists of the mpMRI imaging parameters measured at 21,000 to 53,000

voxels distributed across 11 to 16 imaging slices extending from apex to base of the prostate. After surgery, the prostates were sectioned in planes that matched in vivo imaging and digitized, allowing the study pathologist to annotate the cancerous region.

For one slice per patient, the pathologist labeled region is co-registered with imaging data. Thus for each patient, we have one slice out of the 11 to 15 total imaging slices that contains the ground truth cancer status. For one 3D dataset, we use the 34 co-registered slices to train the "mregion" model from from Jin et al. [1] holding out the slice contained in the 3D data. Then we calculate the posterior probability of cancer at each of the voxels in the 3D data. This provides voxel-wise probability estimates to which we will apply BFSP-3D to partition the prostate into two regions.

From the 34 datasets, we choose two which had the best congruence between the voxel-wise probability and the true cancer status. Imaging is less reliable toward the apex and base, so we exclude the first and the last slices of data for each dataset chosen. We omit voxels containing missing mpMRI parameters. The data are scaled such that all voxels fall within the unit cube.

Developing a method that is computationally feasible for our clinical collaborators is of utmost importance. Due to the large size of the data, we assume the voxel-wise probabilities follow distinct independent normal distributions in each region as in Equation 4.3. This drastically improves the computational time making this method clinically usable. Since the strength of the signal relative to noise is quite strong in these data, we do not sacrifice much power when we make this assumption. In simulations, we show little to no loss of accuracy in the presence of a large mean difference relative to the level of noise. See Figure C.1 for details.

We initialize BFSP-3D with a large sphere centered at the centroid of the entire region with means 0 for the outer region and 1 for the inner region. The variance parameters are initialized to be half the total variance of the voxel-wise probabilities. We collect 50,000 samples of the posterior and discard the first 20,000 as burn-in samples.

Figures 4.3 and 4.4 display the partitioning results of BFSP-3D on data from subject 118. Figures 4.5 and 4.6 display the partitioning results of BFSP-3D on data from subject 111. In each figure, the first column shows the voxelwise probabilities of cancer at each slice - the data to which we apply our method. The color represents the probability with red close to 1 and blue close to zero. The second column shows the estimated

cluster membership by BFSP-3D where red indicates that the voxel was included in the cancerous lesion boundary. The third column shows the pathologist annotated slices where available. The blue or purple boundaries indicate the presence of cancer. The co-registered 2D imaging data do not perfectly align with the 3D imaging data. For this reason, we cannot measure the accuracy of our results with respect to the ground truth. Visually, boundaries estimated by our method align with the voxel-wise probability data quite well. BFSP-3D is able to identify the boundary which separates the region of high cancer probability from the surrounding healthy tissue.

4.6   Discussion

We present a novel method to partition a 3D dataset exhibiting spatial non-stationarity due to the presence of an anomalous region located within a larger homogenous space. In simulation, our method outperforms existing image segmentation, spatial partitioning, and clustering methods. Despite simplifications to the likelihood which ignore spatial autocorrelation in the data, our method can accurately capture the lesion boundary from mpMRI derived imaging data. This is due to the relative strength of the signal of the cancerous regions. A more substantial degradation in performance would be expected in the presence of a smaller signal to noise ratio. In simulation, we showed this assumption may cause loss of power when the difference in means between the two regions is small. See Figure C.1 for details.

Due to the protocol for sectioning and mapping the prostate to assess pathology, the resolution of the data in the Z-direction is much lower than in the X-, an Y- directions. When data are standardized to fall within a 1x1 cube, the neighboring voxels in the Z-direction are .1 units apart. Neighboring points in one XY plane are about 10 times closer. This quirk of the data does not preclude the application of BFSP-3D.

Extending the boundary to be a function of two rather than one angle increases the number of boundary coefficients significantly. We used 11 total boundary coefficients in BFSP-1 and to achieve a similar level of detail in BFSP-3D would require 56 boundary coefficients. High dimensional and correlated parameters are known to cause difficulties in typical MH algorithms. To overcome this challenge, we block the boundary

Figure 4.3: For subject 118 slices 1-5: **(a)** Voxelwise probabilities of cancer [1] (red=1, blue=0) **(b)** Clustering results from BFSP-3D **(c)** Pathologist annotated slices [2]
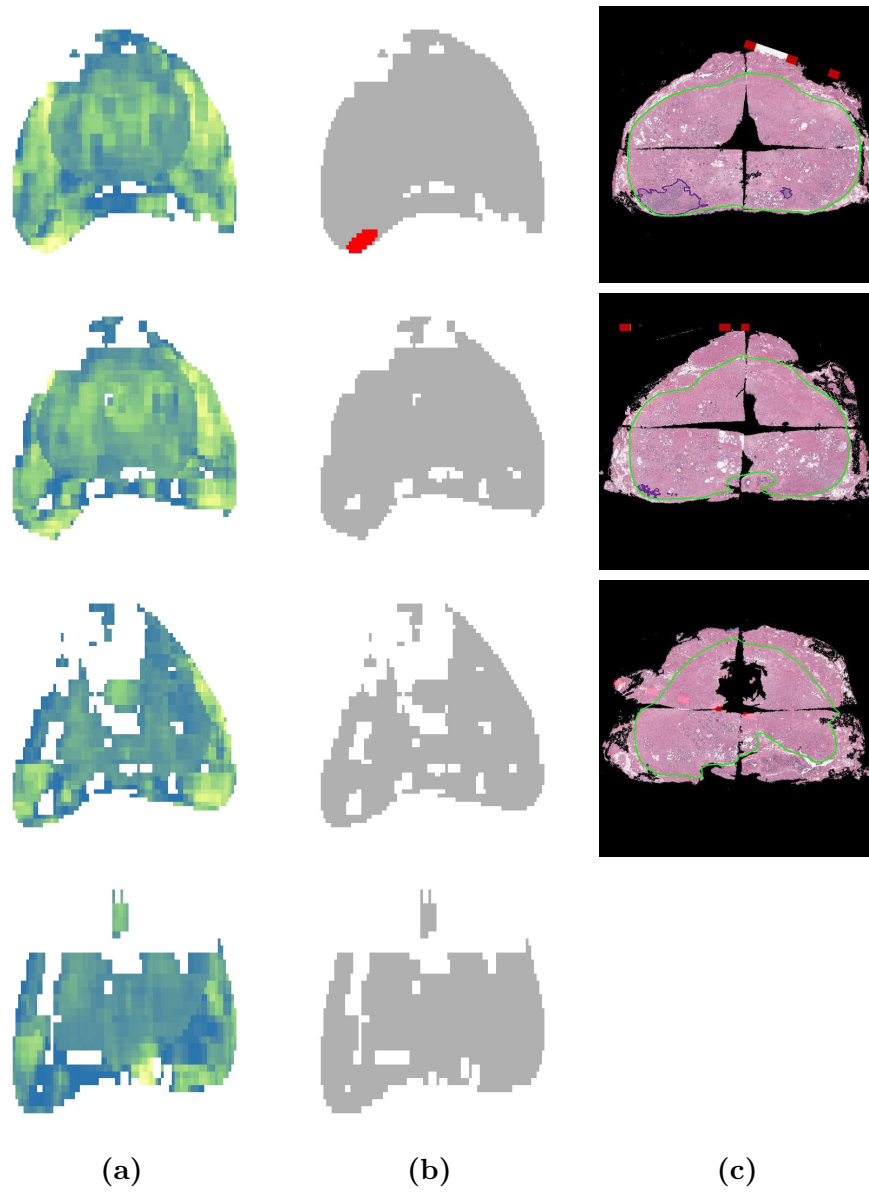
Figure 4.4: For subject 118 slices 6-9: **(a)** Voxelwise probabilities of cancer [1] (red=1, blue=0) **(b)** Clustering results from BFSP-3D **(c)** Pathologist annotated slices [2]
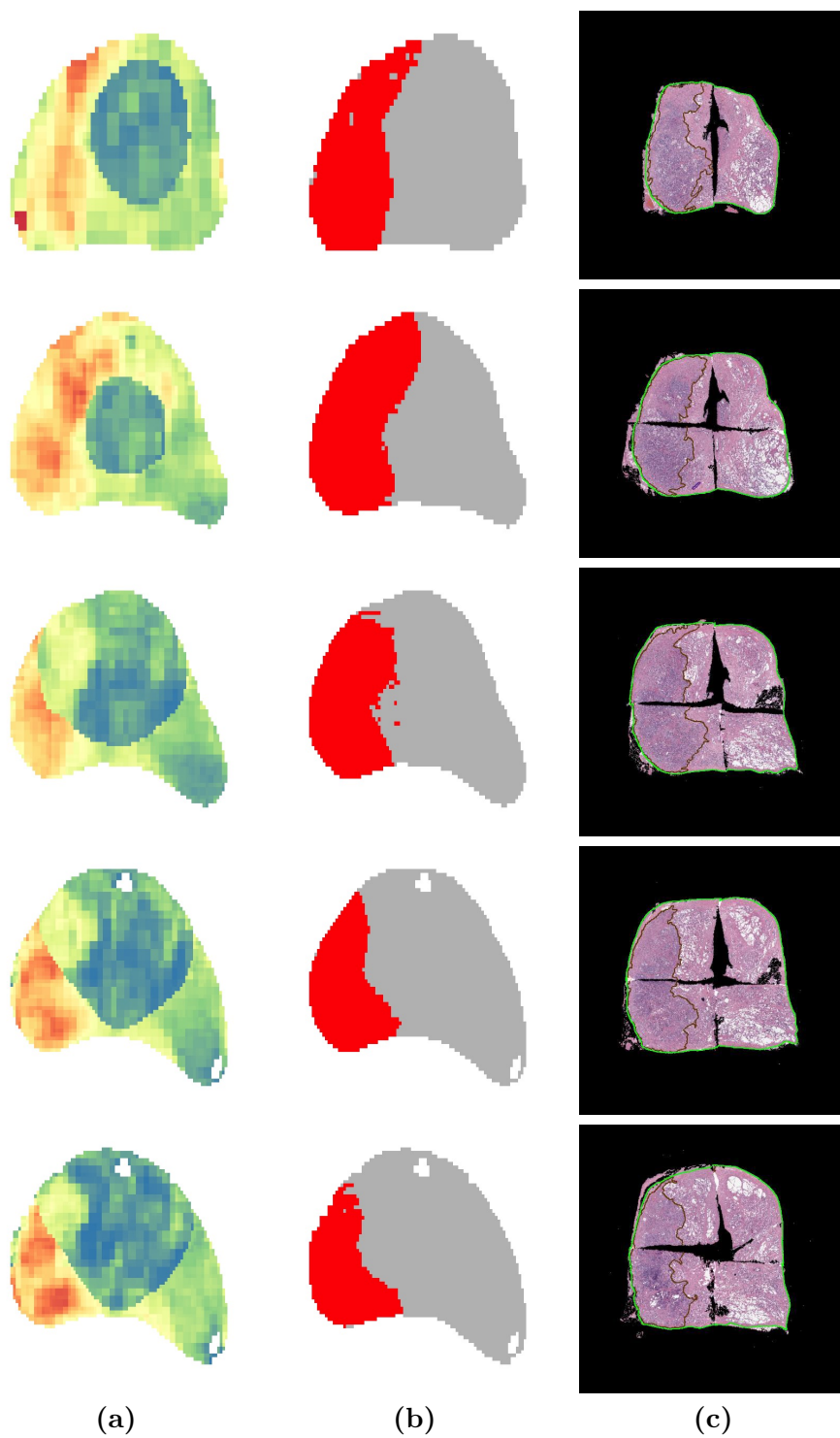
Figure 4.5: For subject 111 slices 1-5: **(a)** Voxelwise probabilities of cancer [1] (red=1, blue=0) **(b)** Clustering results from BFSP-3D **(c)** Pathologist annotated slices [2]

(a)　　　　　　**(b)**　　　　　　**(c)**
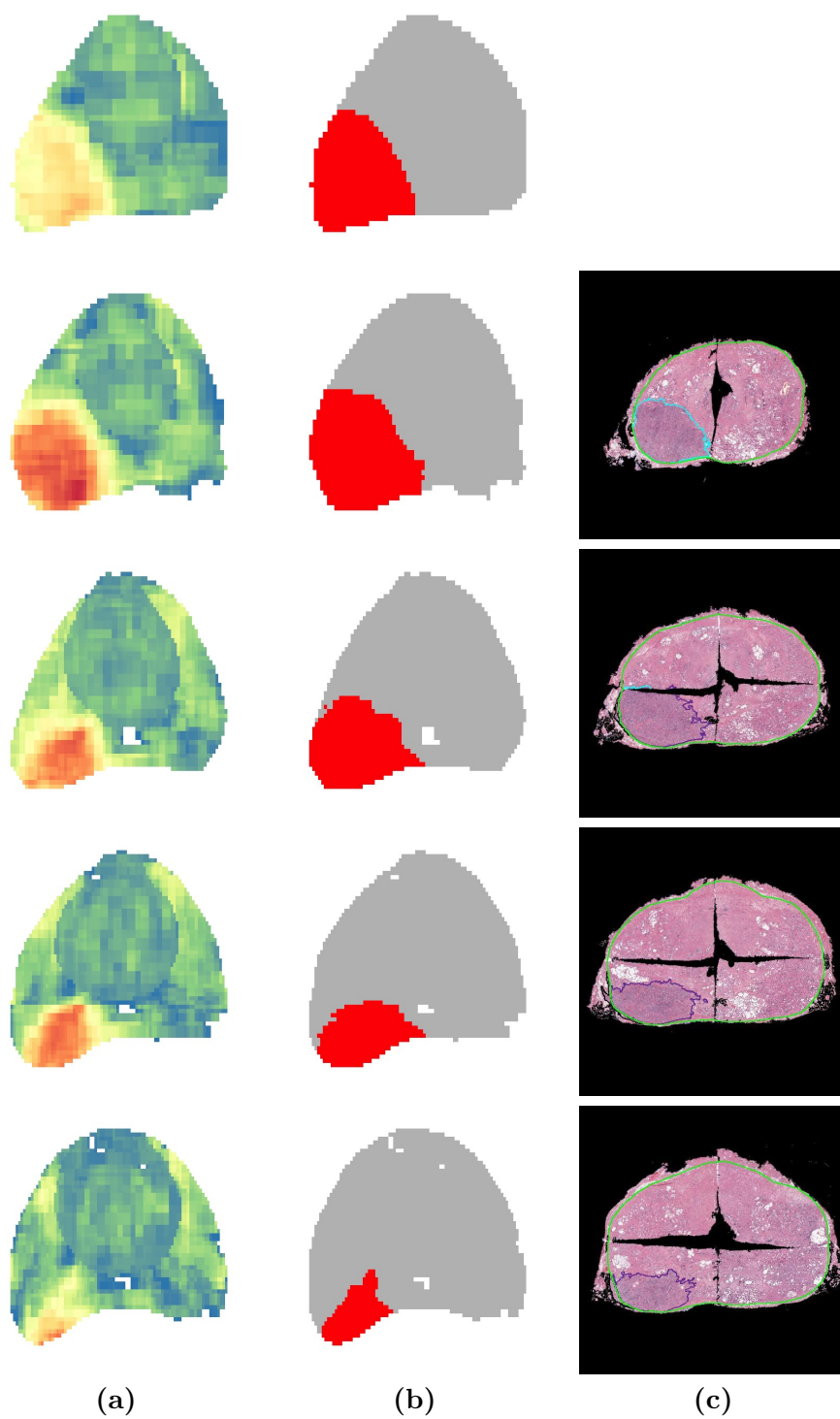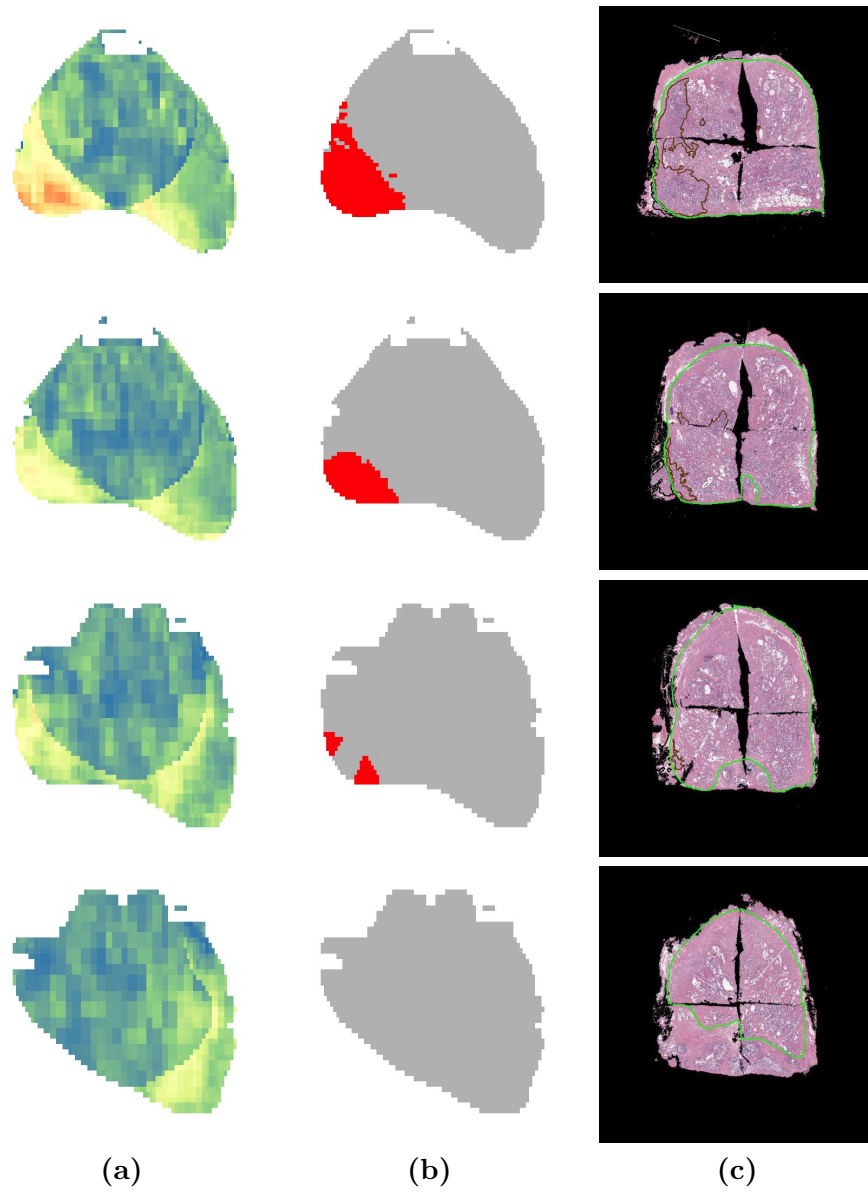
Figure 4.6: For subject 111 slices 6-9: **(a)** Voxelwise probabilities of cancer [1] (red=1, blue=0) **(b)** Clustering results from BFSP-3D **(c)** Pathologist annotated slices [2]

parameters into 11 approximately equally sized blocks. We also utilize adaptive proposal densities which learns the correlation among the blocks and adjusts the step size to increase acceptance rates. Although in theory the value of K could be increased to model greater detail in the lesion boundary, doing so is not advised as the computational burden becomes too intense.

The dimension of the data also increases in the context of 3D imaging data. Modeling the spatial correlation of the data becomes prohibitively computationally expensive as n becomes large. Likelihood approximations increase efficiency but the method remains computationally intensive in the context of large 3D data. Another limitation of this method is the assumption that there is one cancerous lesion or anomalous area. We previously developed a 2D segmentation algorithm that allows for an unknown number of regions, and future will focus on extending BFSP-3D to allow for an unknown number of regions, as well.

Chapter 5

Conclusion

In this dissertation, we have developed novel methods to estimate partitions of a 2D or 3D image space exhibiting non-stationarity due to the presence of anomalous regions or "hot spots." In Chapter 2, we proposed a general framework for boundary estimation using functional tools that can be adapted and extended for various applications. Our spatial modeling framework is similarly flexible and can be adapted by changing the covariance kernel or assuming an independence structure. Though we assume the "hot spot" is defined by a heightened mean, users may adapt our method to search the image for regions of lower mean, or even regions with different variability. In Chapter 3, we extended the general framework of our first method to develop a method for discovering an unknown number of lesions or "hot spots." This method is a clinically useful tool which can identify the number, location, size, and shape of possible cancerous lesions. Finally, in Chapter 4, we built upon the framework laid out in BFSP-1 to extend our method to the 3D image space. Each iteration of this project has prioritized the applicability and interpretability of its clinical use as a tool for non-invasive prostate cancer lesion detection using mpMRI data.

Our novel framework for boundary estimation offers added flexibility over existing methods, which assume linear or piecewise-linear boundaries. Any shape in the star domain can be easily modeled by our method. This has proven to be particularly useful for our application to PCa lesion boundary modeling. Furthermore, although Fourier bases are used for our motivating application, our method allows customization of boundary characteristics by specification of other type and number of basis functions

to suit various problems that require spatial partitioning. As shown in Chapter 4, the framework allows seamless extension to three dimensions.

Another key advantage of our approach is the quantification of uncertainty. Competing methods do not provide clinically relevant estimates of the variability of the estimated lesions. Voxel-wise clustering methods may only provide voxel-level uncertainty, if at all. Our approach provides credible bands showing possible ranges of the cancer. BFSP-M can also estimate lesion specific uncertainty, which measures the proportion of time that our algorithm identified each region as distinct from the healthy tissue.

The main challenge we faced and continue to face as we develop extensions is heavy computational burden. Each iteration of our project added a layer of computational complexity. Computation of the likelihood in the presence of spatial autocorrelation can be prohibitively time consuming due to large dense covariance matrices. To increase efficiency we adopted likelihood approximations such as tapering and partitioning the data. Our boundaries are defined by $2K + 1$ parameters in BFSP-1, $(2K + 1) * M$ in BFSP-M where M is the number of lesions, and $2K^2 + K + 1$ parameters in BFSP-3D. K may be chosen but we found that $K < 5$ was not adequate for our motivating data. Thus, the dimension of the parameter space can quickly become prohibitively large as M increases or in the 3D setting. We employed adaptive MCMC strategies to combat low acceptance and poor mixing.

Despite tapering strategies and partitioning of the likelihood, computation for large 3D datasets is still quite time consuming. A quicker approach involves ignoring the spatial autocorrelation in the data. We found that this solution, although not ideal, estimates lesion boundaries quite well in our motivating data due to the strength of the signal of the cancerous regions.

There are several avenues to explore in extending this work. First, the methods of BFSP-M and BFSP-3D may be combined for the purpose of multiple lesion discovery in 3D imaging data. As both methods are computationally intensive in their own right, the combination may require further simplification and approximation of the likelihood. We are also interested in the idea of modeling lesion boundaries over time using longitudinally captured imaging data. This would allow clinicians to track cancer progression or response to treatment over time.

References

[1] Jin Jin, Lin Zhang, Ethan Leng, Gregory J. Metzger, and Joseph S. Koopmeiners. Detection of prostate cancer with multiparametric MRI utilizing the anatomic structure of the prostate. *Statistics in Medicine*, 37(22), 9 2018.

[2] Gregory J Metzger, Chaitanya Kalavagunta, Benjamin Spilseth, Patrick J Bolan, Xiufeng Li, Diane Hutter, Jung W Nam, Andrew D Johnson, Jonathan C Henriksen, Laura Moench, Badrinath Konety, Christopher A Warlick, Stephen C Schmechel, and Joseph S Koopmeiners. Detection of Prostate Cancer: Quantitative Multiparametric MR Imaging Models Developed Using Registered Correlative Histopathology. *Radiology*, 279(3):805–816, 2016.

[3] Rebecca L Siegel, Kimberly D Miller, Hannah E Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: a cancer journal for clinicians*, 1 2022.

[4] Fritz H. Schröder, Jonas Hugosson, Monique J. Roobol, Teuvo L.J. Tammela, Stefano Ciatto, Vera Nelen, Maciej Kwiatkowski, Marcos Lujan, Hans Lilja, Marco Zappa, Louis J. Denis, Franz Recker, Antonio Berenguer, Liisa Määttänen, Chris H. Bangma, Gunnar Aus, Arnauld Villers, Xavier Rebillard, Theodorus van der Kwast, Bert G. Blijenberg, Sue M. Moss, Harry J. de Koning, and Anssi Auvinen. Screening and Prostate-Cancer Mortality in a Randomized European Study. *New England Journal of Medicine*, 360(13):1320–1328, 3 2009.

[5] Virginia A. Moyer. Screening for prostate cancer: U.S. preventive services task force recommendation statement. *Annals of Internal Medicine*, 157(2):120–134, 2012.

[6] Christopher J D Wallis, Masoom A Haider, and Robert K Nam. Role of mpMRI of the prostate in screening for prostate cancer. *Translational andrology and urology*, 6(3):464–471, 6 2017.

[7] Morgan R. Pokorny, Maarten De Rooij, Earl Duncan, Fritz H. Schröder, Robert Parkinson, Jelle O. Barentsz, and Leslie C. Thompson. Prospective Study of Diagnostic Accuracy Comparing Prostate Cancer Detection by Transrectal Ultrasound–Guided Biopsy Versus Magnetic Resonance (MR) Imaging with Subsequent MR-guided Biopsy in Men Without Previous Prostate Biopsies. *European Urology*, 66(1):22–29, 7 2014.

[8] Shijun Wang, Karen Burtt, Baris Turkbey, Peter Choyke, and Ronald M. Summers. Computer aided-diagnosis of prostate cancer on multiparametric MRI: A technical review of current research. *BioMed Research International*, 2014, 2014.

[9] Baowei Fei. Computer-aided diagnosis of prostate cancer with MRI. *Current opinion in biomedical engineering*, 3:20–27, 9 2017.

[10] Jin Jin, Lin Zhang, Ethan Leng, Gregory J Metzger, and Joseph S Koopmeiners. Bayesian spatial models for voxel-wise prostate cancer classification using multiparametric magnetic resonance imaging data. *Statistics in Medicine*, n/a(n/a):1–17, 2021.

[11] Ethan Leng, Benjamin Spilseth, Lin Zhang, Jin Jin, Joseph S Koopmeiners, and Gregory J Metzger. Development of a measure for evaluating lesion-wise performance of CAD algorithms in the context of mpMRI detection of prostate cancer. *Medical Physics*, 45(5):2076–2088, 2018.

[12] Anders Eklund, Thomas E. Nichols, and Hans Knutsson. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113(28), 7 2016.

[13] Guangpu Shao, Junbin Gao, Tianjiang Wang, Fang Liu, Yucheng Shu, and Yong Yang. Image Segmentation Based on Spatially Coherent Gaussian Mixture Model. *2014 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–6, 2014.

[14] Taisong Xiong, l z, and Zhang Yi. Double Gaussian Mixture Model for Image Segmentation with Spatial Relationships. *Journal of Visual Communication and Image Representation*, 34, 9 2015.

[15] Matthew T Moores, Geoff K Nicholls, Anthony N Pettitt, and Kerrie Mengersen. Scalable Bayesian Inference for the Inverse Temperature of a Hidden Potts Model. *Bayesian Analysis*, 2018.

[16] Hyoung-Moon Kim, Bani K Mallick, and C C Holmes. Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes. *Journal of the American Statistical Association*, 100(470):653–668, 2005.

[17] Robert B Gramacy and Herbert K H Lee. Bayesian Treed Gaussian Process Models With an Application to Computer Modeling. *Journal of the American Statistical Association*, 103(483):1119–1130, 2008.

[18] Bledar A Konomi, Huiyan Sang, and Bani K Mallick. Adaptive Bayesian Nonstationary Modeling for Large Spatial Datasets Using Covariance Approximations. *Journal of Computational and Graphical Statistics*, 23(3):802–829, 2014.

[19] Alan E Gelfand and Sudipto Banerjee. Bayesian wombling: finding rapid change in spatial maps. *WIREs Computational Statistics*, 7(5):307–315, 2015.

[20] Nicholas Syring and Meng Li. BayesBD: Bayesian Inference for Image Boundaries, 2017.

[21] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. Chapman & Hall/CRC, Boca Raton, Florida, 1984.

[22] Zhen Zhang, Chae Young Lim, and Tapabrata Maiti. Analyzing 2000-2010 Childhood Age-Adjusted Cancer Rates in Florida: A Spatial Clustering Approach. *Statistics and Public Policy*, 1(1):120–128, 2014.

[23] Jerome H Friedman. Stochastic Gradient Boosting. *Computational Statistics and Data Analysis*, 38:367–378, 1999.

[24] L Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[25] John A Hartigan and Manchek A Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[26] Leonard Kaufmann and Peter Rousseeuw. Clustering by Means of Medoids. *Data Analysis based on the L1-Norm and Related Methods*, pages 405–416, 9 1987.

[27] Meng Li and Subhashis Ghosal. Bayesian detection of image boundaries. *The Annals of Statistics*, 45(5), 10 2017.

[28] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. *J Digit Imaging*, 30(4):449–459, 8 2017.

[29] Chap T Le, Ping Liu, Bruce R Lindgren, Kathleen A Daly, and G Scott Giebink. Some statistical methods for investigating the date of birth as a disease indicator. *Statistics in Medicine*, 22(13):2127–2135, 2003.

[30] Jerrold H Zar. *Biostatistical analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1974.

[31] Julian Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.

[32] Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. CRC press, Boca Raton, Florida, 2014.

[33] Gareth O Roberts and Jeffrey S Rosenthal. Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

[34] H Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

[35] Aaron A King, Dao Nguyen, and Edward L Ionides. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software*, 69(12):1–43, 2016.

[36] R Core Team. R: A Language and Environment for Statistical Computing, 2020.

[37] Brian Ripley. tree: Classification and Regression Trees, 2019.

[38] Matthew T. Moores, Dai Feng, and Kerrie Mengersen. bayesImageS: Bayesian Mehtods for Image Segmentation, 2021.

[39] Chaitanya Kalavagunta, Xiangmin Zhou, Stephen C Schmechel, and Gregory J Metzger. Registration of in vivo prostate MRI and pseudo-whole mount histology using Local Affine Transformations guided by Internal Structures (LATIS). *Journal of magnetic resonance imaging : JMRI*, 41(4):1104–1114, 9 2015.

[40] Anders Eklund, Thomas E Nichols, and Hans Knutsson. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, 113(28):7900–5, 7 2016.

[41] Hashim U Ahmed, Ahmed El-Shater Bosaily, Louise C Brown, Rhian Gabe, Richard Kaplan, Mahesh K Parmar, Yolanda Collaco-Moraes, Katie Ward, Richard G Hindley, Alex Freeman, Alex P Kirkham, Robert Oldroyd, Chris Parker, and Mark Emberton. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *The Lancet*, 389(10071):815–822, 2 2017.

[42] Maria Masotti, Lin Zhang, Ethan Leng, Gregory J Metzger, and Joseph S Koopmeiners. A novel Bayesian functional spatial partitioning method with application to prostate cancer lesion detection using MRI. *Biometrics*, n/a(n/a):1–12, 2021.

[43] Catherine A. Sugar and Gareth M. James. Finding the Number of Clusters in a Dataset: An Information-Theoretic Approach. *Journal of the American Statistical Association*, 98(463):750–763, 9 2003.

[44] S Salvador and P Chan. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *16th IEEE International Conference on Tools with Artificial Intelligence*, pages 576–584, 2004.

[45] Anil K Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[46] David G T Denison, K Mallick, and Adrian F M Smith. A Bayesian CART algorithm. Technical Report 2, 1998.

[47] Sylvia Fruhwirth-Schnatter, Gilles Celeux, and Christian P Robert. *Handbook of Mixture Analysis*. CRC press, 2019.

[48] Raffaele Argiento and Maria De Iorio. Is infinity that far? A Bayesian nonparametric perspective of finite mixture models. 4 2019.

[49] Radford M Neal. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

[50] J.A. Hartigan. Partition models. *Communications in Statistics - Theory and Methods*, 19(8):2745–2756, 1 1990.

[51] Garritt L. Page and Fernando A. Quintana. Spatial Product Partition Models. *Bayesian Analysis*, 11(1), 3 2016.

[52] Jian Kang, Brian J Reich, and Ana-Maria Staicu. Scalar-on-image regression via the soft-thresholded Gaussian process. *Biometrika*, 105(1):165–184, 3 2018.

[53] D. G. T. Denison and C. C. Holmes. Bayesian Partitioning for Estimating Disease Risk. *Biometrics*, 57(1):143–149, 3 2001.

[54] Hugh A Chipman, Hachipman@uwaterloo Ca, Edward I George, and Robert E Mcculloch. Bayesian Treed Models *. Technical report, 2002.

[55] Sylvia Richardson and Peter J. Green. On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 59(4):731–792, 1997.

[56] Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 1995.

[57] Hyeong-Bin Cheong. Double Fourier Series on a Sphere: Applications to Elliptic and Vorticity Equations. *Journal of Computational Physics*, 157(1):327–349, 2000.

[58] Luca Scrucca, Michael Fop, T Brendan Murphy, and Adrian E Raftery. mclust 5: clustering, classification and density

}estimation using Gaussian finite mixture models. *The R Journal*, 8(1):289–317, 2016.

Appendix A

Supplementary Material for Chapter 2

A.1   Simulation Study Details

The simulated data are generated via two distinct Gaussian processes with Matérn covariance as detailed below. For $j = 0, 1$

$$
\begin{aligned}
\boldsymbol{Z_j}|\boldsymbol{\Phi_j} &\sim N(\mu_j, \Sigma_j), \quad \Sigma = \sigma_j^2 H(\phi_j) + \tau^2 I_{n_j} \\
(H(\phi_j))_{i,i'} &= H(\phi_j; d_{i,i'}) = \frac{2^{1-\nu_j}}{\Gamma(\nu_j)} \left(\sqrt{2\nu_j} \frac{d_{i,i'}}{\rho_j} \right)^{\nu} \mathcal{K}_{\nu_j} \left(\sqrt{2\nu_j} \frac{d_{i,i'}}{\rho_j} \right),
\end{aligned}
$$

where $d_{i,i'}$ is Euclidean distance between spatial locations s_i and $s_{i'}$, Γ is the gamma function, $\mathcal{K}_\nu$ is the modified Bessel function, ν is a smoothness parameter, and ρ is a spatial range parameter. We fix the following settings during simulation: $\nu_0 = 1$, $\rho_0 = \rho_1 = .05$, $\mu_0 = -1$, $\mu_1 = 2$, $\sigma_0^2 = \sigma_1^2 = 1$.

The sensitivity is the proportion of points in the true target region that are identified as being in that region. The specificity is the proportion of points in the true outer region that are identified to be in that region. The Dice coefficient is:

$$
\text{DSC} = \frac{2TP}{2TP + FP + FN}, \tag{A.1}
$$

where TP is the number of true positives, FP is the number of false positives, and

FN is the number of false negatives. The value of the Dice coefficient ranges from 0, indicating no spatial overlap, to 1, indicating complete overlap.

For KM, we take the target region to be the cluster with the higher mean. For the CART algorithm, we take the estimated target region to be the partition with the highest mean and disregard other partitions made by the algorithm. For BayesBD we provided the true centroid of the target region which would be unknown in practice and specified the inner region to have a heightened mean. For BayesImageS we specified neighbors via KNN with k=8 and 4 blocks. BFSP spline utilizes periodic B-splines of degree 3. We set the total number of basis functions to be 11 so that the total number of parameters to be estimated is the same as our original implementation. The periodic B-spline basis functions are computed using R package "pomp".
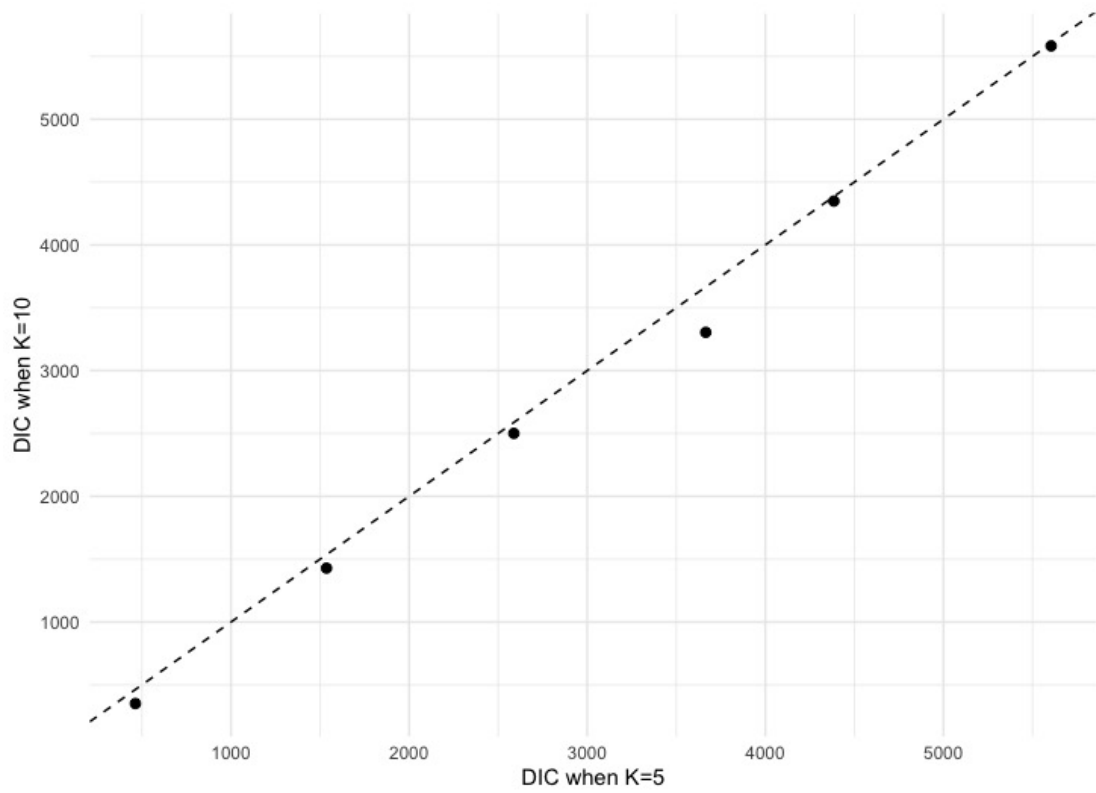
A.2 Supplementary Tables & Figures

Figure A.1: Comparison of DIC when the number of basis functions, K, is 5 or 10 in the data analysis. The dotted line is the line $y = x$.

Table A.1: Results of BFSP and competing methods in simulated data. Mean and SD of sensitivity, specificity, and Dice coefficient for each simulation setting referenced in Section 3 for a square target region where ν_1 and τ^2 are parameters of the simulated Gaussian processes.

Shape	ν_1	τ^2	Method	*Sensitivity* Mean (SD)	*Specificity* Mean (SD)	*Dice* Mean (SD)
Square	0.5	0.1	BFSP	0.981 (0.017)	0.962 (0.042)	0.909 (0.084)
			BFSP Spline	0.976 (0.018)	0.962 (0.059)	0.912 (0.1)
			CART	0.838 (0.27)	1 (0)	0.882 (0.213)
			KM	0.969 (0.019)	0.855 (0.06)	0.718 (0.082)
			BayesBD	0.665 (0.176)	0.997 (0.008)	0.777 (0.124)
			BayesImageS	0.985 (0.016)	0.954 (0.082)	0.905 (0.109)
	0.5	0.5	BFSP	0.969 (0.045)	0.958 (0.055)	0.897 (0.099)
			BFSP Spline	0.969 (0.026)	0.957 (0.054)	0.895 (0.102)
			CART	0.882 (0.23)	1 (0)	0.917 (0.174)
			KM	0.954 (0.021)	0.775 (0.051)	0.613 (0.057)
			BayesBD	0.775 (0.181)	0.992 (0.009)	0.841 (0.125)
			BayesImageS	0.977 (0.027)	0.93 (0.173)	0.889 (0.154)
	1.5	0.1	BFSP	0.968 (0.043)	0.967 (0.046)	0.914 (0.088)
			BFSP Spline	0.967 (0.036)	0.968 (0.05)	0.917 (0.093)
			CART	0.684 (0.301)	1 (0)	0.772 (0.232)
			KM	0.978 (0.028)	0.861 (0.081)	0.739 (0.111)
			BayesBD	0.574 (0.176)	0.999 (0.003)	0.711 (0.139)
			BayesImageS	0.964 (0.057)	0.954 (0.138)	0.914 (0.131)
	1.5	0.5	BFSP	0.949 (0.085)	0.966 (0.048)	0.9 (0.098)
			BFSP Spline	0.959 (0.037)	0.945 (0.071)	0.872 (0.116)
			CART	0.709 (0.31)	1 (0.001)	0.786 (0.245)
			KM	0.962 (0.023)	0.761 (0.075)	0.608 (0.082)
			BayesBD	0.612 (0.227)	0.997 (0.007)	0.724 (0.2)
			BayesImageS	0.947 (0.069)	0.924 (0.203)	0.877 (0.166)

Table A.2: Results of BFSP and competing methods in simulated data. Mean and SD of sensitivity, specificity, and Dice coefficient for each simulation setting referenced in Section 3 for a triangular target region where ν_1 and τ^2 are parameters of the simulated Gaussian processes.

| Shape	ν_1	τ^2	Method	*Sensitivity* Mean (SD)	*Specificity* Mean (SD)	*Dice* Mean (SD)
Triangle	0.5	0.1	BFSP	0.938 (0.035)	0.975 (0.027)	0.868 (0.082)
			BFSP Spline	0.942 (0.041)	0.938 (0.086)	0.787 (0.168)
			CART	0.512 (0.194)	0.996 (0.004)	0.635 (0.188)
			KM	0.992 (0.009)	0.729 (0.075)	0.43 (0.077)
			BayesBD	0.675 (0.177)	0.994 (0.013)	0.766 (0.115)
			BayesImageS	0.985 (0.023)	0.89 (0.193)	0.774 (0.238)
	0.5	0.5	BFSP	0.937 (0.034)	0.957 (0.064)	0.825 (0.131)
			BFSP Spline	0.936 (0.044)	0.916 (0.097)	0.728 (0.172)
			CART	0.515 (0.232)	0.994 (0.008)	0.619 (0.217)
			KM	0.981 (0.016)	0.673 (0.048)	0.374 (0.04)
			BayesBD	0.672 (0.166)	0.992 (0.016)	0.758 (0.109)
			BayesImageS	0.974 (0.025)	0.88 (0.238)	0.779 (0.248)
	1.5	0.1	BFSP	0.92 (0.057)	0.961 (0.055)	0.823 (0.123)
			BFSP Spline	0.918 (0.062)	0.93 (0.088)	0.753 (0.17)
			CART	0.517 (0.205)	0.994 (0.005)	0.63 (0.189)
			KM	0.992 (0.016)	0.729 (0.079)	0.435 (0.108)
			BayesBD	0.57 (0.142)	0.996 (0.014)	0.701 (0.098)
			BayesImageS	0.946 (0.147)	0.897 (0.185)	0.754 (0.234)
	1.5	0.5	BFSP	0.893 (0.141)	0.965 (0.05)	0.816 (0.166)
			BFSP Spline	0.897 (0.088)	0.933 (0.081)	0.742 (0.165)
			CART	0.476 (0.237)	0.993 (0.008)	0.578 (0.237)
			KM	0.98 (0.024)	0.668 (0.062)	0.372 (0.051)
			BayesBD	0.612 (0.148)	0.996 (0.01)	0.73 (0.108)
			BayesImageS	0.948 (0.066)	0.847 (0.297)	0.755 (0.267)
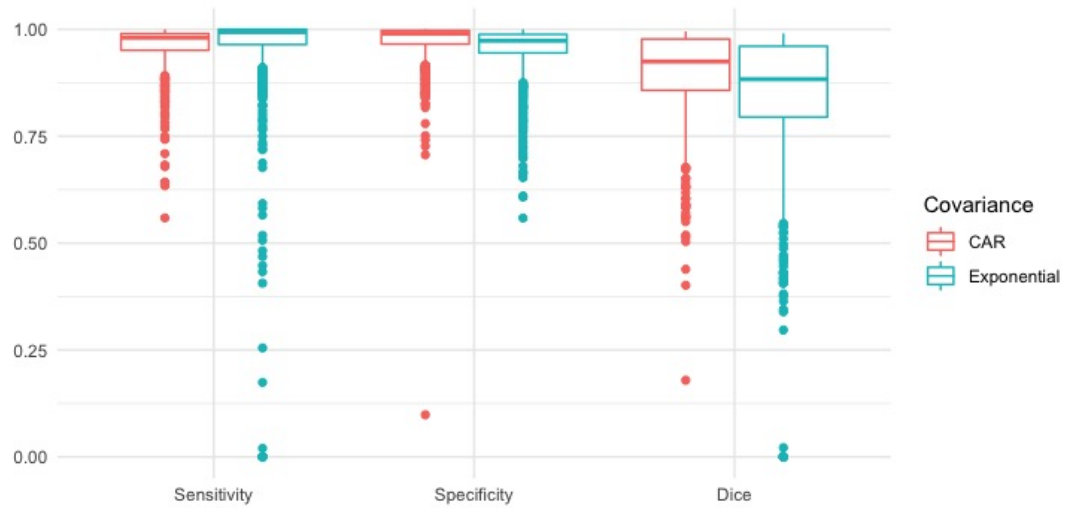
Figure A.2: Results of BFSP on 600 simulated datasets with either tapered CAR or tapered Exponential covariance.
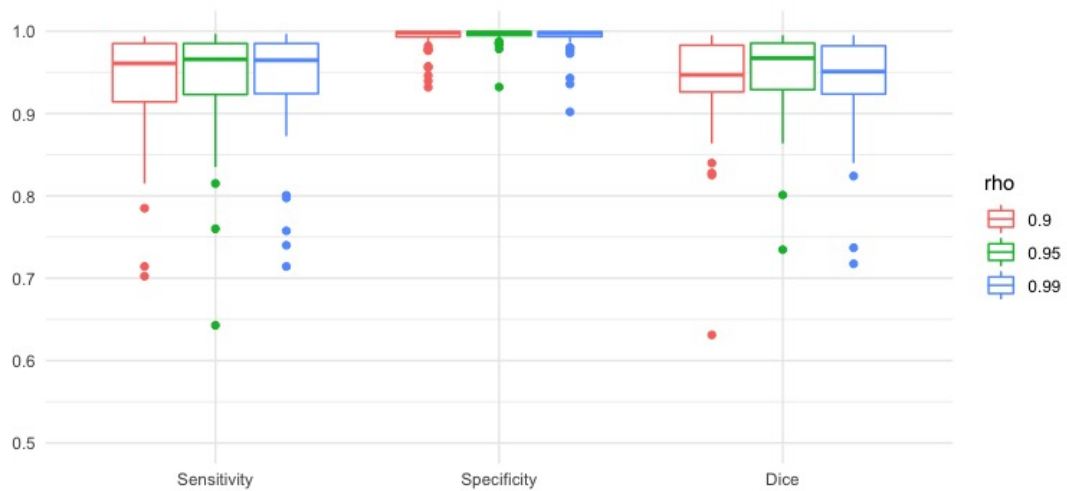

Figure A.3: Sensitivity, specificity, and Dice coefficient of BFSP for different values of CAR parameter ρ on 60 simulated datasets.
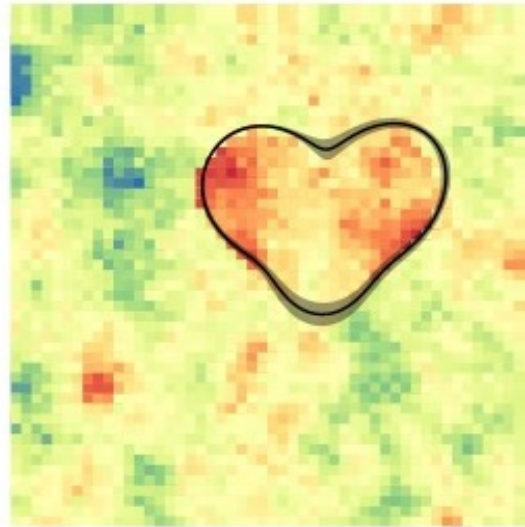
Figure A.4: Example of simulated data with results of BFSP (black line) and 95% credible band (grey)
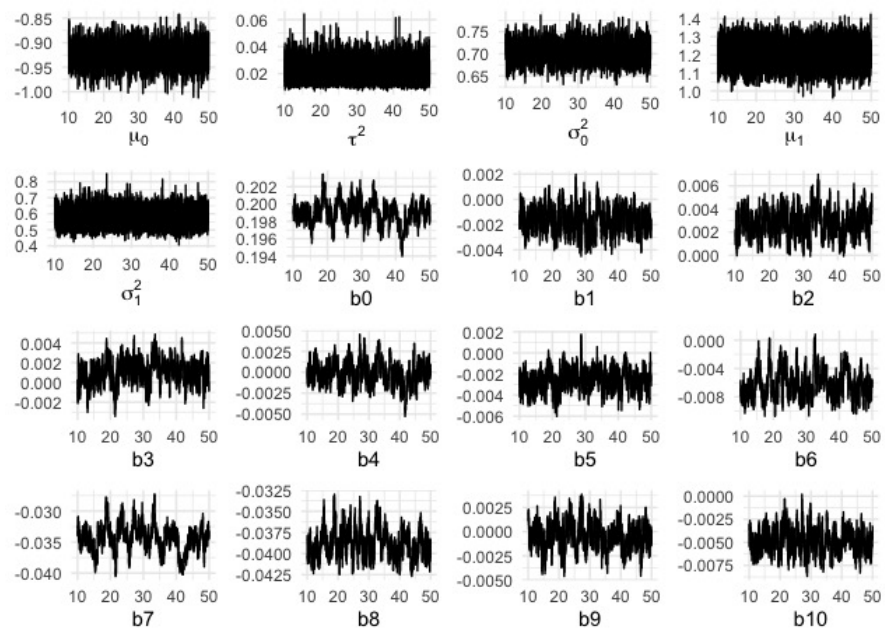

Figure A.5: Post burn-in trace plots of all parameters for one simulated data set. X-axes in thousands.
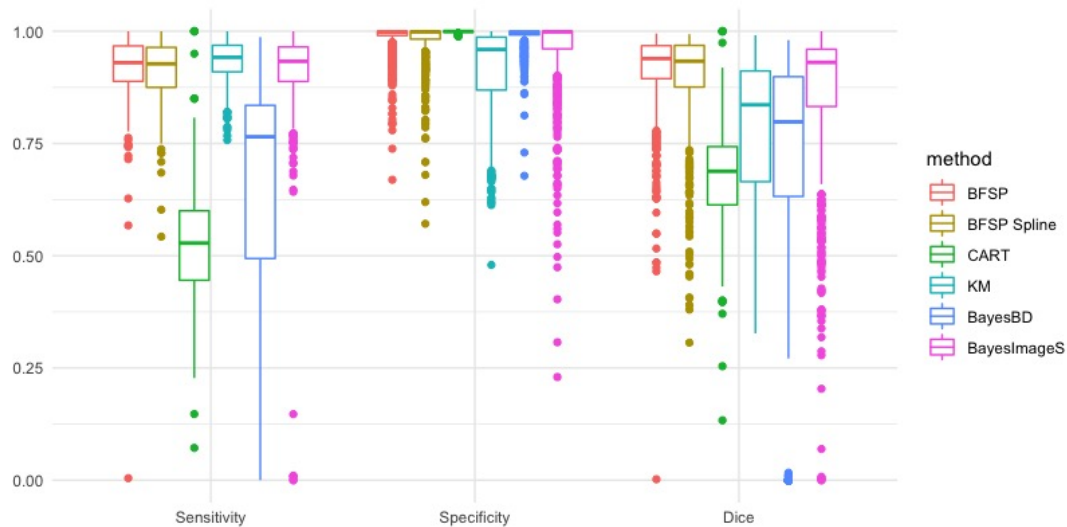
Figure A.6: Results of BFSP and competing methods when model is misspecified. The mean within the target region is a function of the distance from the centroid of the region to mimic the voxel-wise prostate cancer probability data.

Appendix B

Supplementary Material for Chapter 3

B.1 Simulation settings and results for data containing zero target regions

We consider a unit square image space D of 40 by 40 resolution and generate data from a Gaussian processes with Matérn covariance structure exhibiting stationarity (i.e. no anomalous regions). We vary the smoothness parameter and nugget parameter of the entire region. We fix the following settings during simulation: $\rho_0 = .05$, $\mu = 0$, $\sigma_0^2 = 1$. We randomly generate 100 spatial data sets for each of the following settings.

1. $\nu_0 = 1.5, \tau^2 = .1$

2. $\nu_0 = 1.5, \tau^2 = .5$

3. $\nu_0 = .5, \tau^2 = .1$

4. $\nu_0 = .5, \tau^2 = .5$

BFSP-M was highly successful in identifying zero anomalous regions when the covariance structure was correctly specified by the model (i.e. exponential spatial correlations given $\nu_0 = .5$). Our method tends to find false positive signals when the covariance structure is misspecified ($\nu_0 = 1.5$). See Figure B.1 for the posterior distribution of
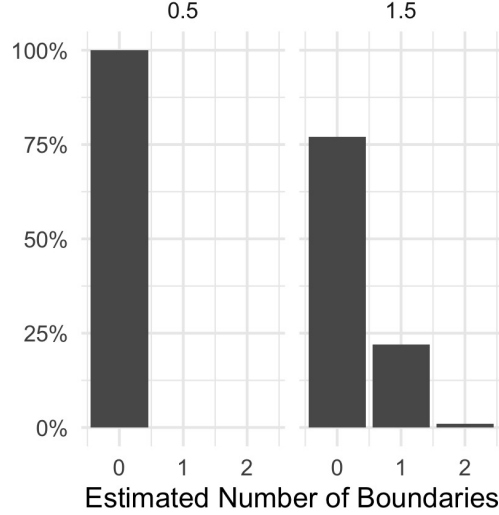
Figure B.1: Distribution of the estimated number of boundaries over 400 simulated data sets containing zero anomalous regions stratified by smoothness parameter of the simulated data.

number of boundaries in the 400 simulated datasets containing zero regions for data simulated with $\nu_0 = .5$ and $\nu_0 = 1.5$. When the model is correctly specified, BFSP-M correctly identified the true number of null lesions 100% of the time.

B.2 Derivation of Acceptance Ratios for RJ-MCMC

B.2.1 Birth Step

For a birth step we have $m' = m + 1$. We generate $\mathbf{v}$, which is the proposed parameters defining the new boundary and spatial process of the new region created by the new boundary. In this scenario, $\mathbf{v}'$ is empty.

$$\mathbf{v} = (c_{m+1}, \beta_{m+1}, \mu_{m+1}, \log(\sigma^2_{m+1}))$$
$$h_{m,m+1}(\mathbf{\Phi}_m, \mathbf{v}) = (\mathbf{\Phi}_m, c_{m+1}, \beta_{m+1}, \mu_{m+1}, \exp(\log(\sigma^2_{m+1})))$$

The Jacobian term is given by:

$$\left| \frac{\partial h_{m,m+1}(\mathbf{\Phi}_m, \mathbf{v})}{\partial(\mathbf{\Phi}_m, \mathbf{v})} \right| = \begin{vmatrix} 1 & 0 & \ldots & 0 \\ 0 & \ddots & 0 & 0 \\ \vdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \sigma^2_{m+1} \end{vmatrix} = \sigma^2_{m+1}$$

B.2.2 Death Step

For a death step we have $m' = m - 1$. This time $\mathbf{v}$ is empty and $\mathbf{v}'$ contains the parameters for the boundary that we propose to delete. Assume that we propose to delete boundary m.

$$\mathbf{v}' = \phi_m$$
$$h_{m,m-1}(\mathbf{\Phi}_m) = (\mathbf{\Phi}_{m-1}, \mathbf{v}')$$

Therefore the Jacobian term is:

$$\left| \frac{\partial h_{m,m-1}(\mathbf{\Phi}_m)}{\partial(\mathbf{\Phi}_m)} \right| = \begin{vmatrix} 1 & 0 & \ldots & 0 \\ 0 & \ddots & 0 & 0 \\ \vdots & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1$$

This holds when proposing deletion for any index $i = 1, ..., m$ after a reordering of labels.

B.2.3 Merge Step

For a merge step we have $m' = m - 1$. This time $\mathbf{v}$ contains the parameters of the new merged boundary and $\mathbf{v}'$ contains the parameters for the boundaries that we propose to merge. Assume that we propose to merge boundaries m and $m - 1$ into boundary m^*.

$$\mathbf{v} = (c_{m*}, \beta_{m*}, \mu_{m*}, \log(\sigma^2_{m*}))$$
$$\mathbf{v}' = (c_{m-1}, \beta_{m-1}, \mu_{m-1}, \sigma^2_{m-1}, c_m, \beta_m, \mu_m, \sigma^2_m)$$
$$h_{m,m-1}(\mathbf{\Phi}_m, \mathbf{v}) = (\mathbf{\Phi}_{m-2}, \mathbf{v}', c_{m*}, \beta_{m*}, \mu_{m*}, \exp(\log(\sigma^2_{m*})))$$

The Jacobian term is:

$$\left| \frac{\partial h_{m,m+1}(\boldsymbol{\Phi}_m, \mathbf{v})}{\partial(\boldsymbol{\Phi}_m, \mathbf{v})} \right| = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 \\ \vdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \sigma^2_{m*} \end{vmatrix} = \sigma^2_{m*}$$

This holds when proposing a merge for any 2 indices from $i = 1, ..., m$ after a reordering of labels.

B.2.4 Split Step

For a merge step we have $m' = m + 1$. This time $\mathbf{v}$ contains the parameters of the two new boundaries resulting from the split and $\mathbf{v}'$ contains the parameters for the boundary that we propose to split. Assume that we propose to split boundary m.

$$
\begin{aligned}
\mathbf{v} &= (c_{m*}, \beta_{m*}, \mu_{m*}, \log(\sigma^2_{m*}), c_{m+1}, \beta_{m+1}, \mu_{m+1}, \log(\sigma^2_{m+1})) \\
\mathbf{v}' &= (c_m, \beta_m, \mu_m, \sigma^2_m) \\
h_{m,m+1}(\boldsymbol{\Phi}_m, \mathbf{v}) &= (\boldsymbol{\Phi}_{m-1}, \mathbf{v}', c_{m*}, \beta_{m*}, \mu_{m*}, \exp(\log(\sigma^2_{m*})), c_{m+1}, \beta_{m+1}, \mu_{m+1}, \exp(\log(\sigma^2_{m+1})))
\end{aligned}
$$

Therefore the Jacobian term is:

$$\left| \frac{\partial h_{m,m+1}(\boldsymbol{\Phi}_m, \mathbf{v})}{\partial(\boldsymbol{\Phi}_m, \mathbf{v})} \right| = \begin{vmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ \vdots & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{m*} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{m+1} \end{vmatrix} = \sigma^2_{m*}\sigma^2_{m+1}$$

This holds when proposing a split for any index $i = 1, ..., m$ after a reordering of labels.
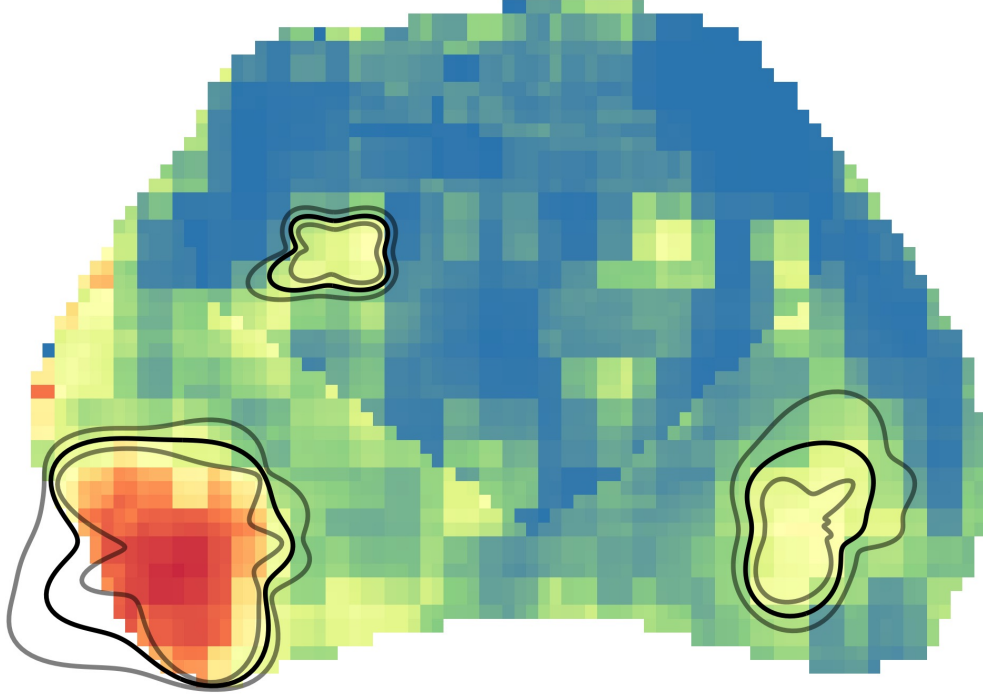
B.3 Additional Tables & Figures

Figure B.2: Estimated partitioning boundary (black) and 95% credible bands (grey) for one slice of prostate imaging data. Color represents voxel probability of cancer.

Table B.1: Mean (SD) of sensitivity (TPR), specificity (TNR), and Dice coefficient for each simulation setting. ν_1 and τ^2 are parameters of the simulated Gaussian processes.

ν_1, τ^2		BFSP-M	BIS	CART+KM	KM	BFSP-1
	Dice	0.887 (0.135)	0.706 (0.226)	0.618 (0.215)	0.595 (0.067)	0.517 (0.162)
.5, .1	TPR	0.891 (0.174)	0.856 (0.224)	0.697 (0.285)	0.902 (0.039)	0.47 (0.263)
	TNR	0.98 (0.034)	0.841 (0.251)	0.866 (0.235)	0.778 (0.064)	0.944 (0.119)
	Dice	0.84 (0.164)	0.644 (0.243)	0.581 (0.225)	0.533 (0.052)	0.52 (0.162)
.5, .5	TPR	0.846 (0.216)	0.836 (0.242)	0.662 (0.307)	0.892 (0.043)	0.464 (0.26)
	TNR	0.974 (0.038)	0.777 (0.304)	0.854 (0.258)	0.719 (0.052)	0.945 (0.138)
	Dice	0.861 (0.118)	0.675 (0.212)	0.625 (0.189)	0.593 (0.087)	0.536 (0.17)
1.5, .1	TPR	0.869 (0.154)	0.805 (0.227)	0.621 (0.268)	0.904 (0.074)	0.462 (0.256)
	TNR	0.972 (0.045)	0.829 (0.269)	0.927 (0.155)	0.774 (0.073)	0.961 (0.106)
	Dice	0.779 (0.204)	0.608 (0.238)	0.591 (0.227)	0.53 (0.062)	0.511 (0.164)
1.5, .5	TPR	0.757 (0.266)	0.747 (0.278)	0.685 (0.29)	0.888 (0.07)	0.451 (0.26)
	TNR	0.976 (0.046)	0.777 (0.344)	0.836 (0.275)	0.716 (0.059)	0.947 (0.131)

			Dice		
slice	BFSP-M	BIS	CART+KM	KM	BFSP-1
1	0.69	0.43	0.63	0.59	0.75
2	0.60	0.54	0.59	0.56	0.63
3	0.64	0.62	0.78	0.71	0.78
4	0.89	0.81	0.82	0.83	0.89
5	0.46	0.37	0.43	0.36	0.51
6	0.76	0.72	0.82	0.79	0.79
7	0.52	0.48	0.53	0.57	0.04
8	0.56	0.50	0.47	0.52	0.53
9	0.53	0.43	0.41	0.38	0.12
10	0.32	0.21	0.36	0.36	0.68
11	0.57	0.60	0.63	0.69	0.63
			Sensitivity		
slice	BFSP-M	BIS	CART+KM	KM	BFSP-1
1	0.60	0.28	0.73	0.49	0.69
2	0.58	0.42	0.59	0.50	0.70
3	0.60	0.46	0.69	0.64	0.76
4	0.91	0.75	0.92	0.92	0.92
5	0.39	0.27	0.52	0.28	0.57
6	0.72	0.64	0.88	0.87	0.77
7	0.41	0.36	0.44	0.57	0.02
8	0.43	0.35	0.34	0.43	0.39
9	0.69	0.30	0.50	0.29	0.08
10	0.19	0.12	0.22	0.22	0.52
11	0.44	0.45	0.70	0.69	0.49
			Specificity		
slice	BFSP-M	BIS	CART+KM	KM	BFSP-1
1	0.98	0.99	0.95	0.96	0.98
2	0.95	0.96	0.94	0.95	0.94
3	0.88	0.98	0.96	0.93	0.93
4	0.94	0.95	0.89	0.90	0.94
5	0.94	0.94	0.92	0.93	0.93
6	0.95	0.95	0.95	0.94	0.95
7	0.97	0.97	0.97	0.96	0.71
8	0.95	0.96	0.94	0.92	0.95
9	0.90	0.94	0.88	0.90	0.76
10	0.99	1.00	1.00	1.00	1.00
11	0.99	0.99	0.98	0.98	0.99

Table B.2: Dice score, Sensitivity, and Specificity for BFSP-M and competing methods applied to voxel-wise cancer probabilities for 11 slices.

Appendix C
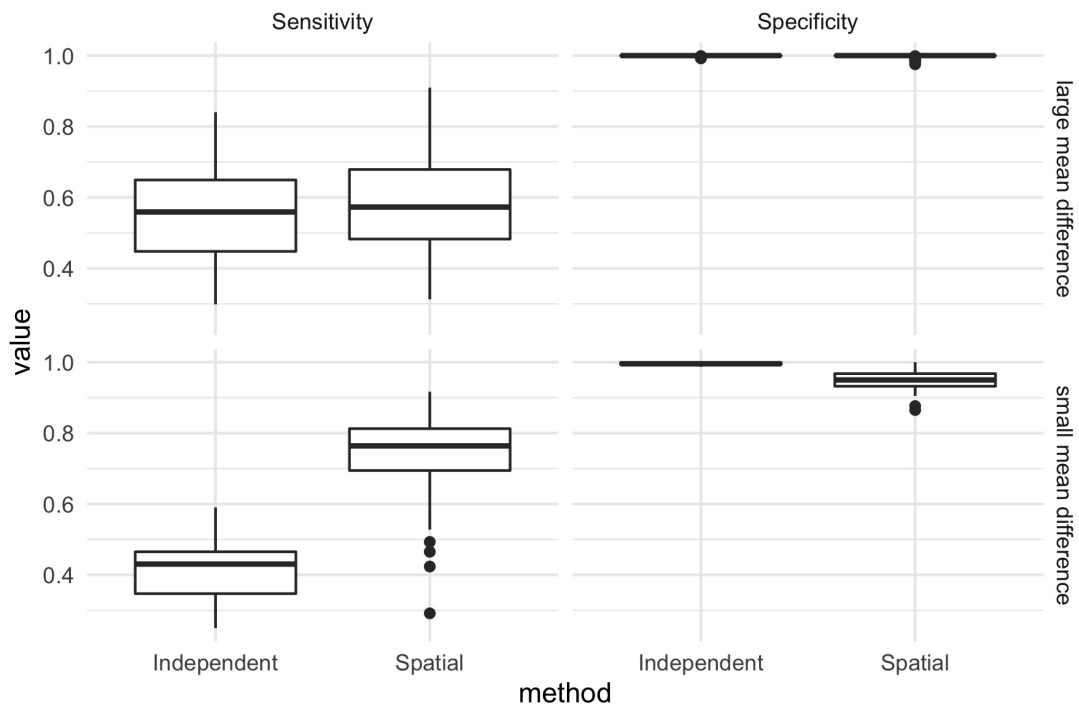
Supplementary Material for Chapter 4

Figure C.1: Sensitivity and specificity of BFSP-3D assuming distinct GPs (Spatial) or distinct independent gaussian distributions (Independent). Data is simulated according to Section 4.4 for the retangular shapes only with a small ($\mu_1 = .5$) or large ($\mu_1 = 2$) difference in means.