# Dynamic Diseased Region Detection for Longitudinal Medical Imaging Data


**Abstract**

The text of your abstract. 200 or fewer words.



*Keywords:* 3 to 6 keywords, that do not appear in the title

# 1 Introduction

# 2 Method

Suppose that we observe a longitudinal imaging dataset for $n$ unrelated subjects observed at $m$ time points. Let $S \subset \mathbb{R}^d$, $d = 2, 3$, be a common template from the dataset, and $\boldsymbol{S}_0 = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\}$ be the set of $k$ voxels in $\mathcal{S}$. The longitudinal imaging measurements for the $i$-th subject at the voxel $\boldsymbol{s}$ are denoted as $\boldsymbol{y}_i(\boldsymbol{s}) = (y_{i,1}(\boldsymbol{s}), \ldots, y_{i,m}(\boldsymbol{s}))^T$, $i = 1, \ldots, n$. Furthermore, let $\boldsymbol{x}_{i,j}$ be a $p \times 1$ vector including the demographic and clinical covariates (e.g., age, gender, scan date or biomarkers) for the $i$- th subject at the $j$-th time point, and $z_i$ be the status of some given disease (e.g., stroke or cancer) such that $x_i = 0$ and 1, respectively, represent normal control and patient. Furthermore, for the $i$-th subject at $j$-th time point, we assume that $\mathcal{S}$ can be decomposed into the union of normal region $\mathcal{R}^0_{i,j}$ and diseased region $\mathcal{R}^1_{i,j}$, that is

$$\mathcal{S} = \mathcal{R}^0_{i,j} \cup \mathcal{R}^1_{i,j} \text{ and } \mathcal{R}^0_{i,j} \cap \mathcal{R}^1_{i,j} = \emptyset.$$

Here we also assume that: (i) normal controls are expected to be perfectly healthy, i.e. do not have any diseased regions; (ii) for patients, the size and location of $\mathcal{R}^1_{i,j}$ may vary across subjects and time points.

## 2.1 Dynamic Spatial Random Effects Model

Our dynamic spatial random effects model (DSREM) consists of a spatial random effects models (SREM) [Besag, 1974, Diggle and Ribeiro, 2007, Geman and Geman, 1984, Huang et al., 2015, Li and Singh, 2009] and a dynamic conditional random field (DCRF) model [Sutton et al., 2007, Wang and Ji, 2005, Wang et al., 2006, Yin et al., 2009].

We first propose SREM to model the conditional distribution of the observed imaging measurements given two sets of random effects, i.e., $\{\boldsymbol{b}_i(\boldsymbol{s})\}_{i=1}^n$ and $\{\boldsymbol{\gamma}_i(\boldsymbol{s})\}_{i=1}^n$. In particular, $\boldsymbol{b}_i(\boldsymbol{s}) = (b_{i,1}(\boldsymbol{s}), \ldots, b_{i,m}(\boldsymbol{s}))^T$ is a $m \times 1$ vector with $b_{i,j}(\boldsymbol{s}) = 0$ if $\boldsymbol{s} \in \mathcal{R}^0_{i,j}$ for the $i-$th subject at $j$-th time point, otherwise $b_{i,j}(\boldsymbol{s}) = 1$. The other random effect $\boldsymbol{\gamma}_i(\boldsymbol{s})$ is a $p \times 1$ vector indicating the subject-specific random effect. Given $b_{i,j}(\boldsymbol{s})$ and $\boldsymbol{\gamma}_i(\boldsymbol{s})$, the SREM is given as

$$y_{i,j}(\boldsymbol{s}) = \boldsymbol{x}_{i,j}^T\boldsymbol{\beta}(\boldsymbol{s}) + \boldsymbol{x}_{i,j}^T\boldsymbol{\gamma}_i(\boldsymbol{s}) + \alpha(\boldsymbol{s})b_{i,j}(\boldsymbol{s})z_i + \epsilon_{i,j}(\boldsymbol{s}), \tag{1}$$

where $\boldsymbol{\beta}(\boldsymbol{s})$ is a $p \times 1$ vector representing the fixed effect at voxel $\boldsymbol{s}$, while $\alpha(\boldsymbol{s})$ represents the additional effect caused by the diseased region. $\epsilon_{i,j}(\boldsymbol{s})$s are independent measurement errors across subjects, time points, and voxels, following the Normal distribution $N(0, \sigma^2(\boldsymbol{s}))$. Based on the proposed model, the pontantial heterogeneity (among different voxels, subjects, and time points) are mainly captured by the term "$\alpha(\boldsymbol{s})b_{i,j}(\boldsymbol{s})z_i$". Additionaly, for all voxels in normal controls and the voxels belonging to normal regions in patients, SREM (1) can be simplifid into

$$y_{i,j}(\boldsymbol{s}) = \boldsymbol{x}_{i,j}^T\boldsymbol{\beta}(\boldsymbol{s}) + \boldsymbol{x}_{i,j}^T\boldsymbol{\gamma}_i(\boldsymbol{s}) + \epsilon_{i,j}(\boldsymbol{s}), \tag{2}$$

Then, we model the random effects $\boldsymbol{\gamma}_i(\boldsymbol{s})$ and $\boldsymbol{b}_i(\boldsymbol{s})$ as follows. First, it is assumed that $\{\boldsymbol{\gamma}_i(\boldsymbol{s}), \boldsymbol{s} \in \boldsymbol{S}_0\}$, $\boldsymbol{b}_i(\boldsymbol{s})$ and $\{\epsilon_{i,j}(\boldsymbol{s})\}$ are mutually independent. Second, $\{\boldsymbol{\gamma}_i(\boldsymbol{s}), \boldsymbol{s} \in \boldsymbol{S}_0\}$ are assumed to be mutually independent across voxels and follows $N(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{s}))$. Moreover, $\{\boldsymbol{b}_i\}_{i=1}^n$ are assumed independent across subjects and each $\boldsymbol{b}_i$ follows a DCRF model [Sutton et al., 2007, Wang et al., 2006], which is given by

$$\begin{aligned}
p(\boldsymbol{b}_i|\tau,\eta) &\propto p(\boldsymbol{b}_{i,1}|\tau)\prod_{j=2}^{m} p(\boldsymbol{b}_{i,j}|\boldsymbol{b}_{i,j-1},\tau,\eta), && (3)\\
p(\boldsymbol{b}_{i,1}|\tau) &= \exp\Big\{-\tau\sum_{s\in S_0}\sum_{s'\in N_s} U(b_{i,1}(\boldsymbol{s}), b_{i,1}(\boldsymbol{s}'))\Big\}, && (4)\\
p(\boldsymbol{b}_{i,j}|\boldsymbol{b}_{i,j-1},\tau,\eta) &= \exp\Big\{-\sum_{s\in S_0}\big[\tau\sum_{s'\in N_s} U(b_{i,j}(\boldsymbol{s}), b_{i,j}(\boldsymbol{s}'))\\
&\quad +\eta\sum_{s'\in M_s} U(b_{i,j}(\boldsymbol{s}), b_{i,j-1}(\boldsymbol{s}'))\big]\Big\}, && (5)
\end{aligned}$$

where

$$U(b_{i,j}(\boldsymbol{s}), b_{i,j'}(\boldsymbol{s}')) = \frac{1-\delta(b_{i,j}(\boldsymbol{s}), b_{i,j'}(\boldsymbol{s}'))}{||\boldsymbol{s}-\boldsymbol{s}'||^2},$$

and $\delta(\cdot)$ is the Kronecker delta function. $\tau$ is introduced to encourage spatial smoothness in homogeneous regions while $\eta$ influences the strength of temporal dependencies. Moreover, both $N_s$ and $M_s$ denote the neighboring sites of vexel $\boldsymbol{s}$. It should be noted that $M_s$ is not equivalent to the neighborhood $N_s$: (i) $M_s$ and $N_s$ may have different sizes, and (ii) $\boldsymbol{s} \notin N_s$ while $\boldsymbol{s} \in M_s$.

To distinguish them, $N_s$ is called the spatial neighborhood and $M_s$ the temporal neighborhood. Throughout the paper, we consider $N_s$ is the set of the closest $3^d - 1$ neighbors of pixel $\boldsymbol{s}$, while $M_s = N_s \bigcup \{\boldsymbol{s}\}$.

## 2.2 Estimation Procedure

Our next task is to estimate the random effects $\{\boldsymbol{b}_i(\boldsymbol{s}), \boldsymbol{s} \in \boldsymbol{S}_0\}_{i=1}^n$ and all unknown parameters consisting of $\tau, \eta, \boldsymbol{\beta}(\boldsymbol{s}), \alpha(\boldsymbol{s}), \sigma^2(\boldsymbol{s})$, and $\boldsymbol{\Sigma}(\boldsymbol{s})$ for $\boldsymbol{s} \in \boldsymbol{S}_0$. We decompose these parameters into three parts: (i) $\boldsymbol{\beta}(\boldsymbol{s}), \alpha(\boldsymbol{s}), \sigma^2(\boldsymbol{s}), \boldsymbol{\Sigma}(\boldsymbol{s})$, and (ii) $\tau, \eta$. For part (i), the maximum likelihood estimate (MLE) can be calculated by using the expectation-maximization (EM) algorithm [Huang et al., 2015]. In particular, the MLEs of $\boldsymbol{\beta}(\boldsymbol{s}), \sigma^2(\boldsymbol{s}), \boldsymbol{\Sigma}(\boldsymbol{s})$ can be derived based on the whole dataset (or only the normal controls for computational efficiency) while the MLE of $\alpha(\boldsymbol{s})$ based on the subpopulation (only including the patients). For part (ii), $\tau$ and $\eta$ can be predefined or dertermined by some data-driven method. In this paper, they are estimated by using a pseudo-likelihood method [Geman and Graffigne, 1986] since the MLEs of $\tau$ and $\eta$ are generally difficult to compute due to the normalizing part of the probability function in (3). In addition, the random effects $\{\boldsymbol{b}_i\}_{i=1}^n$ can be estimated via the maximum a posteriori on Markov random field (MRF-MAP) method.

### 2.2.1 EM algorithm

To derive the EM algorithm for the parameters in part (i), we need to derive the complete-data log-likelihood function of DSREM as follows. Let $\boldsymbol{x}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,m})$ and $\boldsymbol{y}_i(\boldsymbol{s}) = (y_{i,1}(\boldsymbol{s}), \ldots, y_{i,m}(\boldsymbol{s}))^T$. Thus, the distribution of $\boldsymbol{y}_i(s)$ conditional on random effects is given by $N(\boldsymbol{\mu}_i(\boldsymbol{s}), \boldsymbol{\Omega}(\boldsymbol{s}))$, where $\boldsymbol{\mu}_i(\boldsymbol{s}) = \boldsymbol{x}_i^T(\boldsymbol{\beta}(\boldsymbol{s}) + \boldsymbol{\gamma}_i(\boldsymbol{s})) + \alpha(\boldsymbol{s})\boldsymbol{b}_i(\boldsymbol{s})z_i$, and $\boldsymbol{\Omega}(\boldsymbol{s}) = \sigma^2(\boldsymbol{s})\boldsymbol{I}_m$. Thus, the complete-data log-likelihood function is given by

$$\begin{aligned}\log L \quad \propto \quad & -\frac{n}{2}\sum_{l=1}^{k}\log|\boldsymbol{\Omega}(\boldsymbol{s}_l)| - \frac{n}{2}\sum_{l=1}^{k}\log|\boldsymbol{\Sigma}(\boldsymbol{s}_l)| \\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{k}\boldsymbol{\mu}_i^T(\boldsymbol{s}_l)\boldsymbol{\Omega}^{-1}(\boldsymbol{s}_l)\boldsymbol{\mu}_i(\boldsymbol{s}_l) - \sum_{i=1}^{n_0}\log p(\boldsymbol{b}_i|\tau,\eta) \\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{k}\boldsymbol{\gamma}_i^T(\boldsymbol{s}_l)\boldsymbol{\Sigma}^{-1}(\boldsymbol{s}_l)\boldsymbol{\gamma}_i(\boldsymbol{s}_l), \end{aligned} \tag{6}$$

where $n_0 = \#\{z_i : z_i \neq 0\}$ denotes the number of patients. Given the current estimate of $\boldsymbol{\theta} = \{\boldsymbol{\beta}(\boldsymbol{s}), \alpha(\boldsymbol{s}), \sigma^2(\boldsymbol{s}), \boldsymbol{\Sigma}(\boldsymbol{s})\}$ at iteration $r$, denoted as $\boldsymbol{\theta}^{(r)}$, their updates are obtained via maximizing the following Q-function $Q_{\theta^{(r)}}(\boldsymbol{\theta}) \doteq E_{\theta^{(r)}}(\log L|\boldsymbol{y}, \boldsymbol{x})$ with respect to $\boldsymbol{\theta}$:

$$\begin{aligned} & Q_{\theta^{(r)}}(\boldsymbol{\theta}) \propto -\frac{n}{2}\sum_{l=1}^{k}\log|\boldsymbol{\Omega}(\boldsymbol{s}_l)| - \frac{n}{2}\sum_{l=1}^{k}\log|\boldsymbol{\Sigma}(\boldsymbol{s}_l)| \\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{k}E\Big[\boldsymbol{\mu}_i^T(\boldsymbol{s}_l)\boldsymbol{\Omega}^{-1}(\boldsymbol{s}_l)\boldsymbol{\mu}_i(\boldsymbol{s}_l)\Big|\boldsymbol{y}_i(\boldsymbol{s}_l), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\Big] \\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{k}E\Big[\boldsymbol{\gamma}_i^T(\boldsymbol{s}_l)\boldsymbol{\Sigma}^{-1}(\boldsymbol{s}_l)\boldsymbol{\gamma}_i(\boldsymbol{s}_l)\Big|\boldsymbol{y}_i(\boldsymbol{s}_l), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\Big]. \end{aligned} \tag{7}$$

We consider the E-step and M-step of the EM algorithm as follows.

**E-step:** In the E-step, we need to calculate four conditional expectations:

$$\begin{aligned} & E\big[\boldsymbol{\gamma}_i(\boldsymbol{s}_l)\boldsymbol{b}_i^T(\boldsymbol{s}_l)\big|\boldsymbol{y}_i(\boldsymbol{s}_l), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\big], E\big[\boldsymbol{\gamma}_i(\boldsymbol{s}_l)\big|\boldsymbol{y}_i(\boldsymbol{s}_l), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\big], \\ & E\big[\boldsymbol{\gamma}_i(\boldsymbol{s}_l)\boldsymbol{\gamma}_i^T(\boldsymbol{s}_l)\big|\boldsymbol{y}_i(\boldsymbol{s}_l), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\big], E\big[\boldsymbol{b}_i(\boldsymbol{s}_l)\big|\boldsymbol{y}_i(\boldsymbol{s}_l), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\big]. \end{aligned} \tag{8}$$

In order to calculate these conditional probabilities, the class labels $\boldsymbol{b}_i$ should be estimated first. Here we consider the MRF-MAP estimation, which is efficient and adopted in many existing papers, e.g., Marroquín et al. [2002], Nie et al. [2009], Zhang et al. [2001]. The detialed derivation of MRF-MAP estimation will be discussed later in next subsection.

Assumed that we have the MRF-MAP estimate of $\boldsymbol{b}_i$ at iteration $r$, i.e., $\hat{\boldsymbol{b}}_i^{(r)}$, the expectation $E\big[\boldsymbol{b}_{i,j}(\boldsymbol{s}_l)\big|\boldsymbol{y}_i(\boldsymbol{s}_l), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\big]$ can be calculated as

$$\frac{f(y_{i,j}(\boldsymbol{s}_l)|x_{i,j}, b_{i,j}(\boldsymbol{s}_l) = 1, \boldsymbol{\theta}^{(r)})P(b_{i,j}(\boldsymbol{s}_l) = 1|\hat{\boldsymbol{b}}_i^{(r)}, \boldsymbol{\theta}^{(r)})}{\sum_{t=0}^{1} f(y_{i,j}(\boldsymbol{s}_l)|x_{i,j}, b_{i,j}(\boldsymbol{s}_l) = t, \boldsymbol{\theta}^{(r)})P(b_{i,j}(\boldsymbol{s}_l) = t|\hat{\boldsymbol{b}}_i^{(r)}, \boldsymbol{\theta}^{(r)})}, \tag{9}$$

where

$$f(\boldsymbol{y}_i(\boldsymbol{s})|\boldsymbol{x}_i, \boldsymbol{b}_i(\boldsymbol{s}), \boldsymbol{\theta}^{(r)}) \sim \mathcal{N}(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)}(\boldsymbol{s}) + \hat{\alpha}^{(r)}(\boldsymbol{s})\boldsymbol{b}_i(\boldsymbol{s})z_i, \boldsymbol{x}_i^T \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i + \hat{\boldsymbol{\Omega}}^{(r)}(\boldsymbol{s})).$$

If $j = 1$,

$$P(b_{i,j}(\boldsymbol{s}) = 1|\hat{\boldsymbol{b}}_i^{(r)}, \boldsymbol{\theta}^{(r)}) \propto \exp\Big\{ -\tau \sum_{s \in S_0} \sum_{s' \in N_s} U(1, \hat{b}_{i,1}^{(r)}(\boldsymbol{s}')) \Big\},$$

otherwise,

$$\begin{aligned} P(b_{i,j}(\boldsymbol{s}) = 1|\hat{\boldsymbol{b}}_i^{(r)}, \boldsymbol{\theta}^{(r)}) \quad &\propto \quad \exp\Big\{ -\sum_{s \in S_0} \big[\tau \sum_{s' \in N_s} U(1, \hat{b}_{i,j}^{(r)}(\boldsymbol{s}')) \\ &\qquad + \eta \sum_{s' \in M_s} U(1, \hat{b}_{i,j-1}^{(r)}(\boldsymbol{s}'))\big] \Big\}. \end{aligned}$$

Recall that, given $\boldsymbol{x}_i$ and $\hat{b}_i^{(r)}(\boldsymbol{s})$, $(\boldsymbol{y}_i^T(\boldsymbol{s}), \boldsymbol{\gamma}_i^T(\boldsymbol{s}))^T$ is normally distributed as

$$\mathcal{N}\left( \begin{pmatrix} \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)}(\boldsymbol{s}) + \hat{\alpha}^{(r)}(\boldsymbol{s})\boldsymbol{b}_i(\boldsymbol{s})z_i \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{x}_i^T \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i + \hat{\boldsymbol{\Omega}}^{(r)}(\boldsymbol{s}_k) & \boldsymbol{x}_i^T \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s}) \\ \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i & \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s}) \end{pmatrix} \right).$$

Then, given $\boldsymbol{y}_i(\boldsymbol{s})$, $\boldsymbol{x}_i$, and $\hat{b}_i^{(r)}(\boldsymbol{s})$, we have

$$\begin{aligned} &E\Big[\boldsymbol{\gamma}_i(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \hat{b}_i^{(r)}(\boldsymbol{s}), \boldsymbol{\theta}^{(r)}\Big] = \\ &\hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i(\boldsymbol{x}_i^T \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i + \hat{\boldsymbol{\Omega}}^{(r)}(\boldsymbol{s}))^{-1}\big(\boldsymbol{y}_i(\boldsymbol{s}) - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)}(\boldsymbol{s}) - \hat{\alpha}^{(r)}(\boldsymbol{s})\boldsymbol{b}_i(\boldsymbol{s})z_i\big), \end{aligned}$$

$$\begin{aligned} &E\Big[\boldsymbol{\gamma}_i(\boldsymbol{s})\boldsymbol{\gamma}_i^T(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \hat{b}_i^{(r)}(\boldsymbol{s}), \boldsymbol{\theta}^{(r)}\Big] = \\ &\hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s}) - \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i(\boldsymbol{x}_i^T \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i + \hat{\boldsymbol{\Omega}}^{(r)}(\boldsymbol{s}))^{-1}\boldsymbol{x}_i^T \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s}) + E\Big[\boldsymbol{\gamma}_i(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \hat{b}_i^{(r)}(\boldsymbol{s}), \boldsymbol{\theta}^{(r)}\Big]^{\otimes 2}. \end{aligned}$$

Finally, the desired expectations can be estimated as

$$\begin{aligned} E\Big[\gamma_{i,j}(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\Big] \quad &= \quad \sum_{t=0}^{1} E\Big[\gamma_{i,j}(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \boldsymbol{b}_{i,j}(\boldsymbol{s}) = t, \boldsymbol{\theta}^{(r)}\Big] \\ &\qquad \times P(b_{i,j}(\boldsymbol{s}) = t|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}), \end{aligned}$$

$$\begin{aligned} E\Big[\boldsymbol{\gamma}_{i,j}(\boldsymbol{s})b_{i,j}(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}\Big] \quad &= \quad E\Big[\boldsymbol{\gamma}_{i,j}(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, b_{i,j}(\boldsymbol{s}) = 1, \boldsymbol{\theta}^{(r)}\Big] \\ &\qquad \times P(b_{i,j}(\boldsymbol{s}) = 1|\boldsymbol{y}_i(\boldsymbol{s}), \boldsymbol{x}_i, \boldsymbol{\theta}^{(r)}), \end{aligned}$$

$$E\Big[\boldsymbol{\gamma}_{i,j}(\boldsymbol{s})\boldsymbol{\gamma}_{i,j}^T(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}),\boldsymbol{x}_i,\boldsymbol{\theta}^{(r)}\Big] = \sum_{t=0}^{1} E\Big[\boldsymbol{\gamma}_{i,j}(\boldsymbol{s})\boldsymbol{\gamma}_{i,j}^T(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}),\boldsymbol{x}_i,b_{i,j}(\boldsymbol{s})=t,\boldsymbol{\theta}^{(r)}\Big] \times P(b_{i,j}(\boldsymbol{s})=t|\boldsymbol{y}_i(\boldsymbol{s}),\boldsymbol{x}_i,\boldsymbol{\theta}^{(r)}). \quad (10)$$

**M-step:** Taking derivatives of (7) with respect to $\boldsymbol{\theta}$ and equating them to zeros, we find the updates of $\boldsymbol{\theta}$ as follows. For $\boldsymbol{\beta}(\boldsymbol{s})$ and $\alpha(\boldsymbol{s})$, we have

$$\hat{\boldsymbol{\beta}}^{(r+1)}(\boldsymbol{s}) = [\sum_{i=1}^{n}\boldsymbol{x}_i\hat{\boldsymbol{\Omega}}^{(r)^{-1}}(\boldsymbol{s})\boldsymbol{x}_i^T]^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\hat{\boldsymbol{\Omega}}^{(r)^{-1}}(\boldsymbol{s}) \Big(\boldsymbol{y}_i(\boldsymbol{s})-\boldsymbol{x}_i^T E\Big[\boldsymbol{\gamma}_i(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}),\boldsymbol{x}_i,\boldsymbol{\theta}^{(r)}\Big] -z_i\hat{\alpha}^r\boldsymbol{x}_i^T E\Big[\boldsymbol{b}_i(\boldsymbol{s}),l)\Big|\boldsymbol{y}_i(\boldsymbol{s}),\boldsymbol{x}_i,\boldsymbol{\theta}^{(r)}\Big]\Big), \quad (11)$$

and

$$\hat{\alpha}^{(r+1)} = \Big[\sum_{i=1}^{n_0}\sum_{l=1}^{m}\hat{\sigma}^{-2(r)}(\boldsymbol{s}_l)E\big[b_i^T(\boldsymbol{s}_l)b_i(\boldsymbol{s}_l)\big|\Big]^{-1} \sum_{i=1}^{n_0}\sum_{l=1}^{m}\hat{\sigma}^{-2(r)}(\boldsymbol{s}_l)\Big\{E[\boldsymbol{b}_i^T(\boldsymbol{s}_l)][\boldsymbol{y}_i(\boldsymbol{s}_l)-\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}^{(r+1)}(\boldsymbol{s}_l)] -E\Big[\boldsymbol{\gamma}_i^T(\boldsymbol{s}_l)\boldsymbol{b}_i(\boldsymbol{s}_l)\Big|\boldsymbol{y}_i(\boldsymbol{s}_l),\boldsymbol{x}_i,\boldsymbol{\theta}^{(r)}\Big]\Big\}. \quad (12)$$

For the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{s})$, we have

$$\hat{\boldsymbol{\Sigma}}^{(r+1)}(\boldsymbol{s}) = \frac{1}{n}\sum_{i=1}^{n}E\Big[\boldsymbol{\gamma}_i(\boldsymbol{s})\boldsymbol{\gamma}_i^T(\boldsymbol{s})\Big|\boldsymbol{y}_i(\boldsymbol{s}),\boldsymbol{x}_i,\boldsymbol{\theta}^{(r)}\Big]. \quad (13)$$

For $\sigma^2(\boldsymbol{s})$ in $\boldsymbol{\Omega}(\boldsymbol{s}i)$,

$$\hat{\sigma}2^{(r+1)}(\boldsymbol{s}) = \frac{1}{nm}\sum_{i=1}^{n}E\Big[[\boldsymbol{\nu}_i^{(r)}(\boldsymbol{s})]^T\boldsymbol{\nu}_i^{(r)}(s_k)\Big|\boldsymbol{y}_i(\boldsymbol{s}),\boldsymbol{x}_i,\boldsymbol{\theta}^{(r)}\Big]. \quad (14)$$

The E-step and M-step are alternately repeated until the difference between $\log L(\widetilde{\boldsymbol{\theta}}^{(r+1)})$ and $\log L(\widetilde{\boldsymbol{\theta}}^{(r)})$ is smaller than a desired value (e.g., $10^{-4}$).

### 2.2.2 MRF-MAP estimation method

The MRF-MAP estimation is an efficient method for many practical applications (e.g., image segmentation) and adopted in many literatures, e.g., [Nie et al., 2009, Zhang et al., 2001]. First, given the current estimate $\widetilde{\boldsymbol{\theta}}^{(r)}$ and $\widetilde{\tau}^{(r)}$ at iteration $r$, the conditional probability density function of $\boldsymbol{y}_i(\boldsymbol{s})$ given $\boldsymbol{x}_i$ and $\boldsymbol{b}_i(\boldsymbol{s})$ is derived as

$$\boldsymbol{y}_i(\boldsymbol{s})|\boldsymbol{x}_i, \boldsymbol{b}_i(\boldsymbol{s}), \boldsymbol{\theta}^{(r)} \sim \mathcal{N}(\hat{\boldsymbol{\nu}}_i^{(r)}, \hat{\boldsymbol{\Lambda}}_i^{(r)}(\boldsymbol{s})), \tag{15}$$

where $\hat{\boldsymbol{\nu}}_i^{(r)} = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}^{(r)}(\boldsymbol{s}) + \hat{\alpha}^{(r)}(\boldsymbol{s})\boldsymbol{b}_i(\boldsymbol{s})z_i$, and $\hat{\boldsymbol{\Lambda}}_i^{(r)}(\boldsymbol{s}) = \boldsymbol{x}_i^T \hat{\boldsymbol{\Sigma}}^{(r)}(\boldsymbol{s})\boldsymbol{x}_i + \hat{\boldsymbol{\Omega}}^{(r)}(\boldsymbol{s})$.

According to the MAP criterion, the estimate $\widetilde{\boldsymbol{b}}_i^{(r+1)}$ is defined as

$$\begin{aligned}
\widetilde{\boldsymbol{b}}_i^{(r+1)} &= \arg\max_{\boldsymbol{b}_i} \left\{ \prod_{l=1}^{k} f(\boldsymbol{y}_i(\boldsymbol{s}_l)|\boldsymbol{x}_i, b_i(\boldsymbol{s}_l), \boldsymbol{\theta}^{(r)}) p(\boldsymbol{b}_i|\widetilde{\tau}^{(r)}) \right\} \\
&= \arg\min_{\boldsymbol{b}_i} \left\{ \frac{1}{2} \sum_{s \in S_0} [\boldsymbol{y}_i(\boldsymbol{s}_l) - \boldsymbol{\nu}_i^{(r)}(\boldsymbol{s}_l)]^T \hat{\boldsymbol{\Lambda}}_i^{(r)}(\boldsymbol{s})^{-1} [\boldsymbol{y}_i(\boldsymbol{s}_l) - \boldsymbol{\nu}_i^{(r)}(\boldsymbol{s}_l)] \right. \\
&\quad + \hat{\tau}^{(r)} \sum_{s \in S_0} \sum_{s' \in N_s} U(b_{i,1}(\boldsymbol{s}), b_{i,1}(\boldsymbol{s}')) + \sum_{s \in S_0} \Big[ \hat{\tau}^{(r)} \sum_{s' \in N_s} U(b_{i,j}(\boldsymbol{s}), b_{i,j}(\boldsymbol{s}')) \\
&\quad \left. + \hat{\eta}^{(r)} \sum_{s' \in M_s} U(b_{i,j}(\boldsymbol{s}), b_{i,j-1}(\boldsymbol{s}')) \Big] \right\}.
\end{aligned} \tag{16}$$

To obtain the optimal solution to (16), in this paper, we adopt the iterated conditional modes (ICM) algorithm Besag [1986], which uses a greedy iterative strategy for minimization. Convergence is achieved after only a few iterations.

### 2.2.3 Pseudo-likelihood method

Since $\tau$ and $\eta$ in model (3) is not the primary parameter of interest, we use an approximate, but computationally efficient method based on a pseudo-likelihood function. A key advantage of using the pseudo-likelihood function is its computational simplicity, since it does not involve the intractable partition function. The pseudo-likelihood at the iteration $r$ is a simple product of the conditional likelihood

$$PL(\widetilde{\boldsymbol{b}}^{(r)}, \tau, \eta) = \prod_{\{i: z_i = 1\}} \prod_{\boldsymbol{s} \in \mathcal{S}_0 - \partial\mathcal{S}_0} PL(\hat{\boldsymbol{b}}_i^{(r)}(\boldsymbol{s})|\hat{\boldsymbol{b}}_i^{(r)}), \tag{17}$$

where $\partial\mathcal{S}_1$ denotes the set of points at the boundaries of $\mathcal{S}_1$, and $PL(\hat{\boldsymbol{b}}_i^{(r)}(\boldsymbol{s})|\widetilde{\boldsymbol{b}}_i^{(r)})$ is given by

$$\frac{p(\hat{\boldsymbol{b}}_i(\boldsymbol{s})|\tau, \eta)}{\sum_{b_{i,1}(\boldsymbol{s})=0}^{1} \cdots \sum_{b_{i,m}(\boldsymbol{s})=0}^{1} p(\boldsymbol{b}_i(\boldsymbol{s})|\tau, \eta)}.$$

Thus, the MPL estimate $\hat{\tau}^{(r+1)}$ and $\hat{\eta}^{(r+1)}$ can be obtained by solving

$$\frac{\partial \ln PL(\hat{\boldsymbol{b}}^{(r)}, \tau, \eta)}{\partial \tau} = 0, \quad \frac{\partial \ln PL(\hat{\boldsymbol{b}}^{(r)}, \tau, \eta)}{\partial \eta} = 0. \tag{18}$$

# References


Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological) 36*(2), 192–236.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B. Methodological 48*(3), 259–302.

Diggle, P. and P. J. Ribeiro (2007). *Model-based geostatistics*. Springer.

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 6*(6), 721–741.

Geman, S. and C. Graffigne (1986). Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians*, Volume 1, pp. 2. AMS, Providence, RI.

Huang, C., L. Shan, H. C. Charles, W. Wirth, M. Niethammer, and H. Zhu (2015). Diseased region detection of longitudinal knee magnetic resonance imaging data. *IEEE transactions on medical imaging 34*(9), 1914–1927.

Li, S. Z. and S. Singh (2009). *Markov random field modeling in image analysis*. Springer.

Marroquín, J. L., B. C. Vemuri, S. Botello, E. Calderon, and A. Fernandez-Bouzas (2002). An accurate and efficient bayesian method for automatic segmentation of brain mri. *Medical Imaging, IEEE Transactions on 21*(8), 934–945.

Nie, J., Z. Xue, T. Liu, G. S. Young, K. Setayesh, L. Guo, and S. T. Wong (2009). Automated brain tumor segmentation using spatial accuracy-weighted hidden markov random field. *Computerized Medical Imaging and Graphics 33*(6), 431–441.

Sutton, C., A. McCallum, and K. Rohanimanesh (2007). Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *Journal of Machine Learning Research 8*(Mar), 693–723.

Wang, Y. and Q. Ji (2005). A dynamic conditional random field model for object segmentation in image sequences. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Volume 1, pp. 264–270. IEEE.

Wang, Y., K.-F. Loe, and J.-K. Wu (2006). A dynamic conditional random field model for foreground and shadow segmentation. *IEEE transactions on pattern analysis and machine intelligence 28*(2), 279–289.

Yin, J., D. H. Hu, and Q. Yang (2009). Spatio-temporal event detection using dynamic conditional random fields. In *Proceedings of the 21st international jont conference on Artifical intelligence*, pp. 1321–1326. Morgan Kaufmann Publishers Inc.

Zhang, Y., M. Brady, and S. Smith (2001). Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on 20*(1), 45–57.