

# Diseased Region Detection of Longitudinal Knee Magnetic Resonance Imaging Data

Chao Huang, Liang Shan, H. Cecil Charles, Wolfgang Wirth, *Member, IEEE*, Marc Niethammer, and Hongtu Zhu\*

**Abstract**—Magnetic resonance imaging (MRI) has become an important imaging technique for quantifying the spatial location and magnitude/direction of longitudinal cartilage morphology changes in patients with osteoarthritis (OA). Although several analytical methods, such as subregion-based analysis, have been developed to refine and improve quantitative cartilage analyses, they can be suboptimal due to two major issues: the lack of spatial correspondence across subjects and time and the spatial heterogeneity of cartilage progression across subjects. The aim of this paper is to present a statistical method for longitudinal cartilage quantification in OA patients, while addressing these two issues. The 3D knee image data is preprocessed to establish spatial correspondence across subjects and/or time. Then, a Gaussian hidden Markov model (GHMM) is proposed to deal with the spatial heterogeneity of cartilage progression across both time and OA subjects. To estimate unknown parameters in GHMM, we employ a pseudo-likelihood function and optimize it by using an expectation-maximization (EM) algorithm. The proposed model can effectively detect diseased regions in each OA subject and present a localized analysis of longitudinal cartilage thickness within each latent subpopulation. Our GHMM integrates the strengths of two standard statistical methods including the local subregion-based analysis and the ordered value approach. We use simulation studies and the Pfizer longitudinal knee MRI dataset to evaluate the finite sample performance of GHMM in the quantification of longitudinal cartilage morphology changes. Our results indicate that GHMM significantly outperforms several standard analytical methods.

**Index Terms**—Diseased regions detection, EM algorithm, Gaussian hidden Markov model, longitudinal cartilage thickness, pseudo-likelihood method.

Manuscript received January 22, 2015; revised March 05, 2015; accepted March 08, 2015. Date of publication March 23, 2015; date of current version August 28, 2015. This work was supported in part by NIH Grants R025747-01, CA142538-01, MH086633, B005149-01, R01 MH091645-01A1, R21 AR059890, 5R21AR059890-02, and the Pfizer longitudinal study on osteoarthritis (PLS-A9001140). *Asterisk indicates corresponding author.*

C. Huang is with the Department of Mathematics, Southeast University, Nanjing 210018 China, and also with the Department of Biostatistics, and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

L. Shan and M. Niethammer are with the Department of Computer Sciences, and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

H. C. Charles is with Department of Radiology, Duke University, Durham, NC 27710 USA.

W. Wirth is with Institute of Anatomy and Musculoskeletal Research, Paracelsus Medical University, 5020, Salzburg, Austria and also with Chondrometrics GmbH, 83404 Ainring, Germany.

\*H. Zhu is with the Department of Biostatistics, and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (e-mail: hzhu@bios.unc.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2015.2415675

## I. INTRODUCTION

OSTEOARTHRITIS (OA) known as degenerative arthritis represents a group of mechanical abnormalities involving degradation of joints, including articular cartilage and subchondral bone. Symptoms include swelling, pain, discomfort, locking, and other problems in mobility and are often characterized by the progressive loss of joint cartilage [1]. Bone may be exposed and damaged without the protection of cartilage. Fig. 1 shows the anatomy of the human knee and illustrates the cartilage loss. OA is the most common form of arthritis and a major cause of chronic disability in the United States [2]. It is estimated that more than 16% of all adults 45 years or older suffer from symptomatic OA of the knee [1]. Cartilage loss [3] is believed to be the dominating factor in OA. Studying cartilage morphological changes can help to understand OA progression and drug development.

In recent years, magnetic resonance imaging (MRI) has increasingly been used in OA diagnosis [4], [5], since MRI offers the distinct advantage of detecting signal and morphologic changes in articular cartilage in 3D. MRI is the most sensitive and accurate test for evaluating the articular cartilage noninvasively [6]. Particularly, MRI can demonstrate reactive bone edema or soft tissue swelling as well as small cartilage or bone fragments in the joint. However, little is known about the development of such cartilage lesions and their natural progression.

Recently, several large MRI databases have been acquired for OA research in order to help understand the progression of cartilage loss across OA subjects and to identify risk factors (such as genetics, dietary intake, and muscle weakness) associated with such loss [7]–[10]. For example, the Pfizer A9001140 is a longitudinal study, including baseline, 3-month, 6-month, 1-year and 2-year follow-up images of a single knee from 155 subjects [7], [8]; the Osteoarthritis Initiative (OAI) is another longitudinal cohort study of risk factors for incidence and progression of OA, where 4796 subjects with or at high risk of knee OA were recruited and image data was collected annually during 4 years of follow-up [9], [10]. Moreover, analyzing such large MRI datasets is necessary to establish knowledge about the expected pace of progression in knees with different exposure to risk factors for structural progression (e.g., radiographic joint space narrowing [JSN]). Such knowledge can then be used for the design of interventional trials.

Statistical analysis of cartilage thickness changes faces two major challenges. The first one is the lack of spatial correspondence across subjects and/or time. Observational studies have shown that cartilage loss in OA may not be uniform throughout

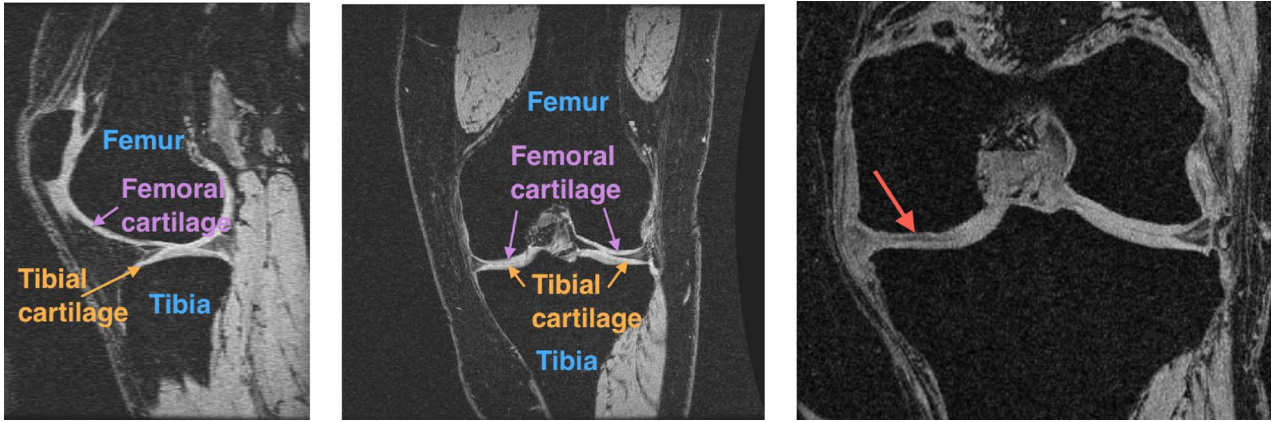


Fig. 1. Anatomy of human knee and illustration of cartilage loss. (left) A sagittal slice of a healthy knee. Bones are annotated in blue, femoral cartilage in purple and tibial cartilage in orange. (middle) A coronal slice of the same healthy knee. (right) A coronal slice of an OA knee with cartilage loss indicated by the red arrow.

the cartilage [11], and advanced statistical analysis methods are necessary to understand localized cartilage thinning. However, the localized analysis requires establishing spatial correspondence across time and subjects, and such analysis is challenging due to the small volume of cartilage in relation to the rest of the knee. The second challenge is the spatial heterogeneity of cartilage progression across subjects. Cartilage thinning may happen at different locations in different subjects, which is usually caused by biomechanical factors [12] or local risk factors [13]. Cartilage thinning at a particular location may only be consistent for a subset of the full subject population. Thus, treating all OA subjects equally and using all of them for standard statistical methods (such as voxel-wise analysis and region-of-interest analysis) would be problematic. In clinical practice, some non-surgical treatments for OA in the knee (e.g., corticosteroid injections) are most likely administered via injection and act globally [14] since the exact locations of lesions are unknown. Therefore, it is of great interest to propose some methods to detect such locations and refine these treatments.

Subregion-based analysis and the ordered value approach are two major statistical methods for quantifying cartilage thickness changes. Subregion-based analysis is one of the most popular approaches. Here, average thickness changes in defined anatomic subregions of the femorotibial joint are studied. In [15], the tibial cartilage was divided into a central area of the total subchondral bone area, and anterior, posterior, internal, and external subregions surrounding it. However, this approach is problematic due to the spatial heterogeneity of cartilage progression discussed above. In another region-wise statistical analysis method [16], a region partitioning process is performed by using a supervised machine learning framework on anatomically aligned medial tibial and medial femoral cartilage thickness maps, and then a point distribution model is used to establish spatial correspondence and a common coordinate system. However, their machine learning procedures are only implemented on the thickness maps at baseline. Buck *et al.* [17] proposed an ordered value approach by sorting subregional changes within each knee by ascending order to focus on the magnitude of the thickness irregardless of the location, in which subregion the change occurs in each subject. The ordered value approach was demonstrated to provide improved

discrimination between healthy subjects and OA participants longitudinally [17]. The key advantage of the ordered value approach is to remove the link between the spatial location and the magnitude/direction of the change in cartilage thickness. But the ordered value approach still relies on subdivisions of the cartilage and the average thickness within each subregion. The subdivisions of the cartilage are purely geometric, and small local changes (that happen to a smaller region than the size of a subregion) can be difficult to recover, since they may be weakened by averaging over a large subregion. Therefore, to fully understand the spatial pattern of OA progression, it is necessary to pinpoint localized cartilage thickness changes.

The paper is organized as follows. Section II summarizes the image preprocessing pipeline and the key contributions in this paper. Section III introduces a novel Gaussian hidden Markov model (GHMM) and its related EM algorithm. In Section IV, simulation studies are performed to show the advantages of our proposed model in diseased regions detection. Section V applies GHMM to the 2D thickness maps derived from the 3D MRI data of the Pfizer Longitudinal Study (PLS-A9001140).

## II. IMAGE PREPROCESSING AND MAIN CONTRIBUTIONS

In this paper, the 3D knee MR images are preprocessed via a set of four image analysis steps including (I.1) segmentation, (I.2) 3D thickness map computation, (I.3) registration, and (I.4) thickness map projection. These four methods summarized in Fig. 2 can successfully establish spatial correspondence across subjects and/or time and therefore address the first challenge in statistical analysis of cartilage thickness changes.

The main contributions of this paper include (i) a statistical method based on GHMM to appropriately address the spatial heterogeneity of cartilage progression across OA subjects and quantify their longitudinal cartilage morphology changes and (ii) a systematic evaluation of our analysis method by using simulations and real data analysis. Moreover, we have developed companion software for GHMM and will release it to the public through <http://www.nitrc.org/> and <http://www.bios.unc.edu/research/bias>.

The path diagrams of GHMM are presented in Fig. 3. Specifically, a novel hidden Markov random field (HMRf) [18]–[27] is proposed to specifically discover the diseased regions for each

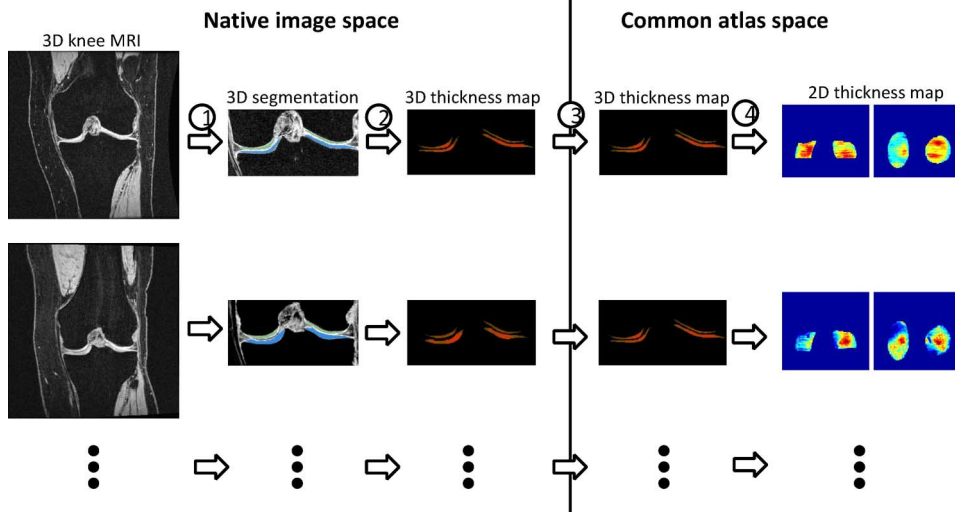


Fig. 2. Path diagram of the 3D knee MR image data preprocessing toolbox: ① extracting the joint region and obtaining the cartilage segmentation; ② computing the 3D cartilage thickness map; ③ registering all thickness maps to a common atlas space; and ④ projecting 3D thickness map onto a 2D plane.

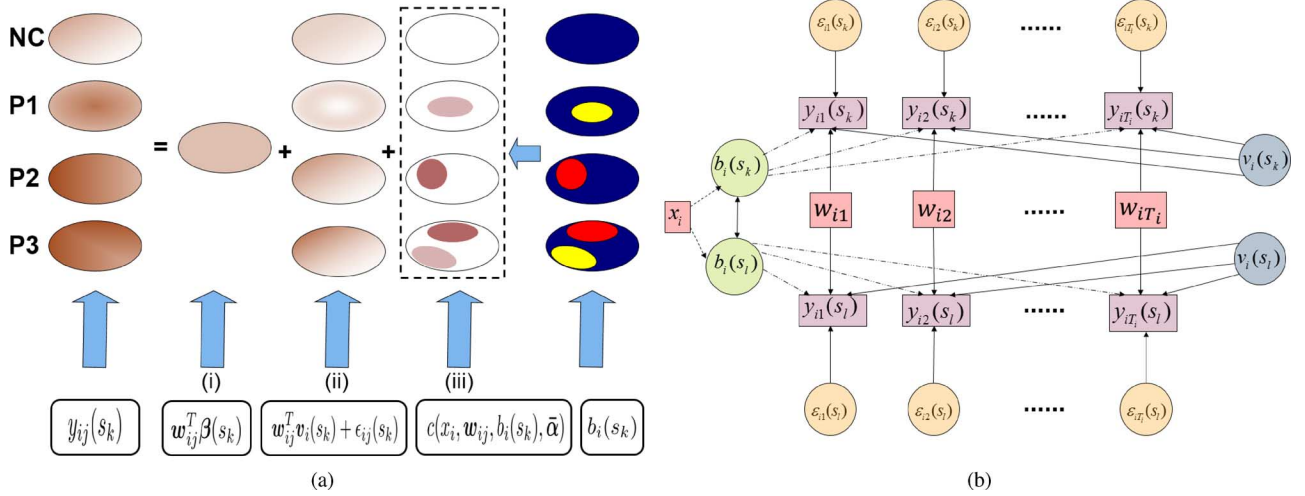


Fig. 3. Path diagrams of GHMM component. (a) Path diagram of GHMM for four different subjects (one normal control and three OA subjects). The change of the map thickness at pixel site  $k$  includes three parts in the proposed model, i.e., (i) the common dynamic change in both normal controls and OA subjects, (ii) the change caused by random effects and measurement errors, and (iii) the change caused by different diseased regions (moderately diseased region (yellow) and severely diseased region (red)). (b) Path diagram of GHMM for one subject at two neighboring pixel sites across time. Following the convention of path diagrams, observed variables are framed with rectangles whereas latent variables are framed with ellipses. Regression associations are depicted as (solid) one-headed arrows (parent=predictor, child=response); correlations between diagnosis result, the diseased region label and the observed thickness are described as (dotted-dashed) one-headed arrows; and the correlation of diseased region labels in the Potts model is shown as (solid) double-headed arrow. For easier interpretation, the following color-codes are used: exogenous variables (light red), endogenous variables (purple), continuous latent variables (blue), discrete latent variables (green), and measurement errors (yellow).

OA subject and to integrate cartilage progression in different diseased regions with different change magnitudes/directions across OA subjects. Moreover, based on the correspondence established in the preprocessing step, our GHMM allows us to carry out a localized analysis of longitudinal cartilage thickness at each voxel. Therefore, GHMM integrates the strengths of both the local subregion-based analysis and the ordered value approach in a single framework.

### III. BACKGROUND AND METHODS

#### A. Cartilage Segmentation

Cartilage segmentation in the knee has been investigated for several decades. Many segmentation approaches have been proposed, e.g., region growing methods [28], [29], watershed

methods [30], live wire methods [31], active contour methods [32], [33], and graph cut methods [34]. However, all the aforementioned methods are semi-automatic, which precludes their applications to large image databases. Recently, several automatic methods have been proposed for cartilage segmentation, including pattern recognition methods [35], model-based segmentation method [36], and graph-based method [37]. However, the spatial interactions between neighboring pixels are neglected in [35]; the model-based segmentation method [36] is prone to local minima in the fitting process; the graph-based method [37] generally suffers from grid bias, also known as metrication errors.

Here we applied the automatic cartilage segmentation method [38], which is an extension of the multi-atlas segmentation approach initially proposed by us for cartilage segmentation in

[39]. The main difference in these two cartilage segmentation approaches is how label-fusion is achieved once multiple segmentation candidates (from the multiple atlases) are available. In [39], a simple locally weighted label-fusion strategy was used, whereas in [38] the locally weighted label-fusion was compared to the baseline majority-voting approach and a more advanced non-local patch-based label fusion approach was also investigated. The latter approach turned out to be superior to the other two label-fusion strategies. In general, the segmentation method [38] has the following advantages. First, the method is fully automatic and requires no user interaction (besides quality control). Secondly, the method is robust as it benefits from multi-atlas-based strategies. Thirdly, both spatial and local appearance information are utilized in the segmentation. Local tissue classification is probabilistic (unlike [35]) and combined with the spatial prior to generate the final segmentation within a segmentation model. Furthermore, the segmentation model is convex and thus allows for the computation of global optimal solutions which cannot be guaranteed by active contour models, active shape models or active appearance models. The segmentation model also allows for the incorporation of spatial and temporal regularization.

The approach that we use for segmenting cartilage and to compute local thicknesses works in a three-dimensional (3D) space. Specifically, our segmentation produces a 3D label map for femoral and tibial cartilage. Given a segmentation, we compute thickness using a Laplace-equation approach [40] through which we define a top and a bottom of the structure and correspondence lines between them. Each point on such a correspondence line is then assigned the thickness value (distance between beginning and end of this line) resulting in a 3D distance map. It should be noted that, all the segmentations are performed for each subject independently and in their own untransformed coordinate systems to avoid any bias introduced by resampling the image. For the same reason, the thickness maps are computed in the native image spaces obtained from cartilage segmentations.

Once these measures have been computed, population comparisons require a common coordinate system for the statistical analysis across time-points and subjects, which requires an additional registration step to this common space. Note that this may cause slight local distortions (depending on the quality of the registration), but it will not change the actual thickness measures as they are computed in the native image spaces. Here, we have used an affine transformation followed by a B-Spline transform based on the bone segmentations for each cartilage thickness volume. Since we are dealing with sheet-like structures, which we have approximately axis-aligned in atlas space, we simply compute the local median values of this 3D thickness map in the axial directions, where thicknesses are approximately constant.

### B. Gaussian Hidden Markov Model

We consider a longitudinal thickness map dataset with thickness data  $\{y_{ij}(s_k) : k = 1, \dots, m\}$  measured at time  $t_{ij}$  for  $j = 1, \dots, T_i$  and  $i = 1, \dots, n$ , where  $n$  is the total number of subjects,  $S_0 = \{s_1, \dots, s_m\}$  is the set of  $m$  pixels in a common template, denoted by  $\mathcal{S}$ , and  $T_i$  is the total number of time points

for the  $i$ -th subject. Let  $x_i$  represent disease status of OA for each subject such that  $x_i = 0$  and 1, respectively, represent normal control and patient with OA.

We introduce a latent partition of  $\mathcal{S}$  for each subject. Specifically, it is assumed that for the  $i$ -th subject,  $\mathcal{S}$  can be decomposed into the union of normal region  $R_{i0}$ , moderately diseased region  $R_{i1}$ , and severely diseased region  $R_{i2}$ , that is

$$\mathcal{S} = R_{i0} \cup R_{i1} \cup R_{i2} \quad \text{and} \quad R_{ic} \cap R_{ic'} = \emptyset \quad \text{for } c \neq c'.$$

Here we also assume that normal controls are expected to be perfectly healthy, i.e., to not have any cartilage loss or lesions. Then, both  $R_{i1}$  and  $R_{i2}$  should be the empty set for normal controls. For OA patients, the size and location of  $R_{i1}$  and  $R_{i2}$  may vary across subjects and play a critical role in establishing spatial correspondences of diseased regions at different locations across subjects. Statistically, we introduce an unobserved random effect  $b_i(s_k) \in L = \{0, 1, 2\}$  to label  $R_{i0}$ ,  $R_{i1}$ , and  $R_{i2}$  at each pixel  $s_k$  of  $\mathcal{S}$  for the  $i$ -th subject. Thus,  $b_i(s_k)$ s are label configurations of three nonoverlapping regions  $\{R_{i0}, R_{i1}, R_{i2}\}$ . It is also possible to extend GHMM to the scenarios with more than two types of diseased regions.

Our Gaussian hidden Markov model consists of a spatial random effect model and a Potts model [20], [21], [23]. Spatial random effects models (SREM), also called hidden Markov models, are very useful for image segmentation and restoration in image analysis [18], [19], [22], [24]–[27].

We first propose SREM to model the conditional distribution of the observed thickness maps given two sets of random effects including  $\mathbf{b}_i = \{b_i(s_k) : s_k \in \mathcal{S}\}$  and  $\mathbf{v}_i = \{\mathbf{v}_i(s_k) : s_k \in \mathcal{S}\}$ , where  $\mathbf{v}_i(s_k)$  will be defined below. We present the path diagrams of SREM in Fig. 3. Given  $\mathbf{b}_i$  and  $\mathbf{v}_i$ , we propose a general spatial random effect model as

$$y_{ij}(s_k) = \mathbf{w}_{ij}^T \boldsymbol{\beta}(s_k) + \mathbf{w}_{ij}^T \mathbf{v}_i(s_k) + \epsilon_{ij}(s_k) + c(x_i, \mathbf{w}_{ij}, b_i(s_k), \bar{\boldsymbol{\alpha}}), \quad (1)$$

where  $\mathbf{w}_{ij}$  is a  $q_w \times 1$  vector of covariates (e.g., time, gender, or genetic marker) that usually includes an intercept 1, but is independent of disease status. Moreover,  $\boldsymbol{\beta}(s_k)$  and  $\mathbf{v}_i(s_k)$  are both  $q_w \times 1$  vectors of regression coefficients. It should be noted that, if  $\mathbf{w}_{ij}$  only comprises an intercept 1 and time,  $\boldsymbol{\beta}(s_k)$  represents both the average thickness at baseline and average thickness change rate at time point  $t_{ij}$ , while the two components in  $\mathbf{v}_i(s_k)$  indicate the individual thickness difference relative to the average thickness at baseline and the individual difference of the thickness change rate relative to the average thickness change rate at time point  $t_{ij}$  respectively. The  $c(x_i, \mathbf{w}_{ij}, b_i(s_k), \bar{\boldsymbol{\alpha}})$  is defined as

$$c(x_i, \mathbf{w}_{ij}, b_i(s_k), \bar{\boldsymbol{\alpha}}) = \begin{cases} 0, & x_i = 0, \\ \sum_{l=1}^2 \mathbf{w}_{ij}^T \bar{\boldsymbol{\alpha}}(l) \delta(b_i(s_k), l), & x_i \neq 0, \end{cases}$$

where  $\delta(\cdot, \cdot)$  is the Kronecker function and  $\bar{\boldsymbol{\alpha}} = (\bar{\boldsymbol{\alpha}}(1)^T, \bar{\boldsymbol{\alpha}}(2)^T)^T$  is a  $2q_w \times 1$  vector of coefficients, in which  $\bar{\boldsymbol{\alpha}}(1)$  and  $\bar{\boldsymbol{\alpha}}(2)$  respectively characterizes the dynamic thickness differences in the diseased regions  $R_{i1}$  and  $R_{i2}$  in addition to the dynamic thickness changes at pixel  $s_k$  in normal regions. Thus,  $c(x_i, \mathbf{w}_{ij}, b_i(s_k), \bar{\boldsymbol{\alpha}})$  equals zero for all pixels

for normal controls and the pixels with  $b_i(s_k) = 0$  for OA patients. For OA patients, pixels in different diseased regions may have different dynamic thickness changes. Throughout the paper, we assume  $\mathbf{w}_{ij} = (1, t_{ij})^T$ ,  $\boldsymbol{\beta}(s_k) = (\beta_1(s_k), \beta_2(s_k))^T$  and set  $q_w = 2$  for simplicity. Generally, the time measurement  $t_{ij}$  in the first two terms of model (1) is not required to be the same as the one in the last item of model (1). In principle, *age* is used as  $t_{ij}$  in the first two terms of (1) to capture the average thickness changes across age, whereas *time to diagnosis* is used in the last term of (1) to character the relative thickness changes in the diseased regions caused by OA. However, neither the age nor time to diagnosis are available in this dataset. Instead, the scan date is the only time measurement in this dataset. Since the range of thickness changes is small compared to the thickness at baseline and the time range (2 years) is also small relative to the age of each subject (all subjects are middle aged [41]), it is reasonable to use the scan date as  $t_{ij}$  throughout the model (1). In this paper, for each subject, the first scan date was set to 0 and the following scan dates were calculated in years, e.g., 0.5 means six months. Therefore, it is straightforward to consider some basis functions (such as spline or wavelet) for  $\mathbf{w}_{ij}$ . Moreover,  $\epsilon_{ij}(s_k)$ s are independent measurement errors across subjects, time, and pixels, following  $N(0, \sigma_{s_k}^2)$ .

Second, we model the random effects  $\mathbf{b}_i$  and  $\mathbf{v}_i$  as follows. First, it is assumed that  $\mathbf{b}_i = (b_i(s_1), \dots, b_i(s_m))^T$  and  $\mathbf{v}_i = \{\mathbf{v}_i(s_k) : k = 1, \dots, m\}$  are mutually independent. Moreover,  $\mathbf{v}_i(s_k)$  are assumed to be mutually independent across pixels and  $\mathbf{v}_i(s_k)$  follows  $N(\mathbf{0}, \boldsymbol{\Sigma}_{v_{s_k}})$  at pixel  $s_k$ . It is assumed that  $\mathbf{b}_i$ s are independent across subjects and each  $\mathbf{b}_i$  follows a Potts model [20], [21], [23], whose Gibbs form is given by

$$p(\mathbf{b}_i|\tau) = \exp\{-U(\mathbf{b}_i)\tau - \log C(\tau)\}, \quad (2)$$

where  $U(\mathbf{b}_i) = -\sum_{s_k \sim s_l} \delta(b_i(s_k), b_i(s_l))$  and  $\tau$  is introduced to encourage spatial smoothness in homogeneous regions. Moreover,  $C(\tau)$  is the partition function such that  $p(\mathbf{b}_i|\tau)$  is a probability function. The notation " $\sum_{s_i \sim s_j}$ " means that  $s_i$  is a neighbor of  $s_j$  and each neighboring pair enters the summation only once. Throughout the paper, we only consider the closest 4 neighbors of each pixel.

To illustrate the key ideas of GHMM, we present two path-diagrams of GHMM in Fig. 3(a) and (b). Model (1) has two key novelties compared with the subregion-based analysis and the ordered value approach. As shown in the subcomponents (i) and (ii) of Fig. 3(a), the first novelty is that  $\mathbf{w}_{ij}^T \boldsymbol{\beta}(s_k)$  allows us to utilize information in all normal regions  $R_{i0}$  across all normal subjects and OA patients and then carry out a localized (or pixel-wise) analysis of longitudinal cartilage thickness at each pixel. As shown in the subcomponent (iii) of Fig. 3(a), the second novelty is that the use of  $c(x_i, \mathbf{w}_{ij}, b_i(s_k), \bar{\alpha})$  and  $\bar{\alpha}$  allows us to integrate information in all diseased regions  $R_{i1} \cup R_{i2}$  across all subjects together and then carry out a joint analysis in the union of all moderately diseased regions  $\bigcup_{i=1}^n R_{i1}$  (or all severely diseased regions  $\bigcup_{i=1}^n R_{i2}$ ) across all OA patients.

### C. Estimation Procedure

Our next problem of interest is to estimate all unknown parameters, denoted as  $\boldsymbol{\theta}$ , and random effects. Here  $\boldsymbol{\theta}$  is a vector

of all unknown parameters consisting of  $\tau, \boldsymbol{\beta}(s_k), \bar{\alpha}, \sigma_{s_k}^2$ , and the covariance matrices  $\boldsymbol{\Sigma}_{v_{s_k}}$  for  $k = 1, \dots, m$ . We decompose the parameter  $\boldsymbol{\theta}$  into two parts:  $\tau$ , and all others, denoted as  $\bar{\boldsymbol{\theta}}$ . The maximum likelihood estimate (MLE) of  $\bar{\boldsymbol{\theta}}$  can be calculated by using the expectation-maximization (EM) algorithm [42], whereas  $\tau$  can be estimated by using a pseudo-likelihood method [43] since the MLE of  $\tau$  is generally difficult to compute due to the normalizing part of the probability function in (2).

To derive the EM algorithm, we need the complete-data log-likelihood function of GHMM as follows. Let  $\mathbf{y}_i(s_k) = (y_{i1}(s_k), \dots, y_{iT_i}(s_k))^T$ ,  $\mathbf{W}_i = (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iT_i})$ ,

$$\begin{aligned} \bar{c}(x_i, \mathbf{W}_i, b_i(s_k), \bar{\alpha}) \\ = (c(x_i, \mathbf{w}_{i1}, b_i(s_k), \bar{\alpha}), \dots, c(x_i, \mathbf{w}_{iT_i}, b_i(s_k), \bar{\alpha}))^T, \end{aligned}$$

$\boldsymbol{\Sigma}_{s_k, i} = \sigma_{s_k}^2 \mathbf{I}_{T_i}$ , where  $\mathbf{I}_{T_i}$  is  $T_i \times T_i$  identity matrix. Thus, the distribution of  $\mathbf{y}_i(s_k)$  conditional on random effects is given by  $N(\mathbf{z}_i(s_k), \boldsymbol{\Sigma}_{s_k, i})$ , where  $\mathbf{z}_i(s_k)$  is given by  $\mathbf{W}_i^T (\boldsymbol{\beta}(s_k) + \mathbf{v}_i(s_k)) + \bar{c}(x_i, \mathbf{W}_i, b_i(s_k), \bar{\alpha})$ . Thus, the complete-data log-likelihood function is given by

$$\begin{aligned} \log \tilde{L}(\boldsymbol{\theta}) \propto & -\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^m \log |\boldsymbol{\Sigma}_{s_k, i}| - \frac{n}{2} \sum_{k=1}^m \log |\boldsymbol{\Sigma}_{v_{s_k}}| \\ & - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^m \mathbf{z}_i^T(s_k) \boldsymbol{\Sigma}_{s_k, i}^{-1} \mathbf{z}_i(s_k) - \sum_{i=1}^{n_0} U(\mathbf{b}_i) \tau \\ & - n_0 \log C(\tau) - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^m \mathbf{v}_i^T(s_k) \boldsymbol{\Sigma}_{v_{s_k}}^{-1} \mathbf{v}_i(s_k), \end{aligned} \quad (3)$$

where  $n_0 = \#\{x_i : x_i \neq 0\}$  denotes the number of OA patients. Given the current estimate  $\tilde{\boldsymbol{\theta}}^{(r)}$  at iteration  $r$ , the value  $\tilde{\boldsymbol{\theta}}^{(r+1)}$  is obtained via maximizing the following Q-function  $Q_{\tilde{\boldsymbol{\theta}}^{(r)}}(\bar{\boldsymbol{\theta}}) \doteq E_{\tilde{\boldsymbol{\theta}}^{(r)}}[\log \tilde{L}(\boldsymbol{\theta})|\mathbf{y}, \mathbf{x}]$  with respect to  $\bar{\boldsymbol{\theta}}$ :

$$\begin{aligned} Q_{\tilde{\boldsymbol{\theta}}^{(r)}}(\bar{\boldsymbol{\theta}}) \propto & -\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^m \log |\boldsymbol{\Sigma}_{s_k, i}| - \frac{n}{2} \sum_{k=1}^m \log |\boldsymbol{\Sigma}_{v_{s_k}}| \\ & - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^m E \left[ \mathbf{z}_i^T(s_k) \boldsymbol{\Sigma}_{s_k, i}^{-1} \mathbf{z}_i(s_k) \middle| \mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)} \right] \\ & - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^m E \left[ \mathbf{v}_i^T(s_k) \boldsymbol{\Sigma}_{v_{s_k}}^{-1} \mathbf{v}_i(s_k) \middle| \mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)} \right]. \end{aligned} \quad (4)$$

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

1) *E-Step*: In the E-step, we need to calculate four conditional expectations:  $E[\mathbf{v}_i(s_k) \delta(b_i(s_k), l) | \mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}]$ ,  $E[\mathbf{v}_i(s_k) | \mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}]$ ,  $E[\mathbf{v}_i(s_k) \mathbf{v}_i^T(s_k) | \mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}]$ , and  $E[\delta(b_i(s_k), l) | \mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}]$ . In order to calculate these conditional probabilities, the class labels  $\mathbf{b}$  should be estimated first. Here we consider the maximum a posteriori on Markov random field (MRF-MAP) estimation, which is efficient and adopted in many existing papers, e.g., [23], [44], [45]. First, the

conditional probability density function of  $\mathbf{y}_i(s_k)$  given  $x_i$  and  $b_i(s_k)$  is derived as

$$f(\mathbf{y}_i(s_k)|x_i, b_i(s_k), \tilde{\boldsymbol{\theta}}^{(r)}) \sim \mathcal{N}(\boldsymbol{\varpi}_i(s_k), \tilde{\boldsymbol{\Lambda}}_{s_k, i}^{(r)}), \quad (5)$$

where  $\boldsymbol{\varpi}_i^{(r)}(s_k) = \mathbf{W}_i^T \tilde{\boldsymbol{\beta}}^{(r)}(s_k) + \bar{\mathbf{c}}(x_i, \mathbf{W}_i, b_i(s_k), \tilde{\boldsymbol{\alpha}}^{(r)})$  and  $\tilde{\boldsymbol{\Lambda}}_{s_k, i}^{(r)} = \mathbf{W}_i^T \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i + \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)}$ .

According to the MAP criterion,  $\tilde{\mathbf{b}}_i^{(r)}$  is defined as

$$\begin{aligned} \tilde{\mathbf{b}}_i^{(r)} &= \arg \max_{\mathbf{b}_i} \left\{ \prod_{k=1}^m f(\mathbf{y}_i(s_k)|x_i, b_i(s_k), \tilde{\boldsymbol{\theta}}^{(r)}) p(\mathbf{b}_i|\tilde{\tau}^{(r)}) \right\} \\ &= \arg \min_{\mathbf{b}_i} \left\{ \frac{1}{2} \sum_{k=1}^m [\mathbf{y}_i(s_k) - \boldsymbol{\varpi}_i^{(r)}(s_k)]^T [\mathbf{W}_i^T \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i \right. \\ &\quad \left. + \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)}]^{-1} [\mathbf{y}_i(s_k) - \boldsymbol{\varpi}_i^{(r)}(s_k)] + U(\mathbf{b}_i) \tilde{\tau}^{(r)} \right\}. \end{aligned} \quad (6)$$

To obtain the optimal solution to (6), in this paper, we adopt the iterated conditional modes (ICM) algorithm [20], which uses a greedy iterative strategy for minimization. Convergence is achieved after only a few iterations. As a result, the conditional expectation  $E[\delta(b_i(s_k), l)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)})]$  can be calculated as

$$\frac{f(\mathbf{y}_i(s_k)|x_i, b_i(s_k) = l, \tilde{\boldsymbol{\theta}}^{(r)}) P(b_i(s_k) = l|\tilde{\mathbf{b}}_i^{(r)}, \tilde{\boldsymbol{\theta}}^{(r)})}{\sum_{l=0}^2 f(\mathbf{y}_i(s_k)|x_i, b_i(s_k) = l, \tilde{\boldsymbol{\theta}}^{(r)}) P(b_i(s_k) = l|\tilde{\mathbf{b}}_i^{(r)}, \tilde{\boldsymbol{\theta}}^{(r)})}, \quad (7)$$

where

$$\begin{aligned} P(b_i(s_k) = l|\tilde{\mathbf{b}}_i^{(r)}, \tilde{\boldsymbol{\theta}}^{(r)}) \\ = \exp \left\{ \tilde{\tau}^{(r)} \sum_{s_j \in \mathcal{N}_k} \delta(l, \tilde{b}_i^{(r)}(s_j)) - \log C(\tilde{\tau}^{(r)}) \right\}. \end{aligned} \quad (8)$$

Here  $\mathcal{N}_k$  is the set of sites neighboring  $s_k$ .

Given  $x_i$  and  $\tilde{b}_i^{(r)}(s_k)$ ,  $(\mathbf{y}_i^T(s_k), \mathbf{v}_i^T(s_k))^T$  is normally distributed as

$$\mathcal{N} \left( \begin{pmatrix} \boldsymbol{\varpi}_i(s_k) \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{W}_i^T \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i + \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)} & \mathbf{W}_i^T \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \\ \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i & \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \end{pmatrix} \right).$$

Then, given  $\mathbf{y}_i(s_k)$ ,  $\mathbf{x}$ , and  $\tilde{b}_i^{(r)}(s_k)$ , we have

$$\begin{aligned} &E[\mathbf{v}_i(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{b}_i^{(r)}(s_k), \tilde{\boldsymbol{\theta}}^{(r)}] \\ &= \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i (\mathbf{W}_i^T \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i + \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)})^{-1} (\mathbf{y}_i(s_k) - \boldsymbol{\varpi}_i(s_k)), \\ &E[\mathbf{v}_i(s_k) \mathbf{v}_i^T(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{b}_i^{(r)}(s_k), \tilde{\boldsymbol{\theta}}^{(r)}] \\ &= \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} - \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i (\mathbf{W}_i^T \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \mathbf{W}_i + \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)})^{-1} \mathbf{W}_i^T \tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r)} \\ &\quad + E[\mathbf{v}_i(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{b}_i^{(r)}(s_k), \tilde{\boldsymbol{\theta}}^{(r)}] \\ &\quad \times E[\mathbf{v}_i^T(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{b}_i^{(r)}(s_k), \tilde{\boldsymbol{\theta}}^{(r)}]. \end{aligned}$$

Finally, the desired expectations can be estimated as

$$\begin{aligned} &E[\mathbf{v}_i(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}] \\ &= \sum_{l=0}^2 E[\mathbf{v}_i(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, b_i(s_k) = l, \tilde{\boldsymbol{\theta}}^{(r)}] \\ &\quad \times P(b_i(s_k) = l|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}), \\ &E[\mathbf{v}_i(s_k) \delta(b_i(s_k), l)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}] \\ &= E[\mathbf{v}_i(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, b_i(s_k) = l, \tilde{\boldsymbol{\theta}}^{(r)}] \\ &\quad \times P(b_i(s_k) = l|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}), \\ &E[\mathbf{v}_i(s_k) \mathbf{v}_i^T(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}] \\ &= \sum_{l=0}^2 E[\mathbf{v}_i(s_k) \mathbf{v}_i^T(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, b_i(s_k) = l, \tilde{\boldsymbol{\theta}}^{(r)}] \\ &\quad \times P(b_i(s_k) = l|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}). \end{aligned} \quad (9)$$

2) *M-Step*: Taking derivatives of (4) with respect to  $\tilde{\boldsymbol{\theta}}$  and equating them to zeros, we find the updates of  $\tilde{\boldsymbol{\theta}}$  as follows. For  $\tilde{\boldsymbol{\beta}}(s_k)$  and  $\tilde{\boldsymbol{\alpha}}(l)$ , we have

$$\begin{aligned} &\tilde{\boldsymbol{\beta}}^{(r+1)}(s_k) \\ &= \left[ \sum_{i=1}^n \mathbf{W}_i \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)-1} \mathbf{W}_i^T \right]^{-1} \sum_{i=1}^n \mathbf{W}_i \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)-1} \\ &\quad \left( \mathbf{y}_i(s_k) - \mathbf{W}_i^T E[\mathbf{v}_i(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}] \right. \\ &\quad \left. - \delta(x_i, 1) \mathbf{W}_i^T \sum_{l=1}^M \tilde{\boldsymbol{\alpha}}^{(r)}(l) E[\delta(b_i(s_k), l)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)})] \right), \end{aligned} \quad (10)$$

and

$$\begin{aligned} &\tilde{\boldsymbol{\alpha}}^{(r+1)}(l) = \left[ \sum_{i=1}^{n_0} \sum_{k=1}^m \mathbf{W}_i \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)-1} \mathbf{W}_i^T \right. \\ &\quad \left. E[\delta(b_i(s_k), l)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}] \right]^{-1} \\ &\quad \sum_{i=1}^{n_0} \sum_{k=1}^m \mathbf{W}_i \tilde{\boldsymbol{\Sigma}}_{s_k, i}^{(r)-1} \left\{ \mathbf{y}_i(s_k) - \mathbf{W}_i^T \tilde{\boldsymbol{\beta}}^{(r)}(s_k) \right. \\ &\quad \left. - \mathbf{W}_i^T E[\mathbf{v}_i(s_k) \delta(b_i(s_k), l)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}] \right\}. \end{aligned} \quad (11)$$

For the covariance matrix  $\boldsymbol{\Sigma}_{v_{s_k}}$ , we have

$$\tilde{\boldsymbol{\Sigma}}_{v_{s_k}}^{(r+1)} = \frac{1}{n} \sum_{i=1}^n E[\mathbf{v}_i(s_k) \mathbf{v}_i^T(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}]. \quad (12)$$

For  $\sigma_{s_k}^2$  in  $\boldsymbol{\Sigma}_{s_k, i}$ ,

$$\tilde{\sigma}_{s_k}^{2(r+1)} = \frac{1}{N} \sum_{i=1}^n E[\mathbf{z}_i^{(r)}(s_k)^T \mathbf{z}_i^{(r)}(s_k)|\mathbf{y}_i(s_k), \mathbf{x}, \tilde{\boldsymbol{\theta}}^{(r)}], \quad (13)$$

where  $N = \sum_{i=1}^n T_i$ ,

$$\begin{aligned} \mathbf{z}_i^{(r)}(s_k) &= \mathbf{y}_i(s_k) - \mathbf{W}_i^T \tilde{\boldsymbol{\beta}}^{(r)}(s_k) \\ &\quad - \bar{\mathbf{c}}(x_i, \mathbf{W}_i, b_i(s_k), \tilde{\boldsymbol{\alpha}}^{(r)}) - \mathbf{W}_i^T \mathbf{v}_i(s_k). \end{aligned}$$



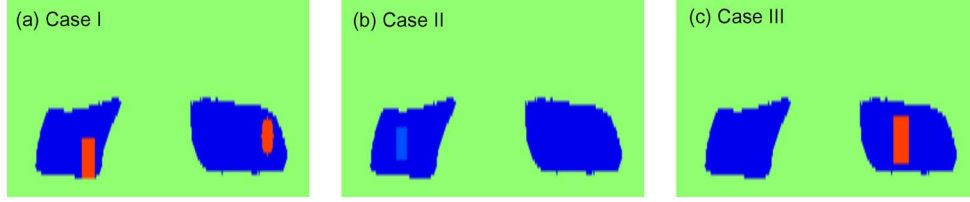


Fig. 4. Visualizations of simulated lesions in femoral cartilage sheets: normal region (*blue*); moderately diseased region (*light blue*); severely diseased region (*orange*).

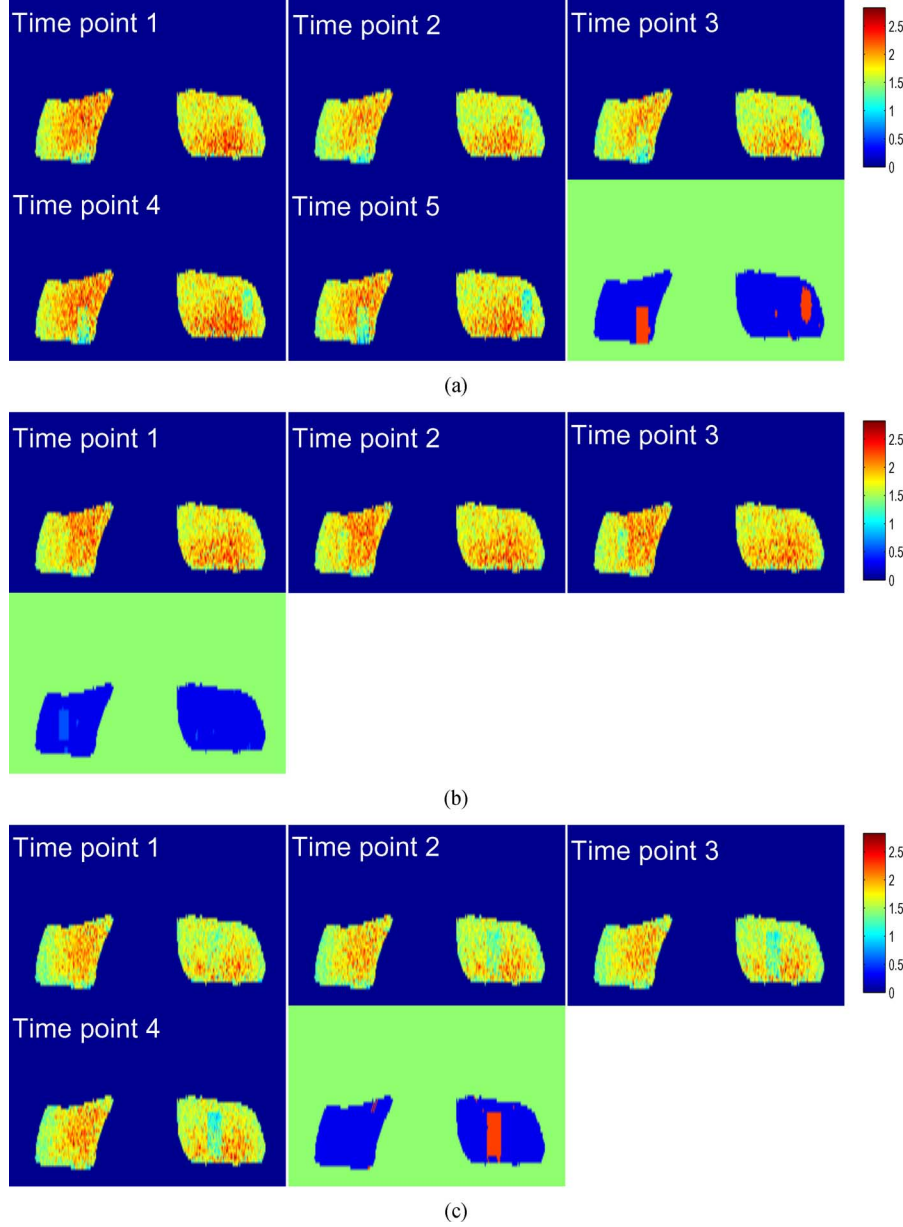


Fig. 5. Diseased region detection for subjects from three groups: (a) subject with 5 time points in group 1; (b) subject with 3 time points in group 2; (c) subject with 4 time points in group 3. This is the simulated result and hence we have an imposed ground truth, which is shown in Fig. 4. In each group, the image with *green* background indicates the detection result, i.e., normal region (*blue*); moderately diseased region (*light blue*); severely diseased region (*orange*), the remaining ones are the observations for the time points, where the color bar for all the observations is presented on the right side of the first row in (a)–(c).

Since  $\tau$  in model (2) 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  $r$ -th step is a simple product of the conditional likelihood

$$PL(\tilde{\mathbf{b}}^{(r)}) = \prod_{i=1}^{n_0} \prod_{s_k \in \mathcal{S} - \partial \mathcal{S}} PL(\tilde{b}_i^{(r)}(s_k) | \tilde{\mathbf{b}}_i^{(r)}), \quad (14)$$

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

$$\frac{\exp\left\{\tau \sum_{s_j \in \mathcal{N}_k} \delta(\tilde{b}_i^{(r)}(s_k), \tilde{b}_i^{(r)}(s_j))\right\}}{\sum_{b(s_k)=1}^M \exp\left\{\tau \sum_{s_j \in \mathcal{N}_k} \delta(\tilde{b}_i^{(r)}(s_k), \tilde{b}_i^{(r)}(s_j))\right\}}.$$

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

$$\frac{\partial \ln PL(\tilde{b}^{(r)}, \tau)}{\partial \tau} = 0. \quad (15)$$

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

#### D. Inference Procedure

After all the parameters are estimated, we carry out formal statistical inference consisting of three different statistical tools below. First, we calculate the standard errors of computed MLEs,  $\hat{\beta}(s_k)$ , at each voxel  $s_k$  and  $\hat{\alpha}$  in the moderately and severely diseased regions. As  $\mathbf{w}_{ij}^T \beta(s_k)$  captures the information in all normal regions  $R_{i0}$  across all normal subjects and OA patients, we calculate the standard errors of  $\hat{\beta}(s_k)$  via pixel-wise analysis of longitudinal cartilage thickness on all subjects with  $\hat{b}_i(s_k) = 0$ , in which model (1) is simplified as follows:

$$\mathbf{y}_{ij}(s_k) = \mathbf{w}_{ij}^T \beta(s_k) + \mathbf{w}_{ij}^T \mathbf{v}_i(s_k) + \epsilon_{ij}(s_k). \quad (16)$$

Thus, the estimated covariance matrix of  $\hat{\beta}(s_k)$  can be approximated by

$$\left[ \frac{1}{\tilde{n}(s_k)} \sum_{i=1}^{\tilde{n}(s_k)} \mathbf{w}_i (\mathbf{w}_i^T \hat{\Sigma}_{v_{s_k}} \mathbf{w}_i + \hat{\Sigma}_{s_k, i})^{-1} \mathbf{w}_i^T \right]^{-1}, \quad (17)$$

where  $\tilde{n}(s_k)$  is the number of subjects with  $\hat{b}_i(s_k) = 0$  at voxel  $s_k$ . Then the standard errors can be estimated by the positive square root of the diagonal elements in (17).

On the other hand, the standard error of  $\hat{\alpha}$  is estimated based on data in the voxels of all diseased regions in all OA patients, since  $\hat{\alpha}$  captures the information in all diseased regions  $R_{i1} \cup R_{i2}$  across all subjects together. Since the information of these diseased regions has to be estimated from the data, the standard errors associated with  $\hat{\alpha}$  are the standard errors conditional on the given diseased regions across subjects. Moreover, the pixel-wise analysis is not suitable here, and thus it is impossible to obtain explicit expressions of the estimated standard errors. In order to tackle this problem, a resampling approach, wild bootstrap [46], is considered here. The idea of wild bootstrap is to leave the regressors at their sample value, but to resample the response variable based on the residuals values. This method proceeds as follows:

- 1) Fit the model with the original data and retain the fitted values  $\hat{\mathbf{y}}_i(s_k) = \mathbf{w}_{ij}^T \hat{\beta}(s_k) + c(x_i, \mathbf{w}_{ij}, \hat{b}_i(s_k), \hat{\alpha})$  and the residuals  $\hat{\epsilon}_i(s_k) = \mathbf{y}_i(s_k) - \hat{\mathbf{y}}_i(s_k)$  for  $i = 1, \dots, n$  and  $k = 1, \dots, m$ ;

TABLE I  
AVERAGE DETECTION PERFORMANCE AMONG SUBJECTS  
IN DIFFERENT GROUPS.

	Group 1	Group 2	Group 3
Jaccard index	0.8509 (0.0045)	0.9200 (0.0113)	0.8683 (0.0040)
Dice index	0.9142 (0.0026)	0.9583 (0.0078)	0.9295 (0.0023)
Rand index	0.9794 (0.0114)	0.9592 (0.0124)	0.9721 (0.0097)
Adjusted Rand index	0.9242 (0.0399)	0.8749 (0.0500)	0.8853 (0.0367)

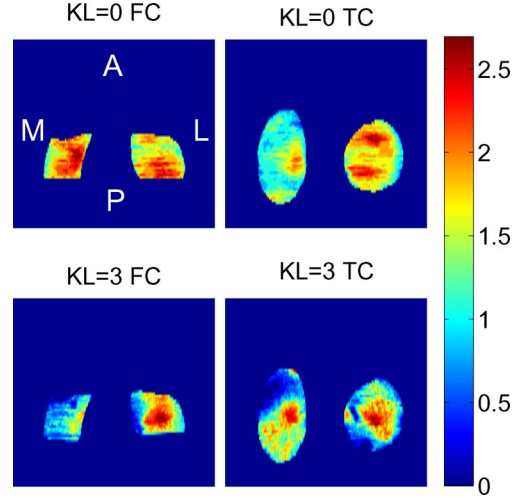


Fig. 6. Local cartilage thickness (mm) for subjects with automatic segmentations at baseline with different KLGs: (top) femoral cartilage (FC) thickness map (left) and tibial cartilage (TC) thickness map (right) for subjects with KLG0; (bottom) femoral cartilage (FC) thickness map (left) and tibial cartilage (TC) thickness map (right) for subject with KLG3. The medial (M), lateral (L), anterior (A), and posterior (P) directions are annotated in the thickness map (top & left).

- 2) Create synthetic response variables  $\mathbf{y}_i(s_k)^* = \hat{\mathbf{y}}_i(s_k) + a_i \hat{\epsilon}_i(s_k)$ , where  $a_i$  is a random variable following standard normal distribution;
- 3) Refit the model using the synthetic response variables  $\mathbf{y}_i(s_k)^*$  and retain the estimates  $\hat{\alpha}^*$ ;
- 4) Repeat Steps 1 and 2 independently  $B$  times to give  $B$  independent realizations of  $\hat{\alpha}^*$ , denoted by  $\hat{\alpha}_1^*, \dots, \hat{\alpha}_B^*$ ;
- 5) The bootstrap covariance matrix of  $\hat{\alpha}(l)$  is estimated by

$$\frac{1}{B-1} \sum_{i=1}^B (\hat{\alpha}_i^*(l) - \overline{\hat{\alpha}^*(l)}) (\hat{\alpha}_i^*(l) - \overline{\hat{\alpha}^*(l)})^T, \quad (18)$$

where  $\overline{\hat{\alpha}^*(l)} = B^{-1} \sum_{i=1}^B \hat{\alpha}_i^*(l)$  for  $l = 1, 2$ . The standard error of  $\hat{\alpha}(l)$  can be estimated by the positive square root of the diagonal element of (18).

It has been shown that 50 to 100 bootstrap replications ( $B = 100$  in this paper) are generally sufficient for standard error estimation [47].

Second, several hypothesis testing problems are quite interesting in real applications. Specifically, we are interested in testing (i) whether there is any cartilage progression across time at each pixel site in normal regions; (ii) whether there is any difference of cartilage thickness at baseline between normal regions and moderately (or severely) diseased regions; and (iii) whether there is any difference of cartilage progression across



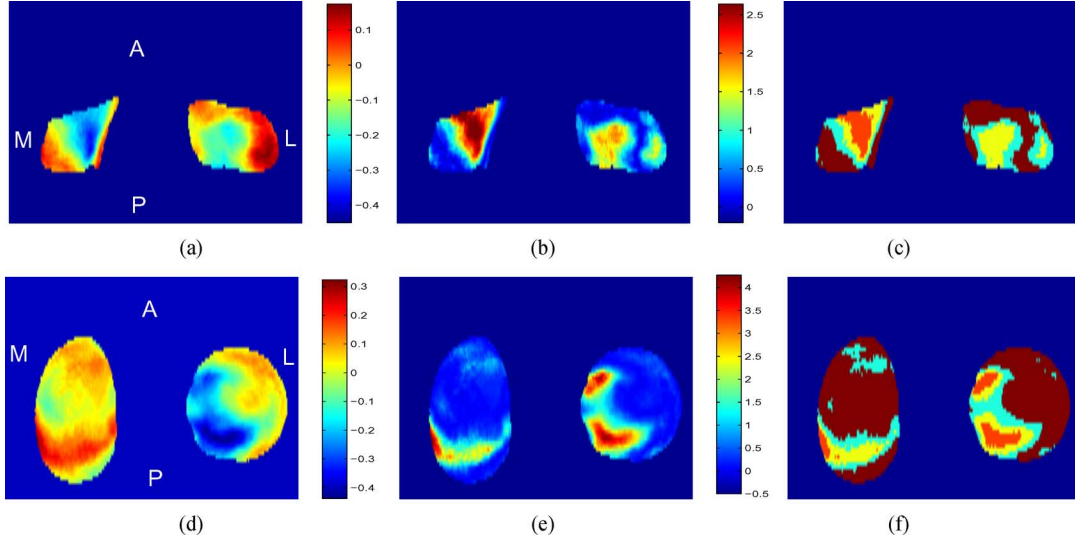


Fig. 7. Voxel-wise analysis based on linear mixed model (25): (a) estimates of  $\alpha_1(s_k)$  in femoral cartilage; (b) FDR adjusted  $-\log_{10}(p)$ -values in femoral cartilage; (c) clustering results based on FDR adjusted  $-\log_{10}(p)$ -values in femoral cartilage: cluster 1 (light blue), cluster 2 (yellow), cluster 3 (orange), cluster 4 (dark red); (d) estimates of  $\alpha_1(s_k)$  in tibial cartilage; (e) FDR adjusted  $-\log_{10}(p)$ -values in tibial cartilage; and (f) clustering results based on FDR adjusted  $-\log_{10}(p)$ -values in tibial cartilage: cluster 1 (light blue), cluster 2 (yellow), cluster 3 (orange), and cluster 4 (dark red).

time between normal regions and moderately (or severely) diseased regions. They can be written as

$$H_0(s_k) : \beta_2(s_k) = 0 \text{ v.s. } H_1(s_k) : \beta_2(s_k) \neq 0; \quad (19)$$

$$H_0 : \bar{\alpha}_1(1) = 0 \text{ v.s. } H_1 : \bar{\alpha}_1(1) \neq 0; \quad (20)$$

$$H_0 : \bar{\alpha}_2(1) = 0 \text{ v.s. } H_1 : \bar{\alpha}_2(1) \neq 0; \quad (21)$$

$$H_0 : \bar{\alpha}_1(2) = 0 \text{ v.s. } H_1 : \bar{\alpha}_1(2) \neq 0; \quad (22)$$

$$H_0 : \bar{\alpha}_2(2) = 0 \text{ v.s. } H_1 : \bar{\alpha}_2(2) \neq 0, \quad (23)$$

where  $\bar{\alpha}_l(l)$  is the  $l^{th}$  component of  $\bar{\alpha}(l)$ ,  $l = 1, 2$ . A sequence of  $t$  tests can be used here. The standard errors of estimates  $\hat{\beta}_2(s_k)$ ,  $k = 1, \dots, m$ , and  $\hat{\alpha}(l)$ ,  $l = 1, 2$  can be obtained as described above. The false discovery rate (FDR) adjustment method [48] is employed here to calculate the adjusted  $p$ -values corrected for the multiple comparison problem (19).

Third, after obtaining the diseased region labels across all pixel sites for each OA subject, we calculate the empirical probabilities of both the moderately diseased region and severely diseased region at each pixel across all OA subjects. The related formula is listed below:

$$\widehat{\Pr}\{b(s_k) = l\} = \frac{\#\{i : \hat{b}_i(s_k) = l\}}{n_0}, \quad l = 1, 2, \quad (24)$$

where  $\#A$  denotes the number of the elements in set  $A$ .

#### IV. SIMULATION STUDIES

We examine the finite sample performance of GHMM for diseased region detection. We first fitted the model (1) to the femoral cartilage thickness data of all the normal controls derived from the 3D MRI data of the Pfizer Longitudinal Study (PLS-A9001140) and then we used the obtained parameter estimators of  $\beta(s_k)$ ,  $\Sigma_{v_{s_k}}$ , and  $\sigma_{s_k}^2$  as their true values for simulations. Moreover,  $\bar{\alpha}(1)$  and  $\bar{\alpha}(2)$  were set  $(-0.17, -0.19)^T$

TABLE II  
MEAN VALUES OF FDR ADJUSTED  $-\log_{10}(p)$ -VALUE  
FOR AT THE PIXEL LOCATIONS IN EACH CLUSTER.

Cartilage	Cluster 1	Cluster 2	Cluster 3	Cluster 4
femoral	1.2565	2.3616	3.8733	0.3111
tibial	1.5198	3.0583	5.0483	0.3491

and  $(-0.26, -0.33)^T$ , respectively. The covariate vector  $w_i$  for each subject was generated according to the real data set analyzed in Section V. We generated 20 subjects from 3 groups of OA patients with 7 patients in group 1, 7 patients in group 2, and 6 patients in group 3. Without loss of generality, the location of all diseased regions is predetermined and does not vary for subjects in the same group, whereas it varies across groups. Fig. 4 presents different lesions that were manually designed and inserted into the thickness maps and the diseased regions associated with the three simulated groups.

We applied GHMM to detect the diseased regions in each subject. We randomly selected one subject from each group and presented its detection results in Fig. 5. For all selected subjects, the diseased regions can be successfully detected with few misclassifications in the results. To show the average detection performance among subjects in each group, the Jaccard index [49], Dice index [50], Rand index [51], and adjusted Rand index [52] were calculated to compare the detection results with the truth. Finally, we calculated the averages of all index values taken over all subjects in each subgroup. For all these indices, a higher index value indicates a more accurate detection result. Simulation results based on all the three groups are presented in Table I, where the related standard deviations are also shown in parentheses. For all these groups, the average of the four index values are all larger than 86 percent, and the related standard deviations are all less than 0.05. Fig. 5 and Table I indicate that GHMM performs very well in diseased region detection of cartilage MRI data.

## V. REAL DATA ANALYSIS

We consider the MRI dataset from the Pfizer longitudinal non-treatment study on osteoarthritis (PLS-A9001140). This dataset contains T1-weighted (3D SPGR) images from 155 subjects imaged at baseline, 3, 6, 12, and 24 months at a resolution of  $1.00 \times 0.31 \times 0.31 \text{ mm}^3$ . All the subjects are females. Some of them have missing scans and thus we have 706 MRIs in total. The Kellgren-Lawrence grades (KLG)<sup>1</sup> [53] were determined for all subjects from the baseline extended weight-bearing radiographs, classifying 81 subjects as normal control (KLG0), 1 subject as KLG1 (mild OA), 40 subjects as KLG2 (severe OA), and 33 subjects as KLG3 (severe OA). It seems that most of the OA patients in this dataset were selected to have predominantly medial OA.

We first computed the cartilage segmentations in the knee MRI data. We randomly chose two subjects with different KLG from the dataset, and presented the related 2D thickness maps for the chosen subjects at baseline in Fig. 6.

After extraction of the 2D thickness maps, we first applied the linear mixed model for voxel-based analysis to the thickness maps from the whole dataset and calculated the parameter estimation at each pixel site. The model is described as below

$$y_{ij}(s_k) = \mathbf{w}_{ij}^T \boldsymbol{\beta}_0(s_k) + \alpha_1(s_k) \delta(x_i, 1) + \mathbf{w}_{ij}^T \mathbf{v}_i(s_k) + \epsilon_{ij}(s_k), \quad j = 1, \dots, T_i; \quad i = 1, \dots, n, \quad (25)$$

where  $\mathbf{w}_{ij}$ ,  $x_i$ ,  $\mathbf{v}_i(s_k)$  and  $\epsilon_{ij}(s_k)$  are all defined the same as those in (1). Compared with GHMM, the reduced model (25) cannot handle the spatial heterogeneity of cartilage progression across subjects, and thus it can only capture the diseased regions at population level instead of individual level. The estimates of  $\alpha_1(s_k)$  across all the pixels in both femoral and tibial cartilages are presented in Fig. 7(a) and (d). We carried out a sequence of  $t$  tests on the hypotheses  $H_0(s_k) : \alpha_1(s_k) = 0$  versus  $H_1(s_k) : \alpha_1(s_k) \neq 0$  across all pixels to identify the regions with substantial differences between OA patients and normal controls. Fig. 7(b) and (e) present the FDR adjusted  $-\log_{10}(p)$ -value [48] at each pixel in both femoral and tibial cartilages. Based on the testing results, we are able to detect some significant subregions in the cartilage. Furthermore, for better visualization, we used the K-means clustering method [54] to cluster voxels into several sub-regions based on the FDR adjusted  $-\log_{10}(p)$ -map. We chose clustering  $-\log_{10}(p)$ -values, since the  $p$ -value indicates the statistical significance of the effect sizes. The clustering results are also presented in Fig. 7(c) and (f), where four subregions are discovered for both femoral and tibial cartilage. The mean values of FDR adjusted  $-\log_{10}(p)$ -value for all the pixel sites in each cluster are also summarized in Table II. Table II reveals that the first three clusters are diseased regions, whereas the fourth one is a normal region. Such results are averages across all the subjects. Since the cartilage thinning happens at different locations for different subjects, the voxel-based analysis based on model (25) fails to capture the diseased regions for individual OA subjects.

<sup>1</sup>According to the definition of KLG [53], the radiographic score is from 0–4, where KLG0 is healthy, KLG1 is considered borderline or mild OA, and KLG2–KLG4 are severe OA.

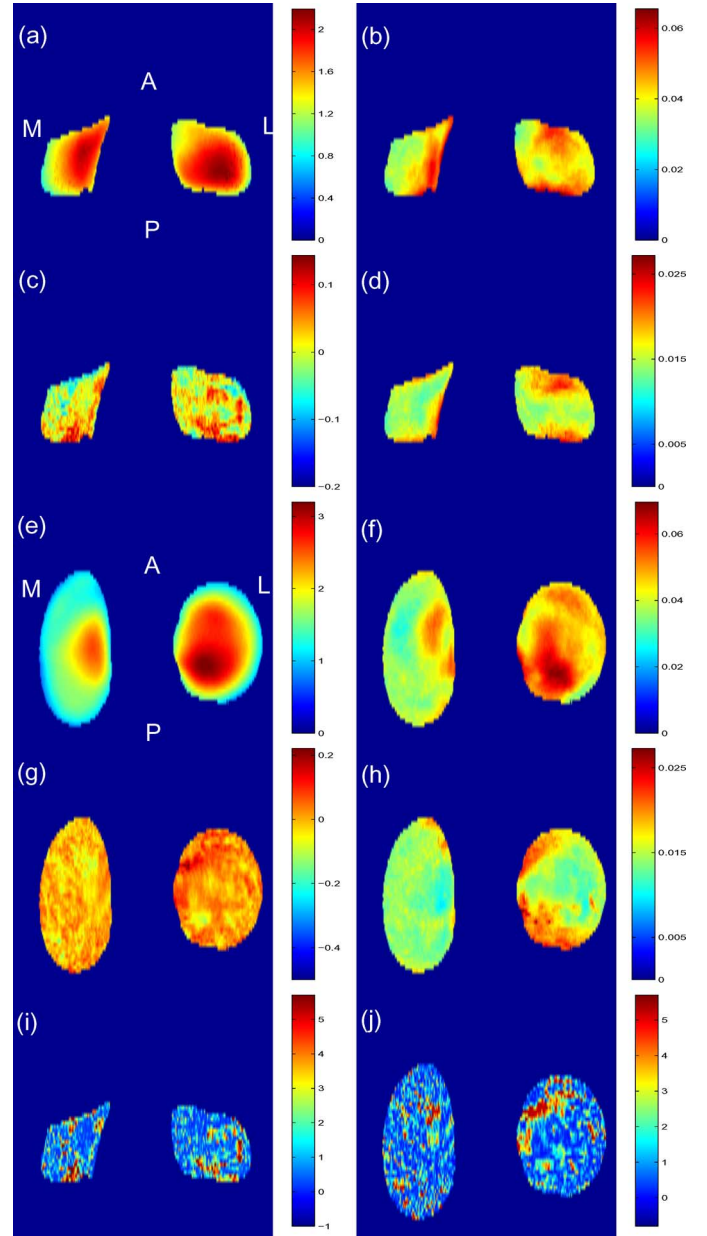


Fig. 8. Inference results of  $\boldsymbol{\beta}(s_k)$  based on GHMM: (a) estimates of  $\beta_1(s_k)$  in femoral cartilage; (b) standard errors of estimates  $\hat{\beta}_1(s_k)$  in femoral cartilage; (c) estimates of  $\beta_2(s_k)$  in femoral cartilage; (d) standard errors of estimates  $\hat{\beta}_2(s_k)$  in femoral cartilage; (e) estimates of  $\beta_1(s_k)$  in tibial cartilage; (f) standard errors of estimates  $\hat{\beta}_1(s_k)$  in tibial cartilage; (g) estimates of  $\beta_2(s_k)$  in tibial cartilage; (h) standard errors of estimates  $\hat{\beta}_2(s_k)$  in tibial cartilage; (i) FDR adjusted  $-\log_{10}(p)$ -values for test (19) across all pixel sites in femoral cartilage; and (j) FDR adjusted  $-\log_{10}(p)$ -values for test (19) across all pixel sites in tibial cartilage.

As a comparison, we applied GHMM to the whole dataset. As an illustration, the estimates of  $\boldsymbol{\beta}(s_k)$  across all the pixels in both femoral and tibial cartilages and their standard errors obtained via wild bootstrap are presented in Fig. 8(a)–(h). From Fig. 8(c)–(d) and (g)–(h), coefficients of most pixels show the range of cartilage thickness changes in normal regions similar to those in [17]. However, coefficients of some pixels show progression in normal regions. One of the reasons for these progressions may be that some regions may undergo tissue swelling,

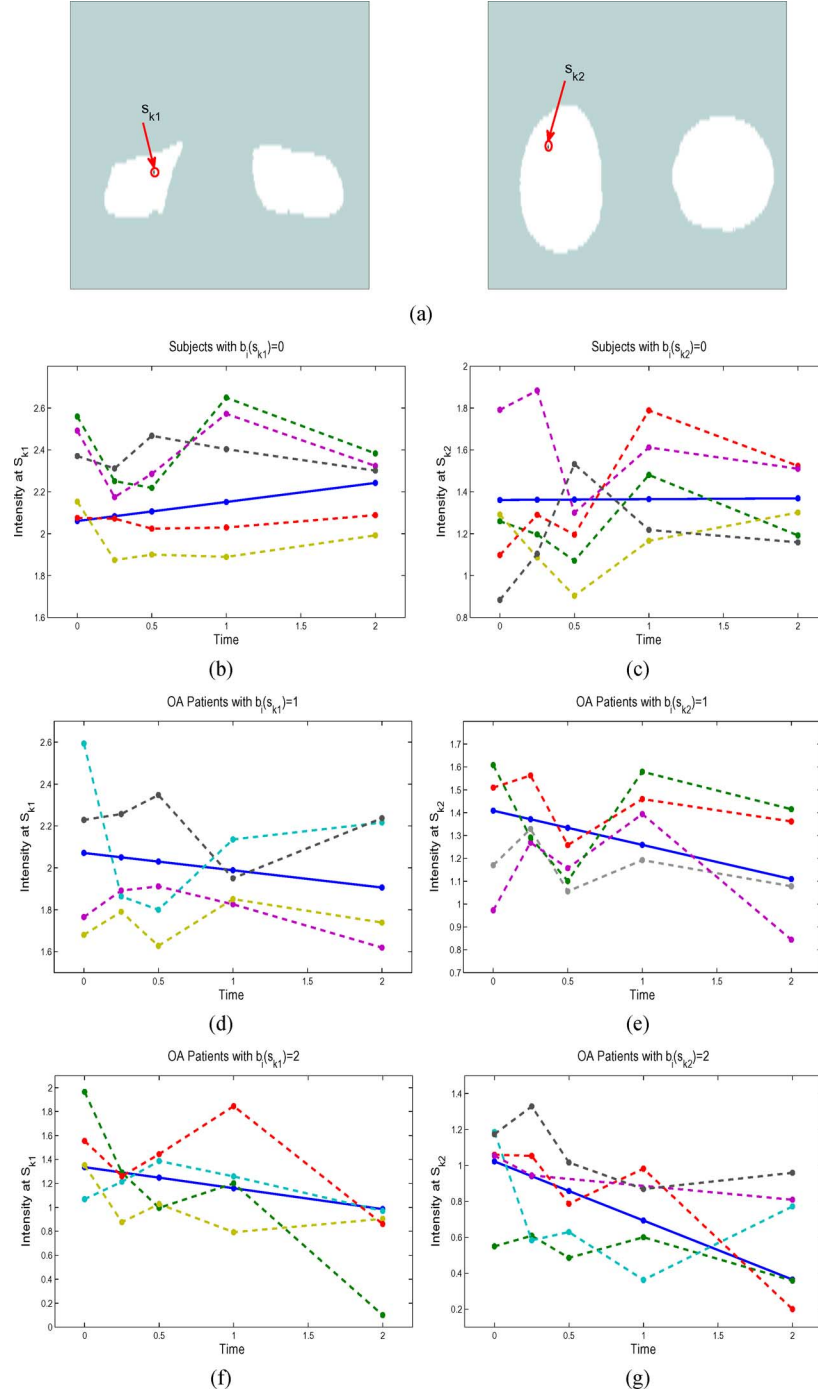


Fig. 9. Dynamic changes of cartilage thickness at randomly selected pixel sites across time: (a) two randomly selected pixel sites on femoral cartilage and tibial cartilage; (b) dynamic changes of cartilage thickness at  $s_{k1}$  of subjects with  $b_i(s_{k1}) = 0$  across time, i.e., estimated curve (solid line) and chosen subjects with  $b_i(s_{k1}) = 0$  (dashed lines); (c) dynamic changes of cartilage thickness at  $s_{k2}$  of subjects with  $b_i(s_{k2}) = 0$  across time, i.e., estimated curve (solid line) and chosen subjects with  $b_i(s_{k2}) = 0$  (dashed lines); (d) dynamic changes of cartilage thickness at  $s_{k1}$  of OA patients with  $b_i(s_{k1}) = 1$  across time, i.e., estimated curve (solid line) and chosen OA patients with  $b_i(s_{k1}) = 1$  (dashed lines); (e) dynamic changes of cartilage thickness at  $s_{k2}$  of OA patients with  $b_i(s_{k2}) = 1$  across time, i.e., estimated curve (solid line) and chosen OA patients with  $b_i(s_{k2}) = 1$  (dashed lines); (f) dynamic changes of cartilage thickness at  $s_{k1}$  of OA patients with  $b_i(s_{k1}) = 2$  across time, i.e., estimated curve (solid line) and chosen OA patients with  $b_i(s_{k1}) = 2$  (dashed lines); and (g) dynamic changes of cartilage thickness at  $s_{k2}$  of OA patients with  $b_i(s_{k2}) = 2$  across time, i.e., estimated curve (solid line) and chosen OA patients with  $b_i(s_{k2}) = 2$  (dashed lines).

diluting, or even masking cartilage loss in adjacent pixels [15]. The FDR adjusted  $-\log_{10}(p)$ -value for test (19) at each pixel is calculated in Fig. 8(i) and (j). The inference results of  $\bar{\alpha}$ , including estimation, standard error and  $-\log_{10}(p)$ -value for tests (20)–(23), are presented in Table III as well. Fig. 8 and Table III reveal that (i) there is significant cartilage progression

across time on only a few pixel sites in normal regions; (ii) the difference of cartilage progression across time between normal regions and moderately (or severely) diseased regions is highly significant with respect to their cartilage thickness at baseline; and (iii) both the difference of cartilage thickness at baseline and difference of cartilage progression across time between normal

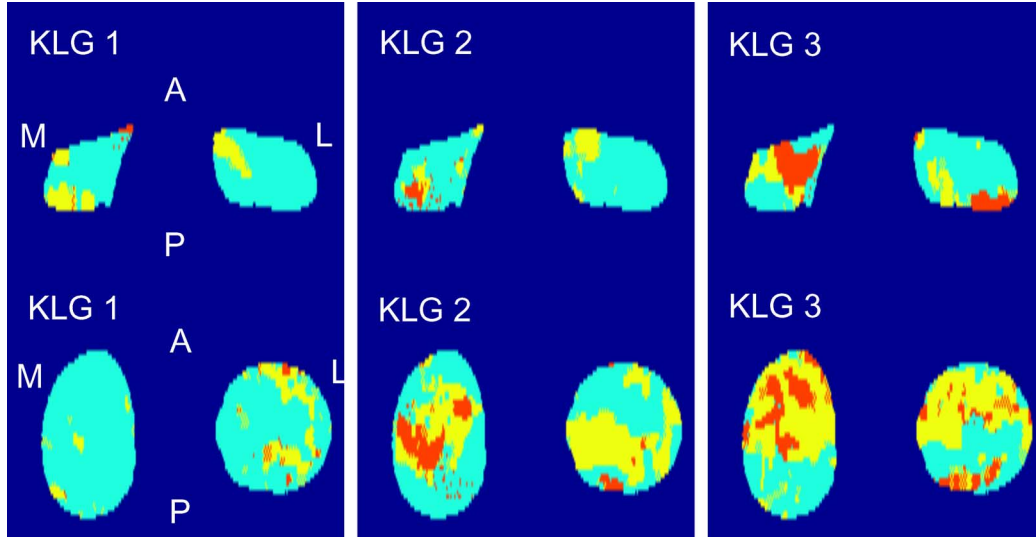


Fig. 10. Diseased region detection for subjects with different KLGs, including 1 KLG1, 1 KLG2, and 1 KLG3. Detection results of femoral cartilage thickness map (*top*) and tibial cartilage thickness map (*bottom*) for every subject are shown in each row, i.e., normal region (*light blue*); moderately diseased region (*yellow*); severely diseased region (*orange*).

regions and severely diseased regions are larger than those between normal regions and moderately diseased regions. Then we chose two pixel sites from the femoral and tibial cartilages, respectively, and presented them in Fig. 9(a). Fig. 9(b)–(g) show the dynamic changes of cartilage thickness at two given pixel sites across time. We also included both the estimated curves and thickness changes of some chosen subjects. The estimated cartilage thicknesses at both two pixel sites follow a trend of decline across time for OA patients. Moreover, the estimated cartilage thickness in severely diseased regions decreases more quickly than that in moderately diseased regions. We randomly selected 3 subjects with different KLGs, i.e., 1 KLG1 subject, 1 KLG2 subject, and 1 KLG3 subject, and presented their diseased regions in Fig. 10. Fig. 11 presents the empirical probabilities of both moderately diseased region and severely diseased region. The locations of diseased regions change across subjects, and the detected diseased regions are more inclined to locate in some regions (e.g., center) of the cartilage. This indicates that the cartilage loss in OA may not be uniform throughout the cartilage, but OA may affect certain regions more frequently than others [11].

Furthermore, we investigate whether there are some subgroups among patients and delineate their associated diseased region pattern in each subgroup. To achieve this, we used a clustering method to cluster all estimated maps  $\{\hat{\mathbf{b}}_i\}_{i \geq 1}$  at the subject level. Since the number of informative pixels is much larger than the number of OA subjects, e.g.,  $m = 4484$  and  $n = 74$  for tibial cartilage, we adopted a sparse k-means clustering method and its associated R package **sparecl** (<http://www.icesi.edu.co/CRAN/web/packages/sparecl/>) [55], which combines a Lasso-type penalty with the K-means clustering method to cluster observations. We applied **sparecl** to the estimated maps  $\{\hat{\mathbf{b}}_i\}_{i \geq 1}$ . Two subgroups were detected for the femoral cartilage, whereas three subgroups were identified for the tibial cartilage. The KLG information of subjects in each

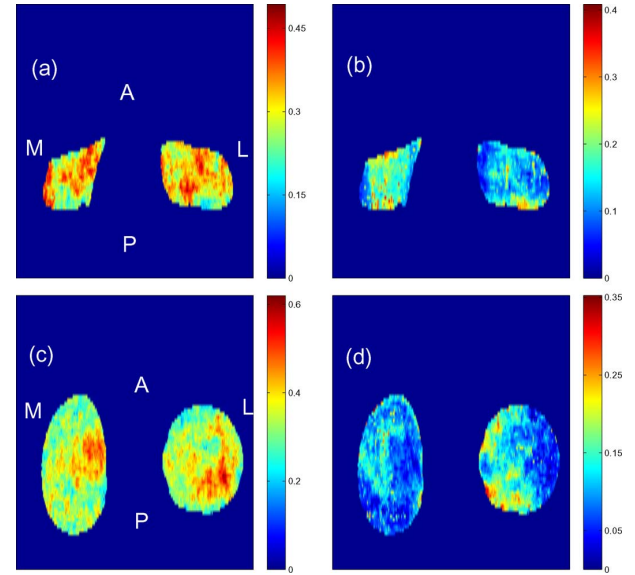


Fig. 11. Empirical probabilities of both moderately diseased region and severely diseased region based on estimated diseased region labels: (a) empirical probability of moderately diseased region in femoral cartilage; (b) empirical probability of severely diseased region in femoral cartilage; (c) empirical probability of moderately diseased region in tibial cartilage; and (d) empirical probability of severely diseased region in tibial cartilage.

subgroup is presented in Table IV. Inspecting Table IV reveals that for the femoral cartilage case, the group membership of most patients is consistent with their KLG information. Specifically, most KLG2 subjects belong to Group 1, whereas most KLG3 subjects fall into Group 2. Similar to the procedures in voxel-based analysis, the diseased region pattern of each group in the average sense can also be calculated based on the detection results. For every group, the diseased-region detection results of two randomly chosen subjects and their related pattern are presented in Fig. A1, which can be found in



TABLE III  
INFERENCE RESULTS OF  $\bar{\alpha}$  BASED ON GHMM.

Parameter	Femoral cartilage			Tibial cartilage		
	Estimation	Standard Error	$-\log_{10}(p)$ -value	Estimation	Standard Error	$-\log_{10}(p)$ -value
$\bar{\alpha}_1(1)$	-0.0104	0.0027	3.9721	-0.0256	0.0080	2.8562
$\bar{\alpha}_2(1)$	-0.1374	0.0353	4.0128	-0.1534	0.0405	3.8111
$\bar{\alpha}_1(2)$	-0.1253	0.0289	4.8366	-0.1387	0.0362	3.9018
$\bar{\alpha}_2(2)$	-0.2619	0.0628	4.5107	-0.2328	0.0584	4.1733

TABLE IV  
KLG INFORMATION OF PATIENTS IN EACH GROUP.

Femoral cartilage				
Group	KLGI	KLGI	KLGI	Total
1	0	36	8	44
2	1	4	25	30
Tibial cartilage				
Group	KLGI	KLGI	KLGI	Total
1	0	26	26	52
2	1	13	0	14
3	0	1	7	8

Supplementary Materials. It reveals the spatial heterogeneity of cartilage progression across OA subjects.

## VI. CONCLUSION

We have developed a GHMM for longitudinal cartilage quantification in OA patients, while addressing these two issues: the lack of spatial correspondence across subjects and time and the spatial heterogeneity of cartilage progression across subjects. Our GHMM is designed to efficiently account for the spatial heterogeneity of cartilage progression across subjects. We have developed an efficient algorithm to estimate unknown parameters in the GHMM. The proposed model can effectively detect diseased regions in each OA subject and present a localized analysis of longitudinal cartilage thickness within each latent subpopulation. Our GHMM integrates the strengths of two standard statistical methods: the local subregion-based analysis and the ordered value approach. Simulation studies and real data analysis have demonstrated that GHMM can effectively detect diseased regions in localizing longitudinal cartilage thickness.

Several important issues need to be addressed in our future research. First, it is assumed that the disease effect  $b_i$  is an ordinal variable at each location and follows a Potts model. However, if the disease effect is strongly associated with the continuous loss of cartilage thickness, then it may be more reasonable to treat them as continuous variables. In this case, computational and theoretical challenges may arise. For example, we may need to introduce another hidden Markov model for continuous disease effects, and thus both the estimation and inference procedures have to be rederived. Second, we plan to apply GHMM to other knee MRI datasets, such as OAI, to investigate other clinical questions of interest. Third, it is interesting to extend GHMM to the group analysis of various neuroimaging data sets, such as the Alzheimer's Disease Neuroimaging Initiative

(ADNI) database[56]. Most group analysis methods (e.g., Statistical Parametric Mapping (SPM)) assume the spatial homogeneity and spatial correspondence of effect regions across subjects, which may not be valid for many neuroimaging studies. For instance, in Alzheimer's disease (AD) studies, quantifying patterns of structural change during early stages of AD is a major challenge. Brain atrophy in the early stages of AD may be relatively subtle and spatially distributed over different brain regions, and the patterns of atrophy have been found spatially heterogeneous[57]. Moreover, the error associated with structural measurements can vary throughout the brain and across subjects and time points. In this case, our GHMM can be used to capture complex morphological characteristics that distinguish AD from normal controls, furthermore to determine which patterns are most informative from a diagnostic perspective.

## ACKNOWLEDGMENT

The authors would like to thank the Editor, the Associate Editor, and two anonymous referees for valuable suggestions, which greatly helped to improve our presentation.

## REFERENCES

- [1] *Arthritis Related Statistics*, Oct. 2013 [Online]. Available: [www.cdc.gov/arthritis/data\\_statistics/arthritis\\_related\\_stats.htm](http://www.cdc.gov/arthritis/data_statistics/arthritis_related_stats.htm)
- [2] A. D. Woolf and B. Pfleger, "Burden of major musculoskeletal conditions," *Bull. World Health Org.*, vol. 81, no. 9, pp. 646–656, 2003.
- [3] D. T. Felson *et al.*, "Osteoarthritis: New insights. Part 1, The disease and its risk factors," *Ann. Internal Med.*, vol. 133, no. 8, pp. 635–646, 2000.
- [4] F. Cicuttini, J. Hankin, G. Jones, and A. Wluka, "Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis," *Osteoarthritis Cartilage*, vol. 13, no. 8, pp. 722–727, 2005.
- [5] J.-P. Raynauld, "Quantitative magnetic resonance imaging of articular cartilage in knee osteoarthritis," *Curr. Opinion Rheumatol.*, vol. 15, no. 5, pp. 647–650, 2003.
- [6] S. Biswal *et al.*, "Risk factors for progressive cartilage loss in the knee," *Arthritis Rheumatism*, vol. 46, no. 11, pp. 2884–2892, 2002.
- [7] R. J. Buck, D. Dreher, and F. Eckstein, "Femorotibial cartilage thickness change distributions for subjects without signs, symptoms, or risk factors of knee osteoarthritis," *Cartilage*, vol. 3, no. 4, pp. 305–313, 2012.
- [8] F. Eckstein *et al.*, "Quantitative imaging of cartilage morphology at 3.0 tesla in the presence of gadopentate dimeglumine (GD-DTPA)," *Magn. Reson. Med.*, vol. 58, no. 2, pp. 402–406, 2007.
- [9] D. T. Felson *et al.*, "Physical activity, alignment and knee osteoarthritis: Data from most and the OAI," *Osteoarthritis Cartilage*, vol. 21, no. 6, pp. 789–795, 2013.
- [10] F. Eckstein, W. Wirth, and M. C. Nevitt, "Recent advances in osteoarthritis imaging: the osteoarthritis initiative," *Nature Rev. Rheumatol.*, vol. 8, no. 10, pp. 622–630, 2012.
- [11] S. Biswal *et al.*, "Risk factors for progressive cartilage loss in the knee," *Arthritis Rheumatism*, vol. 46, no. 11, pp. 2884–2892, 2002.
- [12] K. Moio *et al.*, "Varus-valgus alignment: Reduced risk of subsequent cartilage loss in the less loaded compartment," *Arthritis Rheumatism*, vol. 63, no. 4, pp. 1002–1009, 2011.



- [13] A. Chang *et al.*, "Subregional effects of meniscal tears on cartilage loss over 2 years in knee osteoarthritis," *Ann. Rheumatic Diseases*, vol. 70, no. 1, pp. 74–79, 2011.
- [14] S.-S. Seo, C.-W. Kim, and D.-W. Jung, "Management of focal chondral lesion in the knee joint," *Knee Surg. Related Res.*, vol. 23, no. 4, pp. 185–196, 2011.
- [15] W. Wirth and F. Eckstein, "A technique for regional analysis of femorotibial cartilage thickness based on quantitative magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 27, no. 6, pp. 737–744, Jun. 2008.
- [16] D. R. Jørgensen, E. B. Dam, and M. Lillholm, "Predicting knee cartilage loss using adaptive partitioning of cartilage thickness maps," *Comput. Biol. Med.*, vol. 43, no. 8, pp. 1045–1052, 2013.
- [17] R. J. Buck *et al.*, "Does the use of ordered values of subregional change in cartilage thickness improve the detection of disease progression in longitudinal studies of osteoarthritis?," *Arthritis Care Res.*, vol. 61, no. 7, pp. 917–924, 2009.
- [18] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. R. Stat. Soc. Ser. B (Methodol.)*, pp. 192–236, 1974.
- [19] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 6, pp. 721–741, Nov. 1984.
- [20] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc. Ser. B (Methodol.)*, pp. 259–302, 1986.
- [21] W. Qian and D. Titterton, "Estimation of parameters in hidden Markov models," *Phil. Trans. R. Soc. London. Ser. A, Phys. Eng. Sci.*, vol. 337, no. 1647, pp. 407–428, 1991.
- [22] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel cerebral MR images," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 176–186, Apr. 1997.
- [23] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [24] J. L. Marroquin, E. A. Santana, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1380–1387, Nov. 2003.
- [25] H. Zhu, M. Gu, and B. Peterson, "Maximum likelihood from spatial random effects models via the stochastic approximation expectation maximization algorithm," *Stat. Comput.*, vol. 17, no. 2, pp. 163–177, 2007.
- [26] P. Diggle and P. J. Ribeiro, *Model-Based Geostatistics*. New York: Springer, 2007.
- [27] S. Z. Li and S. Singh, *Markov Random Field Modeling in Image Analysis*. New York: Springer, 2009, vol. 26.
- [28] M. A. Piplani, D. G. Disler, T. R. McCauley, T. J. Holmes, and J. P. Cousins, "Articular cartilage volume in the knee: Semiautomated determination from three-dimensional reformations of MR images," *Radiology*, vol. 198, no. 3, pp. 855–859, 1996.
- [29] F. Eckstein *et al.*, "Determination of knee joint cartilage thickness using three-dimensional magnetic resonance chondro-crassometry (3d MR-CCM)," *Magn. Reson. Med.*, vol. 36, no. 2, pp. 256–265, 1996.
- [30] V. Grau *et al.*, "Improved watershed transform for medical image segmentation using prior information," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 447–458, Apr. 2004.
- [31] A. J. Gougoutas *et al.*, "Cartilage volume quantification via live wire segmentation 1," *Acad. Radiol.*, vol. 11, no. 12, pp. 1389–1395, 2004.
- [32] J. A. Lynch *et al.*, "Cartilage segmentation of 3D MRI scans of the osteoarthritic knee combining user knowledge and active contours," in *Med. Imag.*, 2000, pp. 925–935.
- [33] M. Brem *et al.*, "Magnetic resonance image segmentation using semi-automated software for quantification of knee articular cartilage: initial evaluation of a technique for paired scans," *Skeletal Radiol.*, vol. 38, no. 5, pp. 505–511, 2009.
- [34] K. Bae *et al.*, "Intra-and inter-observer reproducibility of volume measurement of knee cartilage segmented from the OAI MR image set using a novel semi-automated segmentation method," *Osteoarthritis Cartilage*, vol. 17, no. 12, pp. 1589–1597, 2009.
- [35] S. Koo, B. A. Hargreaves, and G. E. Gold, "Automatic segmentation of articular cartilage from MRI," U.S. Patent 8 706 188, Apr. 22, 2014.
- [36] G. Vincent, C. Wolstenholme, I. Scott, and M. Bowes, "Fully automatic segmentation of the knee joint using active appearance models," *Med. Image Anal. Clin., A Grand Challenge*, pp. 224–230, 2010.
- [37] Y. Yin *et al.*, "Logismosa layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint," *IEEE Trans. Med. Imag.*, vol. 29, no. 12, pp. 2023–2037, Dec. 2010.
- [38] L. Shan, C. Zach, C. Charles, and M. Niethammer, "Automatic atlas-based three-label cartilage segmentation from MR knee images," *Med. Image Anal.*, vol. 18, pp. 1233–1246, 2014.
- [39] L. Shan, C. Charles, and M. Niethammer, "Automatic multi-atlas-based cartilage segmentation from knee MR images," in *Proc. 9th IEEE Int. Symp. Biomed. Imag.*, 2012, pp. 1028–1031.
- [40] A. J. Yezzi and J. L. Prince, "An eulerian pde approach for computing tissue thickness," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1332–1339, Oct. 2003.
- [41] R. Stahl *et al.*, "Osteoarthritis of the knee at 3.0 T: Comparison of a quantitative and a semi-quantitative score for the assessment of the extent of cartilage lesion and bone marrow edema pattern in a 24-month longitudinal study," *Skeletal Radiol.*, vol. 40, no. 10, pp. 1315–1327, 2011.
- [42] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser. B (Methodol.)*, pp. 1–38, 1977.
- [43] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," in *Proc. Int. Congr. Math.*, Providence, RI, 1986, vol. 1, p. 2.
- [44] J. L. Marroquin, B. C. Vemuri, S. Botello, E. Calderon, and A. Fernandez-Bouzas, "An accurate and efficient Bayesian method for automatic segmentation of brain MRI," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 934–945, Aug. 2002.
- [45] J. Nie *et al.*, "Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov random field," *Comput. Med. Imag. Graph.*, vol. 33, no. 6, pp. 431–441, 2009.
- [46] C.-F. J. Wu, "Jackknife, bootstrap and other resampling methods in regression analysis," *Ann. Stat.*, pp. 1261–1295, 1986.
- [47] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press, 1994, vol. 57.
- [48] D. Yekutieli and Y. Benjamini, "Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics," *J. Stat. Plan. Inference*, vol. 82, no. 1, pp. 171–196, 1999.
- [49] P. Jaccard, "Etude Comparative de la Distribution Florale dans une Portion des Alpes et du Jura," *Bull. Soc. Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579.
- [50] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecol.*, vol. 26, no. 3, pp. 297–302, 1945.
- [51] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Am. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
- [52] L. Hubert and P. Arabie, "Comparing partitions," *J. Classificat.*, vol. 2, no. 1, pp. 193–218, 1985.
- [53] J. Kellegren and J. Lawrence, "Radiological assessment of osteoarthritis," *Ann. Rheum. Dis.*, vol. 16, pp. 494–501, 1957.
- [54] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probabil.*, 1967, vol. 1, pp. 281–297, 14.
- [55] D. M. Witten and R. Tibshirani, "A framework for feature selection in clustering," *J. Am. Stat. Assoc.*, vol. 105, no. 490, 2010.
- [56] S. G. Mueller *et al.*, "The Alzheimer's disease neuroimaging initiative," *Neuroimag. Clin. N. Am.*, vol. 15, no. 4, pp. 869–877, 2005.
- [57] G. Frisoni, F. Sabatoli, A. Lee, R. Dutton, A. Toga, and P. Thompson, "In vivo neuropathology of the hippocampal formation in AD: A radial mapping MR-based study," *Neuroimage*, vol. 32, no. 1, pp. 104–110, 2006.