Hierarchical Spatial Regression Models for Change Point Analysis

Sarat C. Dass Chae Young Lim Tapabrata Maiti *

Abstract

It is well known that cancer incidence rates are heterogeneous across geographical regions. The aim of this paper is to supplement the existing tools for analyzing cancer rates from the Surveillance, Epidemiology, and End Results (SEER) database that are able to find the change points over time. Subsequently, the model can cluster the geographical subregions based on the magnitude and direction of changes of the disease risk. The proposed model to find change-points over time and cluster spatial locations is based on Dirichlet process priors where we consider temporal functions as the random quantities arising from the Dirichlet process prior. Through the analysis of age adjusted lung cancer mortality rates from 1969 to 2006, the proposed model nicely characterized local data features, namely, the local change points, the rate of changes, and clusters of states that exhibited similar trends of cancer incidence rates. The procedure to extend this model to include covariates therefore enabling selection of meaningful covariates are also discussed.

Key Words: Joinpoint analysis; Disease mapping; Bayesian nonparametrics; Dirichlet process priors.

1. Introduction

Statistical methods for analyzing disease incidence or mortality data over geographical regions and time have gained considerable interest in recent years due to increasing concerns of public health, health disparity and legitimate resource allocations. Cancer is a major threat to public health in the United States and in the world. Cancer accounts for nearly one-quarter of deaths in the United States, exceeded only by heart disease (ACS, www.cancer.org). Among many things, the ACS publishes time trends of age-adjusted cancer death rates for different cancer types, and for different sub-populations defined by geographic and socio-demographic characteristics. However, the impact of cancer surveillance is not uniformly effective over geographical regions; see, for example, Figure 1 which displays cancer trends for four different US states and the overall trend for the nation.

One of the scientific objectives of monitoring cancer rates is to detect changes in the trend over time and identify clusters of sub-populations (generally a set of geographical sub-regions) that are affected by changes (increase or decrease) in risk. Also, if covariates have changes in trend over time, it is interesting to understand how those covariates affect the cancer rates over time. A carefully developed procedure that addresses these issues can help administrators find key information for the prevention of cancer.

Several joinpoint models that identify time points associated with a significant change in disease trend have been developed by several authors (See, Carlin et al. (1992), Kim et al. (2000, 2004), Tiwari et al. (2005) and Ghosh et al. (2009)). The models developed by Kim et al. (2000, 2004), for example, are used in cancer statistics review and implemented in the software of the National Cancer Institute (NCI) (Ries et al. 2002). One important question is whether the rates of cancer incidence before and after the joinpoint is significantly different (statistically speaking) from each other. In other words, we wish to determine if there is a significant *change point* in the cancer incidence rates before and after the joinpoint. Another important concern not addressed by joinpoint modeling is whether there are groups (or, clusters) of states exhibiting similar change-points of cancer incidence rates

*Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824

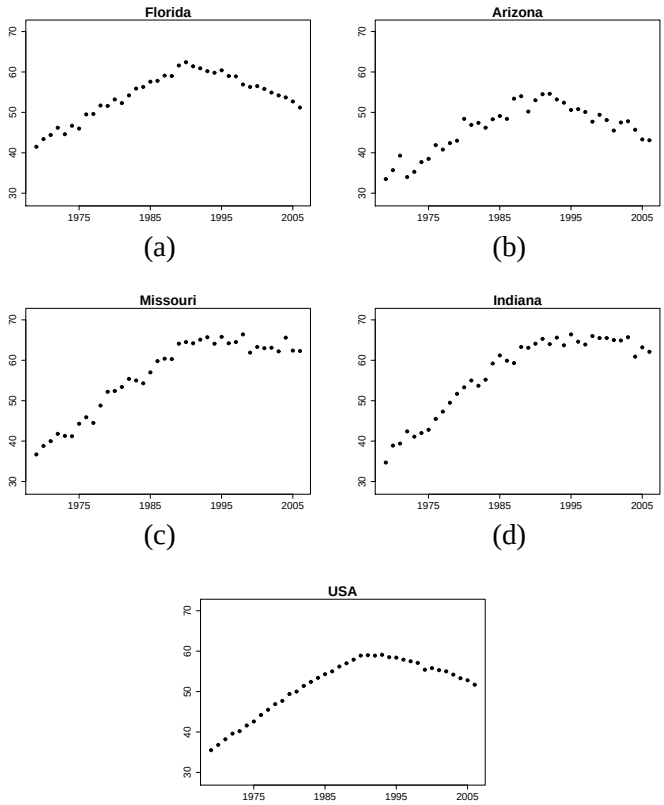

Figure 1: Age-adjusted incidence rates of lung cancer from 1969 to 2006 for four states and the entire US: Florida (a), Arizona (b), Missouri (c), Indiana (d) and entire US (e).

but with significant variations between and within groups. It is well known that the cancer rates in the US vary widely by geographical area. Figure 1 gives one such illustration based on age-adjusted lung cancer mortality rates for four states: Florida, Arizona, Missouri and Indiana. It is evident that Florida and Arizona share the same change-point and rates of change in each time segment while Missouri and Indiana exhibit different levels of these attributes. Figure 1 also demonstrates that Florida and Arizona have different levels of variability around its mean value over time. When we are interested in grouping states by the rates of change, variability is nuisance with respect to the clustering criteria. However, the omission of such variability from the model may result in inefficient estimation. The presence of heterogeneity among states can also give a misleading impression for the rates of change corresponding to the overall US. Figure 1 panel (e) for the entire nation does not reveal the two distinct types of change-points exhibited in the first four panels by the four different states. This indicates that a more efficient estimation procedure is possible by taking into account local geographical effects.

Third, it is also of interest if we can find some relationship between each specific cancer trend and aggregated cancer trend. Changes in trend over time for all cancer types might be different from changes in trend over time for a specific cancer type. Then, which specific cancer rate affects an aggregated cancer rate can be answered by regression type models that incorporates detecting the changes and grouping states with common changes over time and rate of changes. For example, Figure 2 shows mortality rates of all Malignant cancers and several specific cancer types. Starting from early 1990's, most of cancer mortality rates are decreasing including all cancer type except Liver and Intrahepatic Bile Duct and Pancreas for the state of Florida.

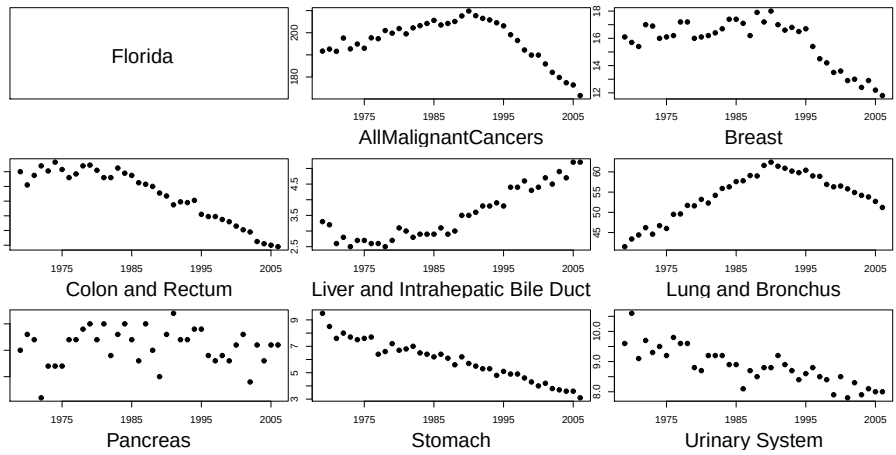

Figure 2: Age-adjusted incidence rates of all cancer and each specific cancer type from 1969 to 2006 for Florida.

To address the above scientific questions, we first developed a *change-point* model to detect changes in multiple time series in presence of heterogeneity. Our proposed model does not assume connectedness at the joinpoint but is able to perform the analysis more locally at this expense. The proposed model is also able to cluster geographical regions which have similar rates. Additional model flexibility for grouping is obtained by including model parameters that represent local data characteristics (for example, variability around the mean trend in Figure 1), which are allowed to vary from site to site. We also incorporate the unknown number of change-points into the estimation scheme, to be inferred from the posterior probabilities.

Our approach is to make use of the Dirichlet Process (DP) methodology in an innovative way to cluster spatio-temporal data. We note that Ghosh et al. (2009) used DP to perform a non-parametric Bayesian analysis of joinpoints where the baseline distribution G_0 is a distribution on R for the errors; this is not for the purpose of clustering but to robustify the analysis with respect to non-normality. Our innovative way of using the DP is to consider realizations from G_0 that are in more general object spaces; in this case, it is the space of all functions over time that represent change points in cancer trends. The advantage of extending DPs in this manner is two-fold: First, the change-points are included as unknown model parameters with a prior distribution governed by G_0. This entails that the uncertainty involved in their estimation is taken into account in the inference, and therefore, represents an improvement over the methodology of Ghosh et al. (2009) and previous approaches by others. Second, even for more general object spaces, the intrinsic property of DPs that assign unit mass to all discrete probability distributions can be utilized to enable clustering of sites with respect to similar cancer trends. For each site on the spatial domain, we have (only one) change-point function. DP-based clustering is obtained for the sites on the spatial domain based on similar change-point functions. Incorporating other aspects of variability via parameters to enhance model flexibility without affecting the DP-based clustering is also an important contribution of this paper.

A primary inferential objective in the analysis of disease data is the summarization and explanation of spatial and spatio-temporal patterns of disease (i.e., disease mapping); see, for example, Elliot et al. (2000), Banerjee et al. (2004) and Lawson (2009) for details and further references. Also of interest is the spatial smoothing, temporal prediction of disease risk and the detection of extremes. Models for inference in this area have been mostly limited by parametric elicitation of dependence structures for pooling spatial information.

On the other hand, the proposed DP-based methodology is free of parametric constraints, and its capability of pooling information via data driven clustering can greatly enhance the analysis of spatial and spatio-temporal patterns. As an illustration, we infer the cluster of US states which correspond to the highest drop in cancer trends in Section 4. We are also able to demonstrate statistical significance of the highest drop compared to other clusters of US states. These types of inference can potentially help policy makers identify factors in the top states that contributed to the highest drop, and subsequently, be implemented as policy or programs in the other states.

The rest of the paper is organized as follows. Section 2 gives the details of the data and application while Section 3 presents the proposed change point model and associated Bayesian inference. Section 4 gives two specific model formulations for the cancer data and demonstrates the superiority of incorporating site specific variability. Section 5 gives a possible extension to incorporate covariate information. Section 6 gives discussion.

2. A Change Point Model for Cancer Incidence Rates

Cancer incidence rates are obtained from the Surveillance, Epidemiology, and End Results (SEER) program (seer.cancer.gov) of the National Institute of Cancer (NCI). The SEER program is an authoritative source of information on cancer incidence and survival in the US. The SEER program currently collects and publishes cancer mortality and survival data from population-based cancer registries covering approximately 26 percent of the population. An age-adjusted incidence/mortality rate is a primary measure for monitoring cancer trends over time and over geographical locations since cancer is a disease where age is a determining factor. An age-adjusted rate is a weighted average of the age-specific (crude) rates, where the weights are the proportions of persons in the corresponding age groups of a standard population. The potential confounding effect of age is reduced when comparing age-adjusted rates computed using the same standard population. Several sets of standard population data are available in SEER which include the 2000 US standard population as well as the standard US populations for the years 1940, 1950, 1960, 1970, 1980, and 1990. The age-adjusted rate using age groups A through B is calculated using the following formula:

$$aarate_{A-B} = \sum_{i=A}^{B} \left[\left(\frac{count_i}{pop_i} \right) \times 100,000 \times \left(\frac{stdmil_i}{\sum_{i=A}^{B} stdmil_i} \right) \right], \qquad (1)$$

where $count_i$, pop_i and $stdmil_i$ are, respectively, the number of incidence/mortality due to a cancer, the population and the choice of a standard population in the age group i. Nineteen age groups and the 2000 US standard population are considered in this study.

We consider lung cancer age-adjusted mortality rates from 1969 to 2006 for the 48 contiguous states in continental United States (excluding Alaska and Hawaii) and Washington D.C. Thus, observations are the age-adjusted lung cancer mortality rates for $t = 1969, 1970, \cdots, 2006$ and $s = 1, 2, \cdots, 49$. Four states and overall USA plots were given in Figure 1 as an example. It is clear from the panels in Figure 1 that there is at least one change-point in the rate of change (i.e., slope) of lung cancer mortality rates for the four states. There are several specific aims of this paper: We would like to determine (i) all possible change-points of slopes of lung cancer mortality rates corresponding to each state, (ii) determine simultaneously if the slopes exhibit some clustering over the states (i.e., different states have identical slope values), and (iii) identify clusters with the highest changes in slope over time, and (iv) whether this highest change is significant compared to the other remaining clusters. To model exponential growth or decay of the age-adjusted

rates, we model the logarithm of the age-adjusted rates as a linear function of time as is done in Ghosh et al. (2011), Clegg et al. (2009) and Ghosh et al. (2009). Slopes over the different time segments capture the essential growth rate (positive or negative) pattern of cancer incidence/mortality rates. The change-point model we develop subsequently is in terms of these slopes and the variability of the observations around the log mean trend.

2.1 Change Point Likelihood Based on Observables

The subsequent discussion applies to both cancer incidence and mortality rates, and therefore, we refer to them as just rates. Let W_{st} denote the logarithm of the observed cancer rate at site s and time t for the collection of sites $s = 1, 2, \cdots, N$ and time points $t = U_0, U_0 + 1, U_0 + 2, \cdots, U_1$. Assume that a site s has k change-points in terms of the slope of the log rates; thus, in Figure 1, s may be Florida with $k = 1$ change points. For fixed k, let $[T_{l-1}, T_l)$, $l = 1, 2, \cdots, k + 1$ be the time intervals where no changes in the disease trend occur (i.e., no change point). To extract the slope and the variability of the observations around the mean trend in each segment, we consider the following regression model on each $[T_{l-1}, T_l)$:

$$W_{st} = \alpha + \beta t + \epsilon_t, \tag{2}$$

for $t = T_{l-1}, T_{l-1} + 1, \cdots, T_l - 1$ with ϵ_t iid $N(0, \sigma^2)$ for the observed data in $[T_{l-1}, T_l)$; thus, in (2), the log rates are modeled as a linear function of time with intercept and slope α and β, respectively, and σ^2 represents the unknown error variance around the mean linear trend. The dependence on s and l is suppressed for the moment. The following results are well known in regression analysis:

$$(\widehat{\alpha}, \widehat{\beta})^T \sim N\left((\alpha, \beta)^T, \sigma^2 (X^T X)^{-1}\right), \quad \text{and} \tag{3}$$

$$\frac{RSS}{\sigma^2} \sim \chi^2_{T_l - T_{l-1} - 2}, \tag{4}$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are the least squares estimators of α and β (which are also the maximum likelihood estimates (MLEs) under the normal error model), RSS is the residual sum of squares given by

$$RSS = \sum_{t=T_{l-1}}^{T_l - 1} (W_{st} - \widehat{\alpha} - \widehat{\beta}\, t)^2, \tag{5}$$

χ^2_ν is the chi-square distribution with ν degrees of freedom, and X is $(T_l - T_{l-1}) \times 2$ matrix whose first and second columns is the vector of ones and $\boldsymbol{t}_l \equiv (T_{l-1}, T_{l-1}+1, \cdots, T_l - 1)^T$, respectively. Also, in (3) and (4), the statistic $(\widehat{\alpha}, \widehat{\beta})$ is independent of RSS. We emphasize here that the number of joinpoints, k, the time intervals $[T_{l-1}, T_l)$, $l = 1, 2, \cdots, k + 1$ and σ^2 are all parameters that are unknown, to be inferred from the subsequent Bayesian analysis. The purpose of introducing these unknown parameters here is to describe the likelihood given the unknown parameters at site s:

$$\widehat{\beta}_l, RSS_l \,|\, \beta_l,\, k, \boldsymbol{t}_l, \sigma_l^2 \overset{ind}{\sim} f_{1l} \times f_{2l} \tag{6}$$

independently for $l = 1, 2, \cdots, k + 1$. In (6), $f_{1l}(\widehat{\beta}_l \,|\, \sigma_l^2)$ is the normal pdf with mean β_l and variance $\sigma_l^2 \cdot v$ where v is the $(2, 2)$-th entry of $(X^T X)^{-1}$; the explicit form of f_{1l} is

$$f_{1l}(\widehat{\beta}_l \,|\, \beta_l, \sigma_l^2) = \frac{1}{\sqrt{2\,\pi\, v\, \sigma_l^2}} \exp\left\{-\frac{1}{2\, v\, \sigma_l^2}(\widehat{\beta}_l - \beta_l)^2\right\}. \tag{7}$$

The density $f_{2l}(RSS_l \,|\, \sigma_l^2)$ in (6) is σ_l^2 times the chi-square density with $m_l = T_l - T_{l-1} - 2$ degrees of freedom whose explicit form is given by

$$f_{2l}(RSS_l \,|\, \sigma_l^2) = \frac{1}{2^{(m_l/2)}\Gamma(m_l/2)} \left(\frac{RSS_l}{\sigma_l^2}\right)^{\frac{m_l}{2}-1} \exp\left\{-\frac{RSS_l}{2\sigma_l^2}\right\} \frac{1}{\sigma_l^2}. \tag{8}$$

Subsequently, we consider the site-wise functions

$$\boldsymbol{\theta}_s(t) = \beta_{sl} \quad \text{if } T_{l-1} \le t \le T_l - 1 \tag{9}$$

where β_{sl} is the true but unknown slope in the interval $[T_{l-1}, T_l)$ at site s. Thus, the functions $\boldsymbol{\theta}_s(t)$ are step-functions of t with k change points at times T_l, $l = 1, 2, \cdots, k$. Denote the set of all observables by $\boldsymbol{Y} = \{\, \boldsymbol{Y}_{sl},\, l = 1, 2, \cdots, k_s,\, s = 1, 2, \cdots, N$ where $\boldsymbol{Y}_{sl} \equiv (\widehat{\beta}_{sl}, RSS_{sl})$ with $\widehat{\beta}_{sl}$ and RSS_{sl} as in (6) for each site s, and k_s is the number of joinpoints for site s. Let $\boldsymbol{\beta}$ denote the collection of all true slope parameters β_{sl}, $l = 1, 2, \cdots, k_s$, $s = 1, 2, \cdots, N\,\}$. Also, denote by $\boldsymbol{K}$, $\boldsymbol{T}$ and $\boldsymbol{\sigma}$ to be the collection of parameters k, $T_l, l = 1, 2, \cdots, k+1$ and σ_l^2 for all the N sites. Assuming independence between the N sites, the likelihood is given by

$$\boldsymbol{f}(\boldsymbol{Y} \,|\, \boldsymbol{\beta},\, \boldsymbol{K},\, \boldsymbol{T},\, \boldsymbol{\sigma}\,) = \prod_{s=1}^{N} \prod_{l=1}^{k_s} f_{1l}^{(s)} \times f_{2l}^{(s)} \tag{10}$$

where $f_{1l}^{(s)}$ and $f_{2l}^{(s)}$ are f_{1l} and f_{2l} corresponding to site s.

As mentioned in the Introduction, the change point analysis here is different from joinpoint modeling. The latter assumes that the cancer incidence rates are continuous at the joinpoints but with different slopes to the left and right of the joinpoint. In our case, we make no assumption on the continuity of the regression at the time points T_l. However, at this expense, the current formulation allows us to infer different slopes for the different sites, and therefore, enable clustering of these slopes based on the DP-methodology. The proposed model also allows the unknown number of change-points and clusters to be inferred concurrently with parameters based on Bayesian posterior probabilities (details in the subsequent sections).

3. Bayesian Inference Using Functional DP Priors

3.1 Functional DP Prior

Let $\boldsymbol{\Theta}$ denote the set of all step functions $\boldsymbol{\theta}_s$ as described in the previous section. We introduce the functional DP as a prior on space of all distributions on $\boldsymbol{\Theta}$. The $DP \equiv DP(\alpha_0 G_0)$ depends on two hyper-parameters, namely, $\alpha_0 > 0$ the precision parameter, and G_0 the baseline (or centering) distribution on $\boldsymbol{\Theta}$. Recall that a randomly generated distribution F from $DP(\alpha_0 G_0)$ is almost surely discrete and admits the representation

$$F = \sum_{i=1}^{\infty} \omega_i\, \delta_{\theta_i}, \tag{11}$$

where δ_z denotes a point mass at z, $\omega_1 = \eta_1$, $\omega_i = \eta_i \prod_{k=1}^{i-1}(1 - \eta_k)$, for $i = 2, 3, \cdots$ with $\theta_1, \theta_2 \cdots$ iid from G_0 (Sethuraman, 1994). Traditionally, θ_i was assumed to be scalar or vector-valued taking values in R^p. To model the observational process via change-points, we conceptually extend θ_is in (11) to functions $\boldsymbol{\theta}_i \equiv \{\, \theta_i(t) \,:\, t = U_0, U_0 + 1, \cdots, U_1\}$. The notations $\boldsymbol{\theta}$, $\boldsymbol{\theta}(t)$ and θ, therefore, will be taken to denote, respectively, a function, the

value of $\boldsymbol{\theta}$ evaluated at time t and a possible realization taken by $\boldsymbol{\theta}(t)$. These notations will be used throughout the paper subsequently. For an integer $k \geq 0$, $\boldsymbol{\theta}$ with k change-points has the form

$$\boldsymbol{\theta}(t) = \theta_l \qquad \text{if} \qquad T_{l-1} \leq t < T_l, \tag{12}$$

for $l = 1, 2, \cdots, k+1$ with $U_0 \equiv T_0 < T_1 < \cdots T_k < T_{k+1} \equiv U_1$ as seen earlier. The notation $F \sim DP(\alpha_0 G_0)$ in this context will be taken to mean

$$F = \sum_{i=1}^{\infty} \omega_i \, \delta_{\boldsymbol{\theta}_i}, \tag{13}$$

where $\delta_{\boldsymbol{z}}$ is now a point mass on the step function $\boldsymbol{z}$, ω_is are as before, and the $\boldsymbol{\theta}_i$s are iid from a distribution G_0 on $\boldsymbol{\Theta}$. To specify G_0, the baseline distribution on $\boldsymbol{\Theta}$, it is convenient to utilize a hierarchical structure: (1) Let $K \sim Poisson(\lambda)$. (2) Fix an integer $w > 0$. Given $K = k$, let

$$(n_1, \cdots, n_{k+1}) \sim \text{Multinomial}\left(n_0, \frac{1}{k+1}, \cdots, \frac{1}{k+1}\right),$$

where $n_0 = U_1 - U_0 - (k+1)w = n - 1 - (k+1)w$. (3) Define T_l recursively as $T_0 = U_0$, $T_l = n_l + T_{l-1} + w$ for $l = 1, 2, \cdots, k+1$. Given $T_1, \cdots, T_k$, generate $\theta_1, \cdots, \theta_{k+1}$ iid from the (univariate or multivariate) density π_0 on R^d, and set

$$\boldsymbol{\theta}(t) = \theta_l \text{ if } T_{l-1} \leq t < T_l, \tag{14}$$

for $l = 1, \cdots, k+1$. Note that $T_k \leq t \leq T_{k+1}$ for $l = k+1$, K is the number of change-points, T_ls for $l = 1, \cdots, K$ are the time points when a change is made and n_l is the number of time points in the interval $[T_{l-1}, T_l)$ for $l = 1, \cdots, K+1$. Note that again, for $l = K+1$, the interval becomes $[T_K, T_{K+1}]$. By introducing $w > 0$, we avoid zero-length interval since each time interval $[T_l, T_{l+1})$ is at least w units. From the hierarchical specification above, it follows that the infinitesimal measure is given as

$$G_0(d\boldsymbol{\theta}_s) = \left(\frac{e^{-\lambda}\lambda^k}{k!}\right)\left(\frac{\Gamma(n_0+1)}{\prod_{i=1}^{k}\Gamma(n_i+1)}\left(\frac{1}{k+1}\right)^{n_0}\right)\prod_{l=1}^{k+1}\pi_0(\theta_l)\,d\theta_l. \tag{15}$$

3.2 Incorporating Site-specific Variability

The prior development thus far has been on the change point functions $\boldsymbol{\theta}_s$. The variance parameters σ_{sl}^2 represent the extent of variability of the log rates around the mean trend. Note from Figures 1 (a) and (b) that although Florida and Arizona have the same cancer trends, the variability around this common mean trend is different for the two states. This necessitates the incorporation of σ_{sl}^2 as site specific parameters independent of the clustering. In fact, we demonstrate in Section 4, the exclusion of such consideration (that is, allowing σ to be common for all the sites in a cluster but different for the different time segments) results in poor clustering of cancer trends. Thus, for additional flexibility, the likelihood component of $\boldsymbol{Y}_{sl}$ incorporates a site-specific variability parameter $\xi_s = \sigma_{sl}^2$ for all $l = 1, 2, \cdots, k$ (that is, one common site-wise variance parameter), for each $s = 1, 2, \cdots, N$. For the subsequent Bayesian analysis, the parameters $\xi_s \in \Xi$, where Ξ is its parameter space, are assumed to be iid from the pdf π_1. Note that ξ_s can be different for each s, and therefore, are not subject to site-based clustering as the change-point functions $\boldsymbol{\theta}_s$. The infinitesimal measure in (15) is now extended to include the site-wise parameters ξ_s and is given by

$$\tilde{G}_0(d\boldsymbol{\theta}_s, d\xi_s) = \left(\frac{e^{-\lambda}\lambda^k}{k!}\right)\left(\frac{\Gamma(n_0+1)}{\prod_{i=1}^{k}\Gamma(n_i+1)}\left(\frac{1}{k+1}\right)^{n_0}\right)\left(\prod_{l=1}^{k+1}\pi_0(\theta_l)\,d\theta_l\right)\pi_1(\xi_s)\,d\xi_s. \tag{16}$$

In what follows, it will be useful to make the following definition: For fixed $\boldsymbol{\theta}_s$, the infinitesimal measure

$$\delta(\boldsymbol{\theta}_s, d\boldsymbol{\xi}_s) = \delta_{\boldsymbol{\theta}_s} \times \pi_1(\xi_s)\, d\xi_s \tag{17}$$

is the product of the point mass measure on $\boldsymbol{\theta}_s$ and the infinitesimal measure $\pi_1(\xi_s)\, d\xi_s$.

Based on the likelihood in (10), the complete hierarchical model specification can now be stated as follows:

$$\begin{aligned} \boldsymbol{Y} \,|\, \boldsymbol{\beta},\, \boldsymbol{K},\, \boldsymbol{T},\, \boldsymbol{\sigma} &\sim \boldsymbol{f} && (18)\\ \boldsymbol{\theta}_s &\overset{iid}{\sim} F, \quad \text{and} && (19)\\ F &\sim DP(\alpha_0\, G_0). && (20) \end{aligned}$$

Note that the set $(\boldsymbol{\beta},\, \boldsymbol{K},\, \boldsymbol{T},\, \boldsymbol{\sigma})$ is in one-to-one correspondence with $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N, \boldsymbol{\xi})$ where $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_N)$.

3.3 Bayesian Inference Methodology

To infer $\boldsymbol{\theta}$s, the standard practice in DP posterior analysis is to integrate out F from the hierarchical specification of (18)-(20) (see, for example, Dey et al. (1998)). The likelihood corresponding to the observables $\boldsymbol{Y}$ in (18) is given by $\ell(\boldsymbol{Y} \,|\, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N,\, \boldsymbol{\xi}) = \prod_{s=1}^{N} \prod_{l=1}^{k+1} f(\, \boldsymbol{Y}_{s,l} \,|\, \boldsymbol{\theta}_s,\, \xi_s)$ where the subscript s on k is suppressed.

The conditional posterior distribution of the pair $(\boldsymbol{\theta}_s,\, \xi_s)$ given the other pairs $(\boldsymbol{\theta}_{-s},\, \boldsymbol{\xi}_{-s}\,)$ can be derived as

$$\begin{aligned} (\boldsymbol{\theta}_s,\, \xi_s \,|\, \boldsymbol{\theta}_{-s},\, \boldsymbol{\xi}_{-s}) &\propto \prod_{l=1}^{k+1} f(\boldsymbol{Y}_{s,l} \,|\, \theta_l,\, \xi_s) \left[\alpha_0 \frac{\tilde{G}_0(d\boldsymbol{\theta}_s,\, d\xi_s)}{\alpha_0 + N - 1} + \frac{1}{\alpha_0 + N - 1} \sum_{s' \neq s} \delta(\boldsymbol{\theta}_{s'},\, d\boldsymbol{\xi}_{s'}) \right], \\ &= \frac{q_{s,0} \tilde{G}_0^*(d\boldsymbol{\theta}_s,\, d\boldsymbol{\xi}_s) + \sum_{s' \neq s} q_{s,s'} \delta(\boldsymbol{\theta}_{s'},\, d\xi_{s'})}{q_{s,0} + \sum_{s' \neq s} q_{s,s'}}, \end{aligned} \tag{21}$$

where the second line is obtained from the first after normalization. The quantities $q_{s,0}$ and $q_{s,s'}$ in (21) have the expressions

$$q_{s,0} = \alpha_0 \int_{\mathcal{S}} \int_{\Xi} \prod_{l=1}^{k+1} f(\boldsymbol{Y}_{s,l} \,|\, \theta_l,\, \xi_s)\, \tilde{G}_0(d\boldsymbol{\theta}_s,\, d\boldsymbol{\xi}_s), \quad \text{and} \tag{22}$$

$$q_{s,s'} = \int_{\Xi} \prod_{l=1}^{k^*+1} f(\boldsymbol{Y}_{s,l} \,|\, \theta_l,\, \xi_s)\, \delta(\boldsymbol{\theta}_{s'}, d\boldsymbol{\xi}_{s'}), \tag{23}$$

where k^* is the number of change-points in $\boldsymbol{\theta}_{s'}$. The distribution

$$\tilde{G}_0^*(d\boldsymbol{\theta}_s, d\xi_s) = \frac{\alpha_0 \prod_{l=1}^{k+1} f(\boldsymbol{Y}_{s,l} \,|\, \theta_l,\, \xi_s)\, \tilde{G}_0(d\boldsymbol{\theta}_s, d\boldsymbol{\xi}_s)}{q_{s,0}}$$

is that of $(\boldsymbol{\theta}_s, \xi_s)$ when a new realization of $(\boldsymbol{\theta}_s, \xi_s)$ (i.e., not belonging to any of the previous clusters) has to be generated. An alternative way of writing (21) in terms of the distinct clusters is

$$(\, \boldsymbol{\theta}_s,\, \xi_s \,|\, \boldsymbol{\theta}_{-s},\, \boldsymbol{\xi}_{-s}\,) = \frac{q_{s,0} \tilde{G}_0^*(d\boldsymbol{\theta}_s, d\xi_s) + \sum_{j=1}^{N^*} N_j\, q_{s,j}\, \delta_{\boldsymbol{\theta}_j}}{q_{s,0} + \sum_{j=1}^{N^*} N_j\, q_{s,j}}, \tag{24}$$

where $\boldsymbol{\theta}_j$, $j = 1, 2, \cdots, N^*$ are the distinct change-point functions for the N^* different clusters, N_j is the number of sites s' for which $\boldsymbol{\theta}_{s'}$ is equal to $\boldsymbol{\theta}_j$, and $q_{s,j}$ is $q_{s,s'}$ in (23) with $\boldsymbol{\theta}_{s'}(t)$ replaced by $\boldsymbol{\theta}_j(t)$. Note that $\sum_{j=1}^{N^*} N_j = N - 1$ since the site s is left out.

Expression (24) explicitly demonstrates the clustering capability of DP. The current value of $\boldsymbol{\theta}_s$ can be selected to be one of the other $\boldsymbol{\theta}_{s'}$ with probability $\sum_{j=1}^{N*} N_j\, q_{s,j}/(q_{s,0} + \sum_{j=1}^{N*} N_j\, q_{s,j})$, this positive probability being the reason for possible clustering of sites in terms of $\boldsymbol{\theta}_s$. Expression (24) also allows for a new $\boldsymbol{\theta}_s^*$ to be generated from the posterior distribution G_0^*; this is the likely scenario if the temporal observations at site s, $\boldsymbol{W}_s = \{\, W_{st},\, t = U_0, U_0 + 1, \cdots, U_1 \,\}$, strongly support a different change-points function compared to the existing $\boldsymbol{\theta}_{s'}$ functions for $s' \neq s$. We note that the above treatment is similar to Gelfand et al. (2005) who extended θ_l to a realization of a random field by replacing it with a surface function on a spatial domain. However, Gelfand et al. (2005) do not consider joinpoint extensions as is done here; see also the related discussion in the Introduction.

The DP prior introduces two other hyper-parameters, namely α_0 and λ, into the inferential framework. In our analysis α_0 is fixed at a known value. We take the prior on λ to be π_2. The priors π_0, π_1 and π_2 are taken to be

$$\pi_0(\theta_l) \propto 1, \quad \pi_1(\sigma^2) = \texttt{igamma}(a_1, b_1) \quad \text{and} \quad \pi_2(\lambda) = \texttt{gamma}(a_2, b_2), \tag{25}$$

where `gamma` and `igamma` are the Gamma and inverse Gamma distributions with shape and scale parameters (a_1, b_1) and (a_2, b_2), respectively. The above choices are conjugate to their respective likelihoods enabling the posteriors to be obtained in closed forms. The reader is referred to the Discussion section of this paper for the motivation of using a flat prior for θ_l from the conjugacy perspective. It turns out that using a common normal prior for θ_l does not allow the integrals in $q_{s,0}$ to be computed in closed form.

For a complete update of all the unknown parameters, please refer to Dass et al. (2010).

3.4 Inference based on Posterior Samples

After convergence is established, we take B samples from the posterior distribution to make inference on all unknown quantities. Let $\mathcal{X}_b^*$, $b = 1, 2, \cdots, B$ be B samples of the posterior obtained from the Gibbs sampler. Components of $\mathcal{X}_b^*$ include N realizations of step functions θ_s and $\boldsymbol{\xi}$ (or equivalently, $\boldsymbol{\beta}, \boldsymbol{K}, \boldsymbol{T}, \boldsymbol{\sigma}$). Thus, marginal posterior inference can be carried out for each of these components. For example, to infer $\boldsymbol{\theta}_s(t)$ for a particular site s and time point t, we extract all $\boldsymbol{\theta}_s(t)$ components from each $\mathcal{X}_b^*$, $b = 1, 2, \cdots, B$. The B realizations of $\boldsymbol{\theta}_s(t)$ are then used to compute the posterior mean, variance and confidence interval. A similar procedure also works for N^* where we can obtain marginal probabilities of $N^* = n^*$ for all non-negative integers n^*. Results for simulation experiments and real data are given in the subsequent sections.

A more challenging inference problem is to obtain results for the clustering tendencies, for example, the "average" clusters. Note that the output of the Gibbs sampler at each iteration is a clustering of the N states, and therefore, it is difficult to obtain a summary posterior measure, such as mean and variance, for the clustering of sites. To get some idea about average clustering tendencies reflected by the posterior distribution, the following methodology is developed: For every pair of sites (s_1, s_2) in $\{1, 2, \cdots, N\}$, define $D_b(s_1, s_2) = 1$ if s_1 and s_2 belong to the same cluster in $\mathcal{X}_b^*$, and 0, otherwise, for $b = 1, 2, \cdots, B$. Subsequently, we construct the average distance measure between the sites s_1 and s_2 using

$$dist(s_1, s_2) = 1 - \bar{D}(s_1, s_2)$$

Number of clusters	4	5	6	7	8
Posterior Prob	0.0688	0.5153	0.3820	0.0335	0.0003

Table 1:
Posterior probabilities of number of clusters for Model 1

where $\bar{D}(s_1, s_2) = \sum_{b=1}^{B} D_b(s_1, s_2)/B$. Based on $dist$, an agglomerative clustering algorithm is performed with the maximum number of clusters threshold in the algorithm fixed at the value of N^* for which the posterior probability has the maximum value. The clustering outputs from this procedure match with our expected scenario. Subsequent sections give results based on real and validation data.

4. Analysis of Cancer Incidence Rates Revisited

We consider two specific choices of models. The site-specific variability model is given by Model 1 below. In Model 2, we assume that σ^2 is cluster-dependent (not site-specific), that is, σ^2 is same over all states in the same cluster (but different for the different clusters). Based on previous discussion, we can write these two models as follows:
Model 1: $\boldsymbol{Y}_{sl} = (\widehat{\beta}_l, RSS_l)^T$, $\theta_l = \beta_l$, $\xi_s = \sigma^2$
Model 2: $\boldsymbol{Y}_{sl} = (\widehat{\beta}_l, RSS_l)^T$, $\theta_l = (\beta_l, \sigma_l^2)$,
suppressing the subscript s on $\widehat{\beta}_l$ and RSS_l.

Note that Model 2 is not a subset of Model 1 or vice versa. In Model 2, σ_l^2 is common to all sites within a cluster but can vary for the different time intervals $[T_{l-1}, T_l)$. In Model 1, one common σ_s^2 is assumed for each site which does not change within each time segment.

The Appendix of Dass et al. (2010) gives the model specific expressions used for the Bayesian inference. We run three Gibbs chains for 10,000 iterations. The convergence is established after 5,000 iterations and we take 2,000 samples from each chain after convergence so that total 6,000 samples are used for further posterior analysis. Specific values of hyper-parameters are set to $a_1 = b_1 = 1$ for π_1, $a_2 = b_2 = 1$ for π_2. α_0 is set to $1/100$. The number of clusters of states based on the highest posterior probability is found to be $N^* = 5$; see Table 1 for posterior probabilities. Using the posterior estimate of N^*, we use the clustering methodology explained in section 3.4 to cluster states into 5 groups.

As mentioned in Introduction, we expect Florida and Arizona to belong to the same cluster while Indiana and Missouri to belong in another. This is what is revealed from the analysis. Marginal posterior analysis on the number of change-points for each state revealed that one change-point corresponds to the highest probability. Further, the posterior probabilities of the time intervals corresponding to no change-point and a single change-point are given in Table 2 for each of the four states. The entries in Table 2 is the marginal posterior probabilities corresponding to the most significant partitions of the interval $[1969, 2006]$ based on output of the Gibbs sampler. Note that both Arizona and Florida showed one change-point, $T_1 = 1989$ while for Missouri and Indiana, the change-point was $T_1 = 1987$. Corresponding to these change-points, the mean posterior estimates of σ_s (site-wise) and β_l (cluster-wise) is given in Table 3.

Next, we demonstrate the superiority of Model 1 over Model 2 based on predictive analysis. A new realization of W_{st}, W^*_{st}, is obtained by sampling from the normal distribution with mean $\alpha_l + \beta_l\, t$ and variance σ_s^2 where β_l and σ_s^2 are posterior realizations from the Gibbs chain and α_l is given from the data for the corresponding time interval and site. The B values of W^*_{st} are then used to construct the 95% credible predictive interval. The confidence bands generated are shown in Figure 3 in the original scale. A similar procedure is repeated for Model 2 to obtain the confidence bands shown in Figure 4. The

Change-Points	Arizona	Florida	Indiana	Missouri
No Change-Points	0.0293	0.0042	0	0
$T_1 = 1994$	0	0	0	0.0010
$T_1 = 1993$	0.0010	0.0002	0.0002	0.0002
$T_1 = 1992$	0.0147	0.0355	0	0.0005
$T_1 = 1991$	0.0657	0.1723	0.0018	0.0028
$T_1 = 1990$	0.0768	0.1533	0.0147	0.0172
$T_1 = 1989$	0.5807	0.6033	0.0188	0.0200
$T_1 = 1988$	0.0662	0.0258	0.0778	0.0975
$T_1 = 1987$	0.0432	0.0007	0.7205	0.7212
$T_1 = 1986$	0.0042	0	0.1345	0.1285
$T_1 = 1985$	0.0010	0	0.0030	0.0032
Two Change-Points	0.1033	0	0.0113	0.0025
Three Change-Points	0.0118	0.0045	0.0172	0.0055

Table 2:
Posterior probabilities of a change-point for Model 1

State	σ_s	Change-Point(s)	(β_1, β_2)
Arizona	0.0416	$T_1 = 1989$	$(0.0209, -0.0107)$
Florida	0.0182	$T_1 = 1989$	$(0.0196, -0.0118)$
Indiana	0.0297	$T_1 = 1987$	$(0.0297, 0.00008)$
Missouri	0.0292	$T_1 = 1987$	$(0.0295, 0.00006)$

Table 3:
Posterior outputs for Model 1

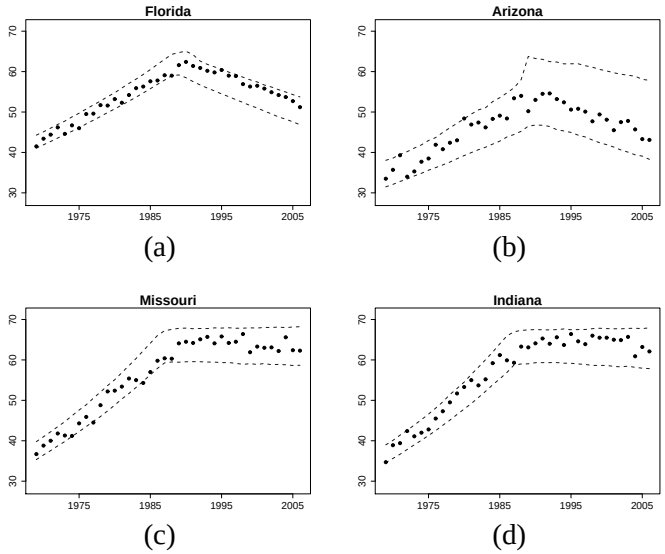

Figure 3: Examples of states belonging to different clusters from the implementation of the change-point methodology. The bands around the observed values (age-adjusted cancer rates) are the 95% predictive credible intervals based on Model 1.

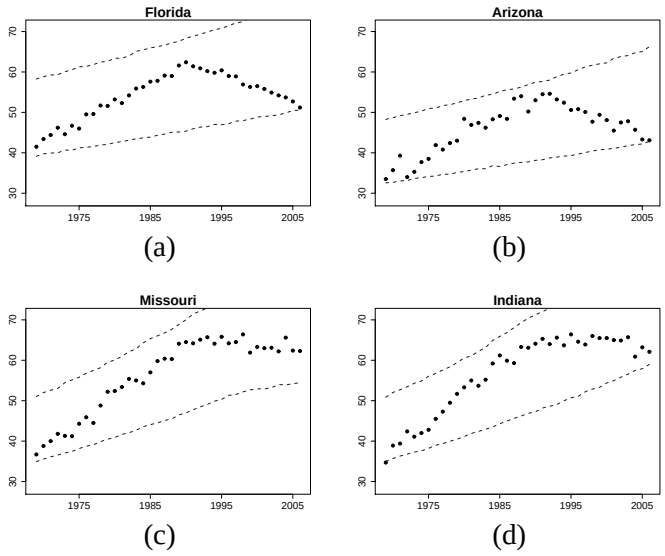

Figure 4: The bands around the observed values (age-adjusted cancer rates) are the 95% predictive credible intervals based on Model 2.

Cluster	Posterior Mean of $\beta_2 - \beta_1$	95% Credible Interval
1	-0.0370	$(-0.0397, -0.0336)$
2	-0.0318	$(-0.0341, -0.0298)$
3	-0.0316	$(-0.0385, -0.0294)$
4	-0.0314	$(-0.0337, -0.0298)$
5	-0.0307	$(-0.0329, -0.0291)$

Table 4:
Clusters (from the agglomerative procedure) with the highest drop in cancer incidence rates measured in terms of $\beta_2 - \beta_1$.

better model will be the one that detects at least one change-point and that gives narrower confidence bands. Note that change-points are not detected and the width of the predictive confidence bands are too large for Model 2. These results indicate that there is significant evidence from the data to suggest heterogeneous (i.e., site-specific) variability around the mean within clusters.

Next, the cluster with the highest drop in cancer incidence rate is identified. The difference $\beta_2 - \beta_1$ in Table 4 is computed using posterior samples for each of the 5 clusters based on Model 1. Table 4 also gives the corresponding 95% credible intervals of the 5 clusters for $\beta_2 - \beta_1$. Note that the top cluster has a drop in rates that is significantly different from clusters 2, 4 and 5. States in this cluster consists of Colorado, Georgia, Oregon and Virginia. One subsequent investigation may, therefore, be to identify the underlying reasons for the highest drop in cancer rates, and to identity and implement effective policies or programs in these states to the other states in the nation.

5. Regression model with Change points and Clustering

In this section, we consider a possible extension to include covariates which also have change points over time. Suppose that we have $n = U_1 - U_0 + 1$ time points given by the index: $t = U_0, U_0 + 1, U_0 + 2, \cdots, U_1$ and p variables on N spatial sites given by $\boldsymbol{x}_s^{(j)}(t)$,

the value of the j-th variable at site s and time point t for $j = 1, \cdots, p$. $\boldsymbol{y}_s(t)$ is a response variable at $s = 1, \cdots, N$ and $t = U_0, U_0+1, U_0+2, \cdots, U_1$. With cancer data, $\boldsymbol{y}_s(t)$ is all malignant cancer mortality rate and $\boldsymbol{x}_s^{(j)}(t)$ is a mortality rate of a specific cancer type.

We first consider a DP change-point model on each $\boldsymbol{x}_s^{(j)}(t)$ independently for each j which is specified in Section 3. The likelihood of $\boldsymbol{x}_s(t)$ is assumed to be

$$(2\pi)^{-n/2}\sigma_s^{-n}\exp\left\{-\frac{1}{2\sigma_s^2}\sum_{l=1}^{k+1}\sum_{t=T_{l-1}}^{T_l-1}\left(\boldsymbol{x}_s(t)-\delta_{l,s}-\theta_{l,s}(t-T_{l-1,s}+1)\right)^2\right\},$$

where $\delta_{l,s}$ is an intercept for the l-th interval which is nuisance to the problem of rates of change. $\theta_{l,s}$ and $T_{l,s}$ were used instead of θ_l and T_l to emphasize dependence on site s. Our approach is to assume that the rate of change of $\boldsymbol{y}_s$ at every time point t is governed by a linear combination of the rates of change of $\boldsymbol{x}_s^{(j)}(t)$. That is, we have

$$\boldsymbol{\mu}_s(t) = \alpha_0 + \sum_{j=1}^{p}\alpha_j\boldsymbol{\theta}_s^{(j)}(t), \tag{26}$$

where $\boldsymbol{\mu}_s(t)$ is the rate of change of $\boldsymbol{y}_s$ at time t. Note that we assume that $(\alpha_0, \cdots, \alpha_p)$ is same over s. Then, we have a regression type model for the rate of changes. Selecting j with significant $\widehat{\alpha}_j$ will enable us to find a subset of covariates, $\boldsymbol{x}_s^{(j)}(t)$, which contributes to the rate of changes of $\boldsymbol{y}_s$.

We further assume the following likelihood for $\boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N)$ to obtain the overall likelihood:

$$\begin{aligned}
&\prod_{s=1}^{N}(2\pi)^{-n/2}\tau_s^{-n}\exp\left\{-\frac{1}{2\tau_s^2}\sum_{t=U_0}^{U_1}\left(\boldsymbol{y}_s(t)-\mu_{0,s}-\int_{U_0}^{t}\boldsymbol{\mu}_s(u)du\right)^2\right\},\\
&=\prod_{s=1}^{N}(2\pi)^{-n/2}\tau_s^{-n}\exp\left\{-\frac{1}{2\tau_s^2}\sum_{t=U_0}^{U_1}\left(\boldsymbol{y}_s(t)-\mu_{0,s}-\alpha_0(t-U_0)\right.\right.\\
&\qquad\left.\left.-\sum_{j=1}^{p}\alpha_j\int_{U_0}^{t}\boldsymbol{\theta}_s^{(j)}(u)du\right)^2\right\},
\end{aligned}$$

where τ_s^2 is a site-specific variability for $\boldsymbol{y}_s$ and $\mu_{0,s}$ is a site-specific intercept term, which is nuisance to the problem. Let $L_s^{(j)}$ be the label of the s-th site for j-th $\boldsymbol{x}$ variable, i.e. $\boldsymbol{x}_s^{(j)}(\cdot)$ such that

$$L_s^{(j)} \sim \sum_{u=1}^{\infty} w_u^{(j)}\boldsymbol{I}(\boldsymbol{\theta}_u^{(j)}(\cdot))$$

independently for all j, where $\boldsymbol{\theta}_u^{(j)}(\cdot) \overset{i.i.d.}{\sim} G_0$ and $w_u^{(j)}$ is a transformed weight from the Beta distribution. That is, $w_1 = \eta_1$, $w_u = \eta_u\prod_{i=1}^{u-1}(1-\eta_i)$ and η_i are i.i.d. from $\mathrm{Beta}(1,\nu)$. Then, the overall likelihood is

$$\begin{aligned}
&\prod_{s=1}^{N}(2\pi)^{-n/2}\tau_s^{-n}\exp\left\{-\frac{1}{2\tau_s^2}\sum_{t=U_0}^{U_1}\left(\boldsymbol{y}_s(t)-\mu_{0,s}-\alpha_0(t-U_0)\right.\right.\\
&\qquad\left.\left.-\sum_{j=1}^{p}\alpha_j\int_{U_0}^{t}\boldsymbol{\theta}_{L_s^{(j)}}^{(j)}(u)du\right)^2\right\}
\end{aligned}$$

$$\times \prod_{j=1}^{p}\prod_{s=1}^{N}(2\pi)^{-n/2}\sigma_{s,j}^{-n}\exp\left\{-\frac{1}{2\sigma_{s,j}^2}\sum_{l=1}^{k_{L_s^{(j)}+1}}\sum_{t=T_{l-1,L_s^{(j)}}}^{T_{l,L_s^{(j)}}-1}\left(\boldsymbol{x}_s^{(j)}(t)-\delta_{l,s}^{(j)}\right.\right.$$

$$\left.\left.-\theta_{l,L_s^{(j)}}^{(j)}(t-T_{l-1,L_s^{(j)}}+1)\right)^2\right\}w_{L_s^{(j)}}^{(j)}$$

$$\times\prod_{j=1}^{p}\prod_{u=1}^{\infty}\pi(w_u^{(j)})\prod_{u=1}^{\infty}dG_0(\boldsymbol{\theta}_u^{(j)}(\cdot))\times\pi_0(\alpha_0,\cdots,\alpha_p)\times\prod_{s=1}^{N}\pi(\mu_{0,s}).$$

With these distributional assumptions on $\boldsymbol{x}_s^{(j)}$ and $\boldsymbol{y}_s$, we can derive conditional distribution of all parameters so that we can develop a Bayesian inferential procedures. Posterior analysis will give us change points and clustering of states for each covariate as well as a subset of covariates which contribute a rate of change of $\boldsymbol{y}_s$.

6. Discussion

In this paper, we propose change-point models for spatio-temporal data that can detect change-points over time and group spatial sites into several clusters with respect to their change-point functions. Clustering is achieved by using a Dirichlet process prior on the space of step functions over time. The model was developed to analyze state-wise age adjusted rates to find local change-points and clusters that have similar changes.

Our analysis based on predictive distribution demonstrate that Model 1 is superior to Model 2. Thus, model flexibility is achieved far more by incorporating site specific parameters which are nuisance to the clustering compared to adding extra parameters for clustering. The latter action may in fact distort true underlying trends as evidenced by Figure 4.

For the real application, we find that state-level and national level age-adjusted lung cancer mortality rates show a clear change-point around late 1980s to early 1990s. Some states like Florida and Arizona follow similar patterns as national level rates while some states like Missouri and Indiana show different patterns from the national level rates (see Figure 1). In particular, Missouri and Indiana have smaller rate of changes after the change-point compared to Florida and Arizona as well as national level (see Table 3). Indeed, we can argue that lung cancer mortality rates have not changed much after 1990s for these states, while the national level seems to significantly decrease. This further indicates that we need different attention on each individual state. Another avenue for this line of research is to incorporate covariate information into the clustering mechanism as we briefly introduced in Section 5 and authors are currently working on refining models and developing an algorithm.

REFERENCES

American Cancer Society (2010) "Cancer Facts & Figures 2010", American Cancer Society, Atlanta, GA.

Banerjee S., Gelfand, A. E., and Carlin, B. P. (2004), Hierarchical Modeling and Analysis for Spatial Data, Chapman & Hall/CRC.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992), "Hierarchical Bayesian Analysis of Changepoint Problems," Applied Statistics, 41, 389-405.

Clegg, L. X., Hankey, B. F. , Tiwari, R., Feuer, E. J. and Edwards, B. K. (2009), "Estimating average annual per cent change in trend analyis", Statist. Med., 28, pp. 3670–3682.

Dass, S. C., Lim, C. and Maiti, T. (2011), "Change Point Analysis of Cancer Mortality Rates for US States using Functional Dirichlet Processes ", Technical Report RM 690, Department of Statistics and Probability, Michigan State University.

Elliott,P., Wakefield, J., Best, N., and Briggs, D. (Eds.) (2000), "Spatial epidemiology: methods and applications," Publisher: Oxford University Press, Oxford.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005), "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing," J. Amer. Statist. Assoc, 100, 1021-1035.

Ghosh, P., Basu, S. and Tiwari, R. C. (2009), "Bayesian Analsysis of Cancer Rates from SEER Program using Parametric and Semiparametric Jointpoint Regression Models," J. Amer. Statist. Assoc., 104, 439-452.

Ghosh, P. and Kaushik, G. and Tiwari, R. (2011), "Bayesian approach to cancer-trend analysis using age-stratified Poisson regression models", Statist. Med., 30, pp. 127–139.

Kim, H. J., Fay, M. P., Feuer, E. J., and Midthune, D. N. (2000), "Permutation Tests for Joinpoint Regression with Applications to Cancer Rates," Statistics in Medicine, 19, 335-351.

Kim, H. J., Fay, M. P., Yu, B., Barrett, M. J., and Feuer, E. J. (2004), "Comparability of Segmented Line Regression Models," Biometrics, 60, 1005-1014.

Lawson, A. B. (2009) Bayesian disease mapping: hierarchical modeling in spatial epidemiology, Boca Raton:CRC Press

Ries, L. A. G., Eisner, M. P., Kosary, C. L., Hanley, B. F., Miller, B. A., Clegg, L., and Edwards, B. K. (2002), SEER Cancer Statistics Review, National Cancer Institute, Bethesda, MD, available at `http://seer.cancer.gov/csr/1973-1999`.

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," Statistica Sinica, 4, 639-650.

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database, "Mortality - All COD, Aggregated With State, Total U.S. (1969-2007) <Katrina/Rita Population Adjustment>", National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released June 2010. Underlying mortality data provided by NCHS (www.cdc.gov/nchs).

Tiwari, R. C., Cronin, K. A., Davis, W., Feuer, E. J., Yu, B., and Chib, S. (2005), "Bayesian Model Selection for Joinpoint Regression with Application to Age-adjusted Cancer Rates," J. R. Statist. Soc. Ser. C, 54, 919-939.