ENVIRONMETRICS
Environmetrics 2007; **18**: 727–744
Published online 2 September 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.876

Spline smoothing in Bayesian disease mapping

Ying C. MacNab*,†

Division of Epidemiology and Biostatistics, Department of Health Care and Epidemiology, University of British Columbia, and Centre for Healthcare Innovation and Improvement, British Columbia Child and Family Research Institute, Vancouver, Canada

SUMMARY

Penalty splines such as smoothing spline and P-spline, as well as unpenalized regression splines, have become increasingly popular methods in contemporary non-parametric and semiparametric regressions, particularly for data arising from longitudinal, multilevel, and spatiotemporal settings. In the recent decade, the development of the Markov chain Monte Carlo (MCMC) methods has facilitated applications of flexible spline fittings via Bayesian hierarchical formulation. In this paper, we study three spline smoothing methods in the context of spatiotemporal modeling of rates and Bayesian disease mapping. We explore their potentials for fully Bayesian (FB) rate and risk trends ensemble estimation and spatiotemporal relative risks inference. In particular, our study presents and compares Bayesian hierarchical formulations of regression B-spline, smoothing spline, and P-spline, explores the connections and distinctions among them, and sheds light on their varying capabilities as 'data-driven' smoothers for risk trends exploration and sequential disease mapping. The methods are motivated and illustrated through a Bayesian analysis of adverse medical events (commonly known as iatrogenic injures) to hospitalized children and youth in British Columbia, Canada. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian disease mapping; spatiotemporal models; regression spline; P-spline; smoothing spline; multivariate conditional autoregressive (MCAR) prior

1. INTRODUCTION

Bayesian disease mapping often analyzes small area data that are aggregated counts or rates of particular disease or health outcomes to 'at risk populations' reside in contiguous subregions such as census divisions or health administrative areas. When such data are made available for a particular time period, say, a 1- or 10-year time period, they are often denoted by y_i, the disease or health outcome occurrence, and n_i, the at risk population, $i = 1, \ldots, N$ for N geographical subregions. The subsequent analysis is often pursued by Bayesian hierarchical modeling of the count data via a spatial prior specification for a 'spatially varying' (log) relative risks ensemble (Lawson *et al*., 1999; Leroux *et al*., 1999, Ugarte, *et al*., 2006; MacNab *et al*., 2006; to name a few).

*Correspondence to: Y. C. MacNab, Centre for Healthcare Innovation and Improvement, British Columbia Child and Family Research Institute, University of British Columbia, Room E417, 4480 Oak Street, Vancouver, BC, Canada V6H 3V4.
†E-mail: ymacnab@interchange.ubc.ac

Received 30 October 2006
Accepted 22 June 2007

Model fitting and posterior risk inference are often implemented using Markov chain Monte Carlo (MCMC) methods.

When such data are made available for mapping of several consecutive time periods T, say annual data for 10 years (i.e., $T = 10$), they are often denoted by y_{it} and $n_{it}, t = t_1, \ldots, t_T$. With the primary objective to highlight the underlying temporal trends and emerging spatial patterns of disease risks, recent literature on Bayesian disease mapping presents spatiotemporal models for the analysis of such data (Waller *et al*., 1997, Böhning *et al*., 2000, MacNab and Dean, 2001; Knorr-Held and Richardson 2003; MacNab and Gustafson, 2007). In particular, MacNab and Dean (2001) proposed a spatiotemporal model framework that incorporates autoregressive local smoothing and temporal regression B-spline smoothing for non-linear small-area risk trends estimation. Poisson mixed effects models were developed with estimation and inference via penalized quasilikelihood method (Breslow and Clayton, 1993).

More recently, MacNab and Gustafson (2007) presented a fully Bayesian (FB) extension to the MacNab–Dean work. They also proposed a broader class of regression B-spline models so that the spatiotemporal relative risks may be modeled by spatially varying or randomly varying regression B-splines. In particular, they introduced spatially structured multivariate conditional autoregressive (multivariate Gaussian Markov random field) priors and spatially unstructured multivariate Gaussian priors for the random spline coefficients ensemble. Region-specific temporal relative risk trends may therefore be modeled by regression B-splines that vary in a spatially structured or spatially unstructured manner, and the interplay between spatial and temporal smoothing may be explored. In addition, to facilitate spline smoothing of greater flexibility, the MacNab–Gustafson study also extended the 1-knot regression B-spline method of MacNab–Dean to present a regression spline approach with adaptive knot selection. A recently developed Bayesian hierarchical model selection criterion, the deviance information criterion (DIC), was used to assess the trade-off between goodness-of-fit and smoothness, to select the number of knots, and for prior and model assessment.

Literature on regularized/penalized spline smoothing methods is extensive; of note is that the smoothing splines and P-splines are two main stream and competing methods. Briefly, smoothing splines are penalized splines of many knots (knots on all data points). The method can be characterized as penalized spline regression with respect to a set of natural basis functions and a roughness penalty—the well-known integrated squared second derivative penalty. The P-splines allow for a smaller number of knots to be placed in a more liberal manner. To regulate goodness-of-fit and smoothness, the P-spline fitting is often formulated in terms of penalized spline regression under a 'difference penalty' on the coefficients of adjacent B-spline bases (Eilers and Marx (1996) or other splines bases such as the truncated power functions (Ruppert and Carroll, 2000; Wand, 2000).

Additive models with smoothing splines and P-splines have been developed for their mixed model representations under appropriate reparameterization; model fittings via empirical Bayes (EB) or FB methods of estimation and inference have been proposed, and both methods have seen recent applications in multiple/ensemble (also named subject-specific) curves fittings to biological, medical, health science, and environmental data (Brumback and Rice, 1998; White *et al*., 1999; Verbyla *et al*., 1999; Durban *et al*., 2004; Lambert and Eilers, 2005). In particular, the methods have also been developed within the generalized linear mixed modeling (GLMM) context for modeling non-Gaussian data (Lin and Zhang, 1999; Zhao *et al*., 2006). With particular prior assumptions that correspond to their penalty specifications, FB implementation of smoothing spline and P-spline fittings have seen increased applications (Hastie and Tibshirani, 1990; Hastie and Tibshirani, 2000; Lang and Brezger, 2004; Brezger and Lang, 2006; Zhao *et al*., 2006).

The temporal B-spline smoothing presented in MacNab and Gustafson (2007) represents a non-regularized (unpenalized) regression spline method for fittings of region-specific rate and risk curves.

DOI: 10.1002/env

Its FB approach to estimation and inference may be readily implemented via GLMM formulation and MCMC, and the method allows for considerably flexible random effects prior specifications. Our main contributions in this paper are (1) extending the MacNab–Gustafson work to present Bayesian disease mapping models with regression B-splines, P-splines, and smoothing splines for temporal rate and risk smoothing, (2) exploring the connections and distinctions among the three smoothing methods, and (3) examining their analytic capabilities as 'data-driven' smoothers for relative risk estimation and inference. In particular, smoothing splines and P-spines are put into a FB inferential framework for spatiotemporal modeling of disease mapping data. A common feature of the Bayesian smoothing spline and P-spline approaches is that the modeling may be implemented via GLMM formulation that places specific assumptions on the design matrix and on the random effects prior so that the corresponding penalty is proportional to the negative logarithm of the prior density. In this study, we show that in comparison with the regression B-spline approach, the smoothing spline and P-spline methods only allow for limited smoothing prior options; and in the FB inferential context, the smoothing spline and P-spline fittings can be heavily influenced by the hyperprior specifications, that is, the prior specifications of the dispersion (or smoothing) parameters. We also propose and explore formulations of 4-level hierarchical smoothing splines and P-spline models that enable posterior estimation of the hyper-parameters (i.e., estimation of the posteriors of the smoothing parameters) and facilitate exploration of the methods as 'data-driven' smoothers for robust Bayesian rate and risk estimation and inference.

The paper is organized as follows. In the following Section 2, we motivate and describe the Bayesian hierarchical model framework for a class of spatiotemporal models with temporal regression B-splines, P-splines, and smoothing splines for fittings of non-linear temporal effects. In Section 3, the methods and their MCMC implementations are illustrated through a detailed Bayesian analysis of the BC AME data. Section 4 provides a summary discussion.

2. THE BAYESIAN HIERARCHICAL MODEL FRAMEWORK

For modeling of rare disease or health outcomes, we assume that conditioning on a vector $\boldsymbol{b}$ of random effects the data $\{y_{it}\}$ arise independently from the Poisson family with intensity μ_{it}^b:

$$\log(\mu_{it}|\boldsymbol{b}) = \log(n_{it}) + m_t + b_{it} \tag{1}$$

where $\log(n_{it})$ is an offset, m_t is the 'global' rate at time t, and b_{it} the area random effect representing log relative risk for the ith area at time t. To enable exploration of the underlying temporal rate and risk trends via non-parametric statistical modeling, the m_t and b_{it} may be replaced by smoothing functions (MacNab and Gustafson, 2007):

$$m_t = a_0 + S_0(t), \quad b_{it} = b_{i0} + RS_i(t)$$

so that

$$\log(\mu_{it}|\boldsymbol{b}) = \log(n_{it}) + a_0 + S_0(t) + b_{i0} + RS_i(t) \tag{2}$$

Here, t is centered; $S_0(t)$ and $RS_i(t)$ represent arbitrary smoothing functions; $a_0 + S_0(t)$ represents 'global' rate trend, with $m = \exp(a_0)$ representing mid-period 'global' rate; $b_{i0} + RS_i(t)$ represents regional (the ith region) risk trend relative to the 'global' rates, with b_{i0} representing the mid-period random area effect. Note that letting $m_{it} = a_0 + S_0(t) + RS_i(t)$, Equation (2) defines a model with

DOI: 10.1002/env

spatially varying rate trend m_{it} and a time-constant spatial effect b_{i0}. Note also that the model framework presented herein may be readily extended to include covariates; the Poisson likelihood may also be readily replaced by a Binomial likelihood to accommodate non-rare rates.

2.1. *Smoothing with unpenalized regression B-splines*

Placed into a Bayesian hierarchical model framework, we briefly review the adaptive regression B-spline method that has been presented in MacNab and Gustafson (2007). The main feature of the method is that the relative risks smoothing functions (i.e., the $RS_i(t)$s) are fitted by random coefficients regression in terms of B-spline basis functions, and the resulting splines are piecewise polynomials with a relatively small number of interior knots.

Specifically, if one assumes a L-knot regression cubic B-spline for the arbitrary smoothing function $S_0(t)$ and a family of L-knot regression cubic B-splines for $RS_i(t)$, the following generalized mixed effects model may be derived:

$$\log(\mu_{it}|\boldsymbol{b}) = \log(n_{it}) + a_0 + \sum_{k=1}^{K} a_k B_k(t) + b_{i0} + \sum_{k=1}^{K} b_{ik} B_k(t) \tag{3}$$

where $\{a_k, k = 1, \ldots, K\}$ are fixed effects; $\{B_k\}$, $k = 1, 2, \ldots, K$, is a set of B-spline basis functions (without the intercept), $B_k(t)$ denoting the kth B-spline basis function evaluated at time t, t is centered; $K = L + 3$; $\boldsymbol{b}_k = (b_{1k}, \ldots, b_{Nk})^\top$ are random spline coefficients.

Using recently developed Bayesian hierarchical model selection criterion, the DIC (Spiegelhalter *et al*., 2003a), knot selection may be carried out as follows: Implementing in sequence the model fitting with B-splines of equally spaced $1, 2, \ldots, L$ interior knots; examining the temporal smoothing and goodness-of-fit via DIC, together with visual inspection of the deviance residuals and the fitted rate and risk splines; and selecting a minimum number of knots based on the best DIC, that is, the fitted splines indicating the best balance between goodness-of-fit and smoothness.

To facilitate spatiotemporal risks smoothing, one places second-level random effects prior(s) on the NK-vector of regression parameters $\boldsymbol{b} = (b_{10}, \ldots, b_{N0}, \ldots, b_{1K}, \ldots, b_{NK})^\top$. Note that the aforementioned regression spline fitting method places no restriction on the random effects prior choice. As such, model (3) facilitates a variety of prior options and accommodates spatially varying and randomly varying smoothing functions for the modeling of temporal risk trends. For example, denoting $\boldsymbol{b}_i = (b_{i0}, b_{i1}, \ldots, b_{iK})^\top$ for $i = 1, \ldots, N$, the regression B-spline (3) estimation and inference may be implemented via spatial priors such as multivariate Gaussian Markov random fields $\boldsymbol{b} \sim \text{MVN}(\boldsymbol{0}, \Omega)$ (Ω is the precision matrix) that are formulated via their corresponding full conditionals with mean $\text{E}(\boldsymbol{b}_i|\boldsymbol{b}_{-i})$ and precision matrix $\text{Prec}(\boldsymbol{b}_i|\boldsymbol{b}_{-i})$, where $\boldsymbol{b}_{-i} = (\boldsymbol{b}_1^\top, \ldots \boldsymbol{b}_{i-1}^\top, \boldsymbol{b}_{i+1}^\top, \ldots, \boldsymbol{b}_N^\top)^\top$ (MacNab and Gustafson, 2007).

In this paper, and without loss of generality, we illustrate the B-spline method using one of the multivariate conditional auto regressive (MCAR) formulations presented in MacNab and Gustafson (2007). Specifically, defining 'neighborhood' by area adjacency, that is, regions that share a common border are neighbors, we consider the following MCAR for spatially varying B-splines:

$$\text{E}(\boldsymbol{b}_i|\boldsymbol{b}_{-i}) = \lambda(1 - \lambda + \lambda m_i)^{-1} \sum_{j \neq i} w_{ij} \boldsymbol{I}_P \boldsymbol{b}_j, \quad \text{Prec}(\boldsymbol{b}_i|\boldsymbol{b}_{-i}) = (1 - \lambda + \lambda m_i)\boldsymbol{\Gamma}, \tag{4}$$

DOI: 10.1002/env

$$\boldsymbol{\Omega} = (\boldsymbol{L} - \lambda \boldsymbol{W}) \otimes \boldsymbol{\Gamma}$$

where m_i is the number of neighbors for the ith region, $\lambda \in (0, 1)$, $\boldsymbol{L} = \text{diag}(1 - \lambda + \lambda m_1, \ldots, 1 - \lambda + \lambda m_N)$, I_P identity matrix of dimension P, and $\boldsymbol{\Gamma}$ is a unstructured P by P symmetric positive definite matrix, $P = K + 1$; $\boldsymbol{W}$ is a N by N matrix with elements $w_{ii} = 0$, $w_{ij} = 1$ if the ith region and the jth region are neighbors and $w_{ij} = 0$ otherwise. Note that the parameter λ may be regarded as a 'spatial' smoothing parameter as it controls the degree of autoregressive 'local' smoothing, and the precision matrix Γ may be regarded as a temporal 'smoothing matrix'. Note also that MCAR (4) is a multivariate extension of the following univariate CAR:

$$\text{E}(b_i|b_{-i}) = \frac{\lambda \sum_{j \neq i} w_{ij} b_j}{(1-\lambda+\lambda m_i)^{-1}}, \quad \text{Prec}(\boldsymbol{b}_i|\boldsymbol{b}_{-i}) = (1 - \lambda + \lambda m_i)\tau,$$
$$\text{Prec}(\boldsymbol{b}) = \tau(\lambda \boldsymbol{Q} + (1 - \lambda) I_N) \tag{5}$$

where $\boldsymbol{Q}$ is a N by N matrix with elements $Q_{ii} = m_i$, $Q_{ij} = -1$ if the ith region and the jth region are neighbors and $Q_{ij} = 0$ otherwise (Leroux *et al*., 1999). When $\lambda = 0$, the MCAR (4) reduced to a non-spatial prior for randomly varying B-splines:

$$\text{E}(\boldsymbol{b}_i) = 0, \quad \text{Prec}(\boldsymbol{b}_i) = \boldsymbol{\Gamma} \tag{6}$$

When $\boldsymbol{\Gamma} = \text{diag}(\gamma_0, \gamma_1, \ldots, \gamma_K)$, the MCAR (4) reduce to an independent CAR with the corresponding block diagonal precision matrices given in Equation (5).

2.2. *Smoothing with penalized alternatives: the smoothing splines and P-splines*

The aforementioned regression splines of a small number of interior knots may facilitate limited flexibility in non-linear modeling of rates and ratios. The smoothing spline and P-spline, on the other hand, are penalty splines with a relatively large number of knots. In this study, we relate and compare cubic regression B-splines to the two smoothing alternatives.

Briefly, well known for several decades in the statistics literature and related applications, a (cubic) smoothing spline (also known as the 'natural cubic spline') is a roughness penalty spline with knots at the data points (i.e., at all distinct design points). The method can be formulated as a basis functions spline fitting $\sum_{l=1}^{L} a_l B_l(t)$ subject to a natural boundary condition ($\sum_{l=1}^{L} a_l B_l''(t) = \sum_{l=1}^{L} a_l B_l'''(t) = 0$ when $t = t_1$ and $t = t_T$) and an integrated squared second derivative roughness penalty $\gamma \int_{t_1}^{t_T} \{\sum_{l=1}^{L} a_l B_l''(t)\}^2 \text{d}t$; $\gamma \geq 0$ is a (temporal) smoothing parameter; $B_l''(t)$ and $B_l'''(t)$ denote the second and third derivatives of the cubic B-spline basis function evaluated at time t (O'Sullivan, 1986; Eilers and Marx, 1996). Note that the *natural boundary condition* (thus the name *natural spline*) puts a particular boundary constraint on the smoothing spline: The spline is linear on the two extreme intervals (Green and Silverman, 2000).

Also well known but comparatively new, the P-splines on the other hand accommodate more flexible placement of the interior knots and are without any boundary constraint. The method is often characterized as a penalized B-spline solution $\sum_{m=1}^{M} a_m B_m(t)$ to a 'difference penalty' on the coefficients of the adjacent (cubic) B-spline basis: $\gamma(\sum_{m=r+1}^{M} \Delta^r a_m)^2$, where Δ^r represents rth-order difference and $\gamma > 0$ is a smoothing parameter (Eilers and Marx 1996). It was shown that with a sufficient number of interior knots, the actual placement of the knots (i.e., knot locations) often have

DOI: 10.1002/env

little influence on the P-spline fitting (Ruppert, 2002). As such, in real-life application one typically places a moderate number of equally spaced interior knots over the full range of the data points.

It is worth noting that smoothing spline is typically introduced and fitted via natural spline basis functions that satisfy the natural spline (boundary) constraints (Silverman, 1985; Green and Silverman, 2000). The P-spline has also been introduced using truncated power basis functions (Ruppert and Carroll, 2000; Wand, 2000) that are readily interpretable. Because of the mathematical equivalence of the basis functions, the fitted curves are not affected by the choice of the basis functions. Nevertheless, for numerical stability B-spline basis functions are typically used for the implementation of the P-spline fittings; see Ruppert *et al*. (2003) for details.

With recent work on mixed-effects representations of the smoothing splines (Brumback and Rice 1998; Verbyla *et al*. 1999; Lin and Zhang 1999) and P-splines (Rupert and Carroll 2000; Wand 2000), additive fittings of smoothing splines and P-splines can be formulated and implemented within linear mixed-effects model and GLMM frameworks; the corresponding Bayesian justifications and formulations for FB implementation may also be developed (Ruppert *et al*., 2003; Zhan *et al*., 2006). In particular, with respect to the penalty specifications, the Bayesian hierarchical smoothing spline and P-spline model formulations may be developed via GLMMs that place specific assumptions on the design matrices of the random effects, say C for the smoothing spline and A for the P-spline, and on the precision matrices of the random effects priors.

Specifically, the mixed effects representations of smoothing spline and P-spline may be placed into the GLMM context (2) and the Bayesian hierarchical model framework for Baysian disease mapping. For example, after some algebra, model (2) with smoothing splines for the temporal effects S_0 and RS_is may be formulated as follows:

$$\log(\mu_{it}|\boldsymbol{b}) = \log(n_{it}) + u_0 + u_1 t + \sum_{k=2}^{T-1} u_k C_k(t) + v_{i0} + v_{i1} t + \sum_{k=2}^{T-1} v_{ik} C_k(t) \tag{7}$$

where (u_0, u_1) are fixed effects, $\boldsymbol{u} = (u_2, \ldots, u_{T-1})^\top$ are random effects, $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma_u^2 I_{T-1})$, σ_u^2 is the smoothing parameter, (v_{i0}, v_{i1}) and $\boldsymbol{v}_i = (v_{i2}, \ldots, v_{iT-1})^\top$ are random effects, $\boldsymbol{v}_i \sim N(\boldsymbol{0}, \sigma_i^2 I_{T-1})$, and $\tau_i = \sigma_i^{-2}$ is a smoothing parameter, $i = 1, \ldots, N$; $\boldsymbol{C} = (C_k(t))$ is the corresponding T by $T-2$ design matrix for the random effects, $\boldsymbol{C} = \boldsymbol{L}(\boldsymbol{L}^\top \boldsymbol{L})^{-1}$, $\boldsymbol{L} = \boldsymbol{Q}\boldsymbol{U}^{-1}$, $\boldsymbol{U}$ is the Choleski decomposition of $\boldsymbol{R}$, $\boldsymbol{R}$ and $\boldsymbol{Q}$ are the two band matrices defined in Green and Silverman (2000).

Similarly, the model (2) with P-splines (with respect to a 2nd-order difference penalty) may be written as

$$\log(\mu_{it}|\boldsymbol{b}) = \log(n_{it}) + \alpha_0 + \alpha_1 t + \sum_{k=2}^{T} \alpha_k A_k(t) + \beta_{i0} + \beta_{i1} t + \sum_{k=2}^{T} \beta_{ik} A_k(t) \tag{8}$$

where (α_0, α_1) are fixed effects, $\boldsymbol{\alpha} = (\alpha_2, \ldots, \alpha_T)^\top$ are random effects, $\boldsymbol{\alpha} \sim N(\boldsymbol{0}, \sigma_\alpha^2 I_T)$, σ_α^2 is the smoothing parameter; (β_{i0}, β_{i1}) and $\boldsymbol{\beta}_i = (\beta_{i2}, \ldots, \beta_{iT})^\top$ are random effects, $\boldsymbol{\beta}_i \sim N(\boldsymbol{0}, \sigma_i^2 I_T)$, $\tau_i = \sigma_i^{-2}$ is the smoothing parameter, $i = 1, \ldots, N$; $\boldsymbol{A} = (A_k(t))$ is the corresponding T by $T-1$ design matrix for the random effects, $\boldsymbol{A} = \boldsymbol{B}D^\top(DD^\top)^{-1}$, $\boldsymbol{B}$ is the corresponding B-spline bases matrix, D is the difference matrix defined in Eilers and Marx (1996). Note that for the comparison of the three smoothing methods, the P-splines, as defined by Equation (8), are penalty B-splines with respect to a 2nd-order difference penalty and a set of interior knots placed at the (unique) values of the time t. For a

DOI: 10.1002/env

10-year mapping of annual rates, for example, 8-knot P-splines are compared with the smoothing splines and the regression B-splines. As shown in Eilers and Marx (1996) and in this study (see Section 3), the smoothing splines and the P-splines are expected to produce comparable smoothing. In Section 3, we also present a brief report on equally spaced K-knot P-spline fitting to the BC AME data, and for $K = 5, 6, 7$, respectively. It is worth mentioning that the P-spline method accommodates higher (or lower) order difference penalty, although the associated mixed-effects model representation will be different. In particular, note that with respect to the 2-order difference penalty, $\alpha_0 + \alpha_1 t$, as well as $\beta_{i0} + \beta_{i1} t$, represents the parametric (linear) part of the P-spline, and $\sum_{k=2}^{T} \alpha_k A_k(t)$, as well as $\sum_{k=2}^{T} \beta_{ik} A_k(t)$, the non-parametric (non-linear) part of the P-spline curve. The parametric part of a P-spline of rth-order difference penalty should be a parametric polynomial of $r - 1$ degree (Ruppert *et al*. 2003; Fahrmeir *et al*., 2004).

Also for comparison purpose, we explore the following second-level prior formulations:

$$\boldsymbol{u} \sim \text{MVN}(\boldsymbol{0}, \sigma_u^2 I_{T-2}),\ \nu_{i0} \sim N(0, \sigma_0^2),\ \nu_{i1} \sim N(0, \sigma_1^2),\ \boldsymbol{v}_i \sim \text{MVN}(0, \sigma_2^2 I_{T-2}) \tag{9}$$

or

$$\begin{aligned} &\boldsymbol{u} \sim \text{MVN}(\boldsymbol{0}, \sigma_u^2 I_{T-2}),\ \nu_{i0} \sim N(0, \sigma_0^2),\ \nu_{i1} \sim N(0, \sigma_1^2), \\ &\boldsymbol{v}_i \sim \text{MVN}(0, \sigma_{i+1}^2 I_{T-2}), i = 1, \ldots, N \end{aligned} \tag{10}$$

for the smoothing splines, and

$$\boldsymbol{\alpha} \sim \text{MVN}(\boldsymbol{0}, \sigma_\alpha^2 I_{T-1}),\ \beta_{i0} \sim N(0, \sigma_0^2),\ \beta_{i1} \sim N(0, \sigma_1^2),\ \boldsymbol{\beta}_i \sim \text{MVN}(\boldsymbol{0}, \sigma_2^2 I_{T-1}) \tag{11}$$

or

$$\begin{aligned} &\boldsymbol{\alpha} \sim \text{MVN}(\boldsymbol{0}, \sigma_\alpha^2 I_{T-1}),\ \beta_{i0} \sim N(0, \sigma_0^2), \\ &\beta_{i1} \sim N(0, \sigma_1^2),\ \boldsymbol{\beta}_i \sim \text{MVN}(\boldsymbol{0}, \sigma_{i+1}^2 I_{T-1}), i = 1, \ldots, N \end{aligned} \tag{12}$$

for the P-splines.

Note that within the GLMM framework the coefficients ν_0, ν_1, β_0, and β_1 are fixed effects. Note also that for single-curve smoothing spline and P-spline fittings, the corresponding mixed-effects representations assume that ν_{i0}s, ν_{i1}s, β_{i0}s, and β_{i1}s, the coefficients with respect to the constant and linear functions, to be fixed effects. In ensemble-curve fitting in general and in this study in particular, they may be treated as random effects with common priors for all local regions to enable 'pooling data' across 'small areas'. For example, the priors (10) and (12) facilitate 'borrowing strength' across the local areas, both for the linear trend components (the intercepts and slops) as well as the non-linear components (the random spline coefficients) of the smoothing functions. Under smoothing spline (9) (smoothing spline 1 hereafter), as well as P-spline (11) (P-spline 1 hereafter), the relative risk splines would be fitted with the same degree of smoothness across the local areas, as we assume a single ('global') smoothing parameter, that is, $\sigma_{i+1} = \sigma_2^2, i = 1, 3, \ldots, N$. Under priors (10) (smoothing spline 2 hereafter) and (12) (P-spline 2 hereafter), on the other hand, the variance parameters $\{\sigma_{i+1}^2, i = 1, \ldots, N\}$ are allowed to vary across the areas and serve as 'local' temporal smoothing parameters, that is the corresponding priors allow the degrees of spline smoothing to differ from area to area, and permit unequal variances for the resulting splines.

DOI: 10.1002/env

In addition, the CAR prior (5) may be considered for spatially varying intercepts or/and slopes ν_{i0}s, ν_{i1}s, β_{i0}s, and β_{i1}s. For example, one may consider

$$\boldsymbol{\nu}_0 \sim \text{MVN}(\mathbf{0}, \Omega_{\boldsymbol{\nu}_0}), \ \boldsymbol{\beta}_0 \sim \text{MVN}(\mathbf{0}, \Omega_{\boldsymbol{\beta}_0}) \tag{13}$$

or

$$\boldsymbol{\nu}_0 \sim \text{MVN}(\mathbf{0}, \Omega_{\boldsymbol{\nu}_0}), \ \boldsymbol{\nu}_1 \sim \text{MVN}(\mathbf{0}, \Omega_{\boldsymbol{\beta}_1}), \ \boldsymbol{\beta}_0 \sim \text{MVN}(\mathbf{0}, \Omega_{\boldsymbol{\beta}_0}), \ \boldsymbol{\beta}_1 \sim \text{MVN}(\mathbf{0}, \Omega_{\boldsymbol{\beta}_1}) \tag{14}$$

where $\boldsymbol{\nu}_0 = (\nu_{10}, \ldots, \nu_{N0})^\top$, $\boldsymbol{\nu}_1 = (\nu_{11}, \ldots, \nu_{N1})^\top$, $\boldsymbol{\beta}_0 = (\beta_{10}, \ldots, \beta_{N0})^\top$, $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{N1})^\top$, and $\Omega_{\boldsymbol{\nu}_0}$, $\Omega_{\boldsymbol{\nu}_1}$, $\Omega_{\boldsymbol{\beta}_0}$, and $\Omega_{\boldsymbol{\beta}_1}$ are the CAR precision matrices as defined in Section 2.1. The MCAR prior discussed in Section 2.1 may also be considered for joint modeling of the random intercepts and slopes:

$$\boldsymbol{\nu} \sim \text{MCAR}(\mathbf{0}, \Omega_{\boldsymbol{\nu}}) \tag{15}$$

where $\boldsymbol{\nu} = (\nu_{10}, \nu_{11}, \ldots, \nu_{N0}, \nu_{N1})^\top$, and

$$\boldsymbol{\beta} \sim \text{MCAR}(\mathbf{0}, \Omega_{\boldsymbol{\beta}}) \tag{16}$$

where $\boldsymbol{\beta} = (\beta_{10}, \beta_{11}, \ldots, \beta_{N0}, \beta_{N1})^\top$.

2.3. The third-level hyperpriors

We explore 'vague' and flat hyperpriors for the third-level hierarchical specifications. In particular, diffuse normal prior $N(0, 0.0001)$ may be considered for all fixed effects, and we assume $\tau = \sigma^{-2} \sim$ Gamma(0.0001, 0.0001), $\lambda \sim$ Unif(0, 1). Under the non-spatial prior (6), as well as the MCAR (4), we assume a vague Wishart prior for the precision matrix: $\Gamma \sim$ Wishart(R^{-1}, p), where R is a $(K+1)$ by $(K+1)$ identity matrix, $p \geq (K+1)$ is the shape parameter ('degrees of freedom') and assumed to be $(K+1)$ to represent a 'non-informative' hyperprior. Since hyperprior choices are known to influence spline smoothing, we present a comprehensive sensitivity analysis for each of the three smoothing methods (see Section 3).

2.4. Four-level hierarchical formulations of smoothing spline and P-spline models

We also propose and explore 4-level hierarchical formulations of the smoothing spline 2 and P-spline 2 for posterior estimation of the smoothing parameters and for robust inference on the small area rate and risk splines. Specifically, we discuss the following hyperprior formulations:

$$\tau_u \sim \text{gamma}(s, r), \ \tau_i \sim \text{gamma}(s, r), \ i = 0, 1, \ldots, N+1 \tag{17}$$

or

$$\tau_i \sim \text{gamma}(s, r), \ i = 2, \ldots, N+1 \tag{18}$$

DOI: 10.1002/env

assuming fourth-level hyper-hyperpriors: $s \sim \text{unif}(0, 10)$ and $r \sim \text{unif}(0, 10)$. Note that for hyperprior (18), we also assume $\tau_u \sim \text{gamma}(0.0001, 0.0001)$, $\tau_i \sim \text{gamma}(0.0001, 0.0001)$, $i = 0, 1$.

3. APPLICATION: THE BRITISH COLUMBIA IATROGENIC INJURY STUDY

To illustrate the methodologies presented in Section 2, we present an in-depth analysis of hospital admission/separation data on non-fatal iatrogenic injuries among male patients of 1–19 years of age in BC's 84 local health areas (LHAs). Hospitalized individuals were aggregated to LHAs based on patient place of residence and to 1991–2000 annually based on date of admission. The main objectives of the spatiotemporal analysis are (1) to explore the underlying temporal trends in the BC rate and in the local health area rates and ratios, (2) to inform sequential spatial risk variations and risk cluster evolution, and (3) to present adequate statistical inference on small area relative risks.

Based on the 1991–2000 data, non-fatal adverse medical events occurred to approximately 3% of the hospitalized males aged 1–19 in BC. The provincial rate trend was moderately non-linear: it rose between 1991 and 1994, dropped low in 1995 and 1996, rose again in 1997 and 1998, and declined again in 1999 and 2000. The observed rates (y_{it}/n_{it}) for the LHAs fluctuated from year to year (see Figure 1, the crude rates are shown in dashed-lines).

The full Bayesian analysis was implemented using WinBUGS 1.4 (Spiegelhalter *et al*., 2003b). For each model fitting, we ran three parallel chains from dispersed starting points. Pilot runs were carried out to set burn-in, thinning, and posterior sample accumulation for adequate posterior prediction and inference. Convergence was assessed by visually assessing the trace, Gelman–Rubin statistics, and the posterior density plots of all parameters. To ensure that representative samples were drawn for posterior inference and cross model comparisons, three chains of 10 000-sample (after thinning) each were retained, setting burn-in at 2000 and thinning at 10. The computing time, under a Pentium 4 CPU 3.20 GHz 100 GB RAM, was about 4 (usually models with non-spatial priors) to 7 (usually models with MCAR prior) hours for the 4-knot B-spline fitting and 8–10 h for the smoothing spline 2 and the P-spline 2.

3.1. The B-spline fittings

We run MCMC simulations for models (1) and (3) and for a variety of spatial (CARs and MCARs) and non-spatial prior options presented in MacNab and Gustafson (2007). To conserve space, however, this paper only present and discuss the B-spline fittings for the selected spatial and non-spatial priors described in Section 2 (see Table 1), as they represent priors of better fits with respect to the Dbar-pD-DIC comparison; see MacNab and Gustafson paper (2007) for a detailed exposition on the spatiotemporal model with B-splines and a variety of spatial (CARs and MCARs) and non-spatial priors.

For selected local health areas, Figure 1 illustrates a comparison of the estimated rate splines for the 1- and 4-knot models with (non-spatial) prior (6). Overall, with increasing knots the B-splines facilitated decreased level of smoothing to the rates; the resulting values of DIC assessment indicated increasing model complexity (higher pD, also a measure of smoothness) and improving goodness-of-fit (smaller Dbar). Increasing knots also resulted in increased uncertainty in the Bayesian estimates; i.e. 'higher knot' splines resulted in higher posterior standard deviation in rate and risk estimates (lower posterior precision). The 1-knot splines explored gradually changing temporal rate and risk trends,

DOI: 10.1002/env

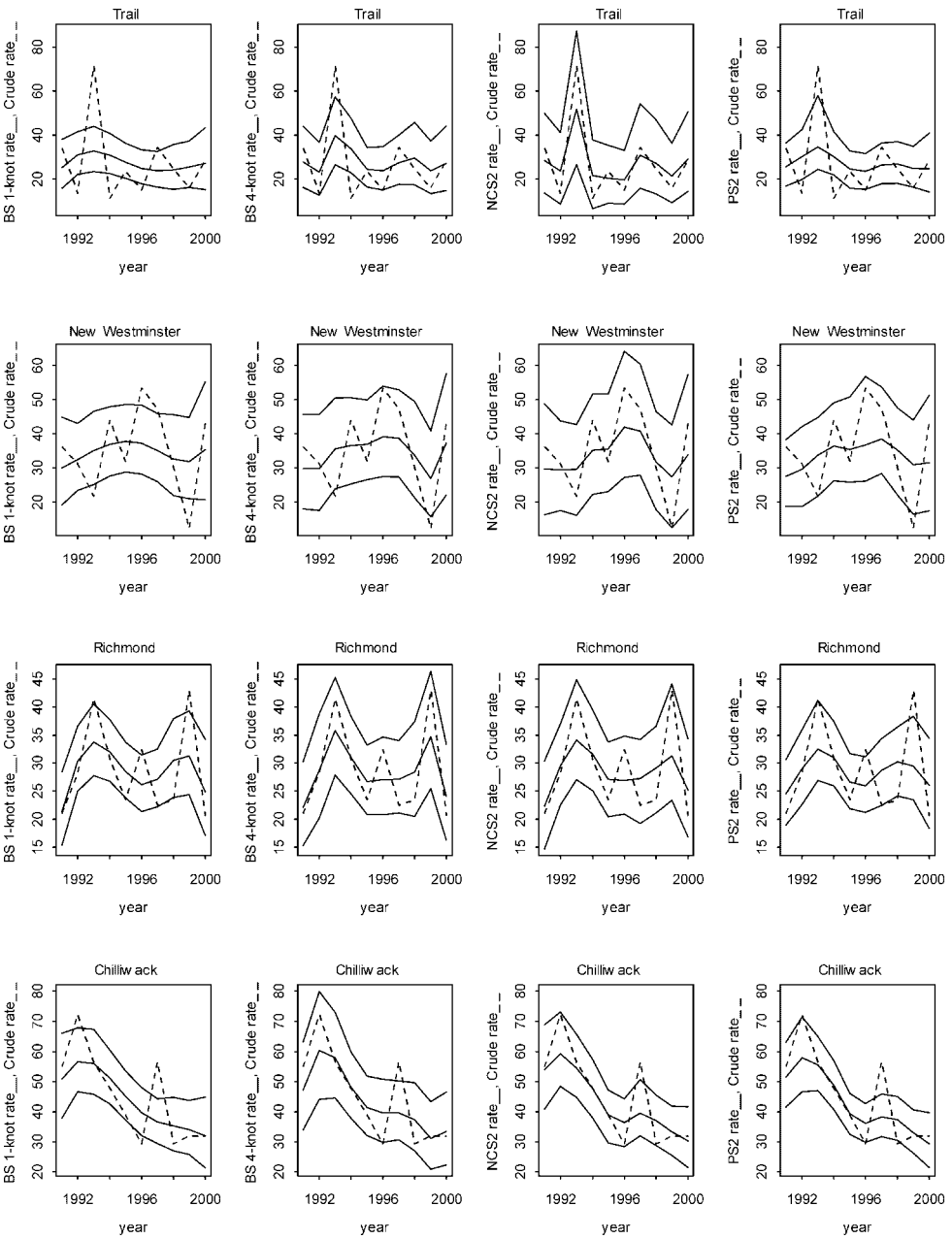

Figure 1. Crude and Bayesian posterior rate (per 1000 patients) estimates for select local health areas: dashed line—crude rates, solid lines—posterior means, 2.5- and 97.5-percentiles. The Bayesian estimates are derived from the spatiotemporal model (3): BS 1-knot—1-knot B-spline model with non-spatial prior (6), BS 4-knot rate—4-knot B-spline model with non-spatial prior (6), NCS2 rate—3-level smoothing spline model 2, PS2 rate—3-level P-spline model 2

DOI: 10.1002/env

Table 1. Deviance information criterion for the spatial model (1) and selected spatiotemporal models with B-splines and for independent Gaussian prior (IID), the CAR prior (Leroux *et al.* CAR (5)), the MCAR prior (multivariate CAR (4)), and the non-spatial prior (6)

Model	Dbar	pD	DIC
Spatial model, IID prior	3952	361	4312
Spatial model, CAR prior	3955	345	4300
1-knot B-spline model, CAR prior	4241	169	4410
1-knot B-spline model, MCAR prior	4203	186	4389
1-knot B-spline model, non-spatial prior	4178	206	4384
2-knot B-spline model, non-spatial prior	4130	236	4366
3-knot B-spline model, non-spatial prior	4041	272	4314
4-knot B-spline model, non-spatial prior	**4014**	**295**	**4310**
5-knot B-spline model, non-spatial prior	4002	311	4313
4-knot B-spline model, MCAR	4063	263	4326

The results are for the adverse medical events to male patients 1–19 years of age, BC, Canada, 1991–2000.

whereas the 4-knot splines explored more 'localized' trends (i.e., the peaks and valleys) but indicated higher posterior uncertainty (i.e., wider credible intervals). The 5-knot splines showed signs of overfit, observed from the visual inspection of the rate and ratio estimates (not shown here to conserve space) and the DIC statistics (see Table 1). The DIC assessment (also see Table 1) favored the non-spatial prior (6) over the rest of the prior options, and suggested that the MCAR performed better than the independence CAR; the changes in the resulting rate and risk spline estimation and inference, however, were relatively moderate.

We run sensitivity analysis with respect to several alternative Wishart specifications. For example, for the 4-knot B-splines, we rerun MCMC simulations for the $\Sigma^{-1} \sim \text{Wishart}(R^{-1}, p)$ options: $R = \text{diag}(1, 8)$, $p = 10$ or $p = 15$ and $p = 8$, $R = \text{diag}(0.1, 8)$ or $R = \text{diag}(0.01, 8)$ or $R = \text{diag}(5, 8)$ or $R = \text{diag}(10, 8)$, respectively. As to be expected, the varying Wishart prior specifications on R and p resulted in moderate changes in the degrees of smoothing imposed on the rates and risks. However, the resulting changes in the rate and risk inference, in terms of suggesting evidence of high/low risk trends and clusters, remained relatively small. For the CAR and MCAR priors, sensitivity analysis with respect to $\lambda \sim \text{beta}(8, 4)$, $\lambda \sim \text{beta}(4, 8)$, $\alpha \sim \text{beta}(2, 18)$, and $\alpha \sim \text{beta}(18, 2)$ indicated moderate changes in posterior inference for the hyperparameters, which were also to be expected. The resulting inferences for the rates and risks, however, were fairly robust to the hyperprior specifications.

3.2. The smoothing spline and P-spline fittings

The 3-level smoothing spline and P-spline fittings were implemented for the aforementioned model, prior, and hyperprior formulations; Table 2 presents the DIC results. By and large, for the same knot-placement (8 interior knots), the two penalized smoothing methods under a 'global' smoothing parameter led to similar posterior predictions of the relative risks ensemble, and they are comparable to the 1-knot B-splines. The two penalty splines under varying 'local' smoothing parameters were also comparable smoothers, although the fitted P-spline 2 was shown to be comparably smoother and closer to the 4-knot B-splines (see Figure 1). Nevertheless, reportable differences from the fitted rate and risk spline inference were only seen from a few LHAs that had sharp rate/ratio change or occasional extreme rates

DOI: 10.1002/env

Table 2. Deviance information criterion for selected 3-level P-spline and smoothing spline models

Model	Hyper-prior	Dbar	pD	DIC
P-spline 1, 8-knot	$\tau_i \sim$ gamma(0.0001, 0.0001)	4233	182	4416
	$\tau_i \sim$ gamma(0.5, 0.0005)	4233	181	4414
	$\tau_i \sim$ gamma(0.01, 0.01)	4222	189	4411
	$\sigma_i \sim$ unif(0, 100)	4229	185	4414
P-spline 2, 8-knot	$\tau_i \sim$ gamma(0.0001, 0.0001)	4118	248	4365
	CAR prior	4154	268	4422
	MCAR prior	4186	339	4525
	$\tau_i \sim$ gamma(0.5, 0.0005)	4110	252	4363
	$\tau_i \sim$ gamma(0.01, 0.01)	4050	301	4350
	$\sigma_i \sim$ unif(0, 100)	3988	344	4332
Smoothing spline 1	$\tau_i \sim$ gamma(0.0001, 0.0001)	4245	173	4419
	$\tau_i \sim$ gamma(0.5, 0.0005)	4256	166	4422
	$\tau_i \sim$ gamma(0.01, 0.01)	4232	181	4413
	$\sigma_i \sim$ unif(0, 100)	4238	178	4417
Smoothing spline 2	$\tau_i \sim$ gamma(0.0001, 0.0001)	4068	270	4337
	$\tau_i \sim$ gamma(0.5, 0.0005)	4175	198	4373
	$\tau_i \sim$ gamma(0.01, 0.01)	3983	334	4316
	$\sigma_i \sim$ unif(0, 100)	3921	380	4301

P-spline 1 (smoothing spline 1), splines with a common smoothing parameter for all LHAs; P-spline 2 (smoothing spline 2), splines with varying smoothing parameters across the LHAs.

over time. Table 3 presents DIC results for the P-spline fitting with 5-, 6-, 7-, and 8-knot, respectively; the 8-knot model was shown to perform the best (best Dbar and DIC). In addition, the 3-level P-spline and smoothing spline models with CAR and MCAR priors for the random intercepts or/and slopes were also explored but showed lack of fit (Table 2).

The 4-level smoothing spline 2 and P-spline 2 fittings produced similar rate and risk estimation and inference; the deviance and DIC slightly favored the prior (17) over the (18) (Table 4). The deviance and DIC results (Table 4), as well as the fitted rate and risk splines, suggested that the penalty splines imposed considerable smoothing and were comparable to the 1-knot B-splines. To illustrate, Table 5 presents the posterior summary statistics for the smoothing spline hyper-hyperparameters.

As shown in Table 2, the 3-level hierarchical formulations of the Bayesian spline fittings under $\tau \sim$ gamma(0.5, 0.0005), $\sigma \sim$ unif(0, 100) indicated considerable sensitivity for both the smoothing splines 2 and the P-splines 2; the smoothing spline 1 and P-spline 1 fittings were comparably robust. In comparison with the P-splines 2, the smoothing splines 2 displayed notably higher

Table 3. Deviance information criterion for 5- to 8-knot P-spline 2 3-stage model fittings

Model	Dbar	pD	DIC
P-spline 2, 5-knot	4176	227	4403
P-spline 2, 6-knot	4166	232	4398
P-spline 2, 7-knot	4140	238	4378
P-spline 2, 8-knot	4118	248	4365

DOI: 10.1002/env

Table 4. Deviance information criterion for the smoothing spline 2 4-stage model fittings

Model	Hyper-hyperprior	Dbar	pD	DIC
Hyperprior (17)	$u \sim$ unif(0,1), $v \sim$ unif(0,1)	4161	209	4370
	$u \sim$ unif(0,10), $v \sim$ unif(0,10)	4164	209	4373
Hyperprior (18)	$u \sim$ unif(0,5), $v \sim$ unif(0,5)	4171	205	4375
	$u \sim$ unif(0,10), $v \sim$ unif(0,10)	4172	205	4377

sensitivity with respect to posterior rate and ratio estimation and inference. As shown in Figure 2, the smoothing spline fitting under the hyperprior $\tau \sim \text{gamma}(0.5, 0.0005)$ was slightly smoother than but very close to the 1-knot B-splines. The (smoothing spline) fitting under the hyperprior $\sigma \sim \text{unif}(0, 100)$, on the other hand, imposed very little smoothing and tended to interpolate the crude rates and risks. For the 4-level models, the resulting rate and risk estimation and inference were relatively robust.

3.3. The BC AME application

Note that adequate smoothing may inform area-specific temporal trends or temporal evolution of the spatial rate and risk patterns. Lack of smoothing could result in estimated rates and risks varying rapidly from 1 year to another, whereas over smoothing could result in estimated rates and risks staying relatively the same over the space and time; both situations often fail to inform and uncover important rate/risk trends and patterns.

In this BC AME application, for example, the fitted 4-knot B-spline model with non-spatial prior (6) (as well as the MCAR prior) informed downward non-linear rate and risk trends for LHAs with a high rate at the early 1990s but upward non-linear rate and risk trends for LHAs with a low rate at the early 1990s. The phenomenon also resulted in changing spatial risk patterns and clustering over time, with the major high and low clusters (formed in the beginning of the 1990s) being persistent only until 1996. By 2000, there were only a very few LHAs with high or low risks. As an example, Figure 3 presents 10 annual risk maps based on the best fitting 4-knot B-splines with the non-spatial prior (6); these maps highlight high/low risk areas and suggest important risk cluster evolution over the 10-year period.

Table 5. Posterior summary statistics of the hyper-hyperparameters s and r, the smoothing spline 2 4-stage model fittings

Model	Parameter	Mean	SD	2.5%	Median	97.5%
Hyperprior (17)						
$s \sim$ unif(0,1), $r \sim$ unif(0,1)	s	0.72	0.15	0.43	0.72	0.98
	r	0.003	0.002	0.000	0.003	0.008
$s \sim$ unif(0,10), $r \sim$ unif(0,10)	s	0.85	0.38	0.44	0.78	1.67
	r	0.004	0.004	0.000	0.003	0.015
Hyperprior (18)						
$s \sim$ unif(0,5), $r \sim$ unif(0,5)	s	0.70	0.32	0.33	0.65	1.34
	r	0.002	0.003	0.000	0.001	0.010
$s \sim$ unif(0,10), $r \sim$ unif(0,10)	s	0.88	1.02	0.33	0.66	3.91
	r	0.004	0.010	0.000	0.002	0.025

DOI: 10.1002/env

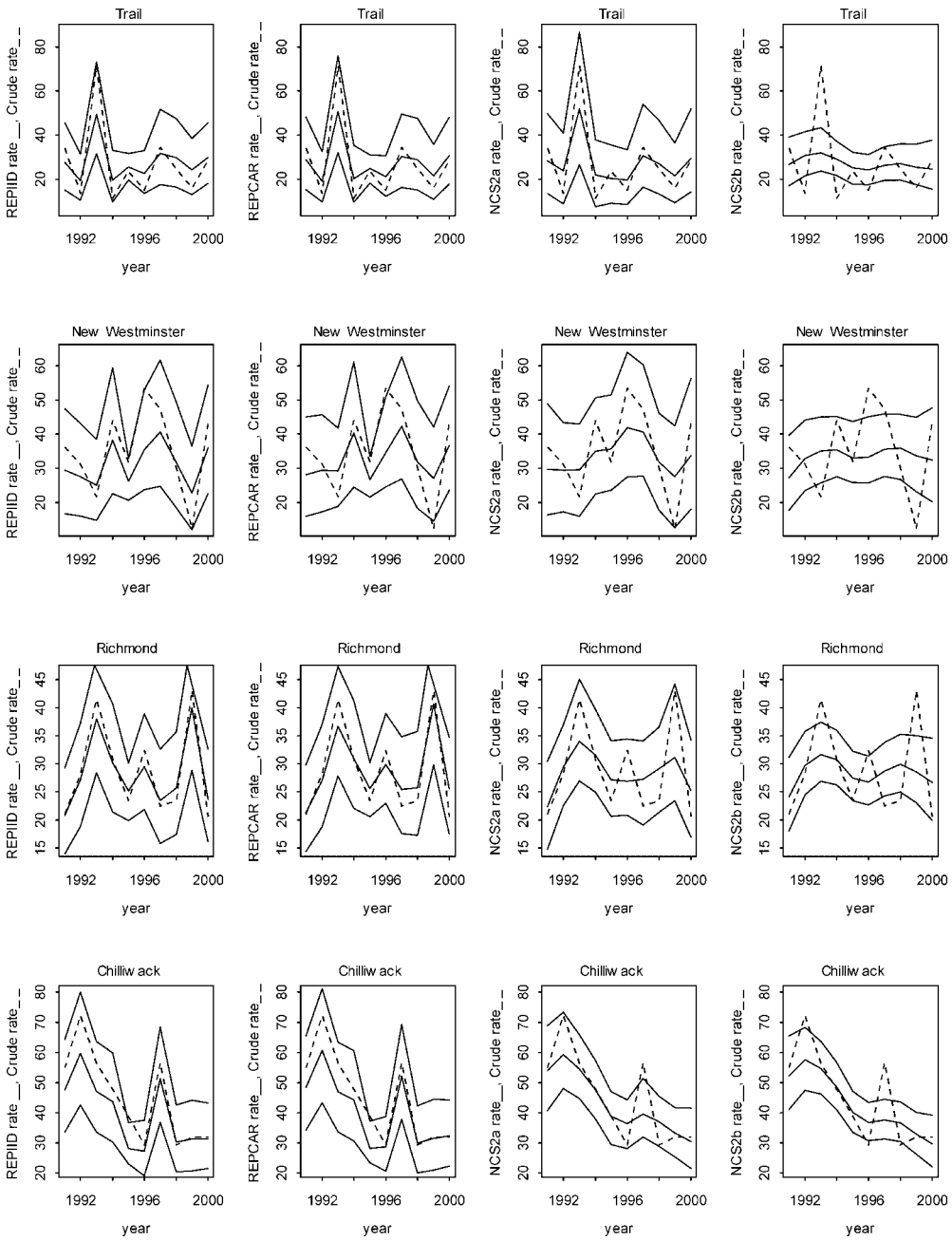

Figure 2. Crude and Bayesian posterior rate estimates for select local health areas: dashed line—crude rates, solid lines—posterior means, 2.5- and 97.5-percentiles. The Bayesian estimates are: REPIID—spatial model (1) with IID prior, REPCAR—spatial model (1) with CAR prior (5), NCS2a—3-level smoothing spline 2 with hyperprior $\sigma_i \sim$ unif(0, 100), NCS2b—3-level smoothing spline 2 with hyperprior $\tau_i \sim$ gamma(0.5, 0.0005)

DOI: 10.1002/env

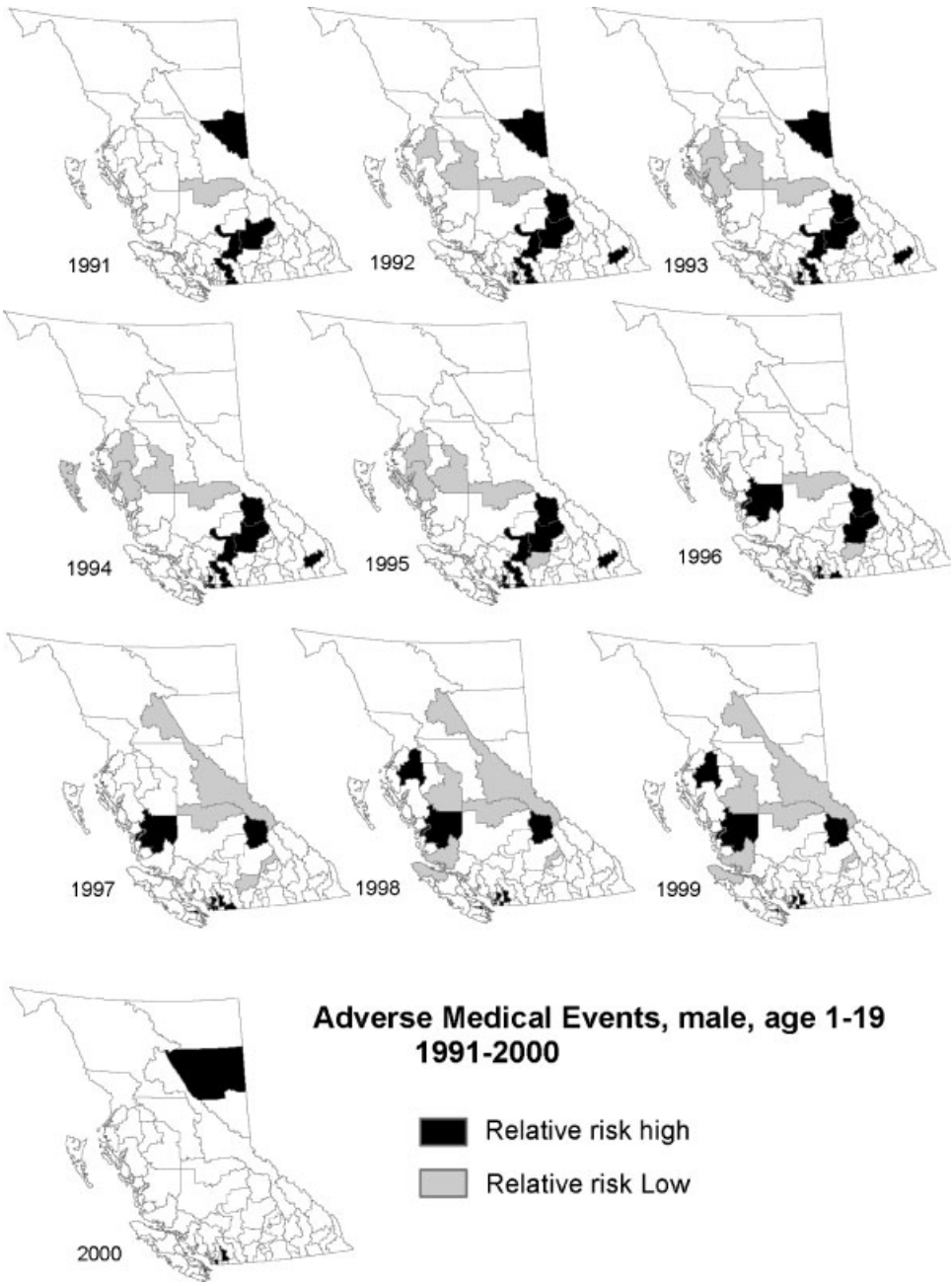

Figure 3. Maps indicating high risk (posterior probability Pr(RR > 1) > 0.975) and low risk (posterior probability Pr(RR < 1) > 0.975) local health areas, 1991–2000, 4-knot B-spline model with non-spatial prior (6)

DOI: 10.1002/env

4. DISCUSSION

The regression spline method presented herein facilitates fitting of the smoothing functions S_0 and RS_is via piecewise cubic polynomials with a relatively small number of interior knots. The key feature of the regression spline method is that the smoothing functions ensemble RS_is is fitted by spatially or randomly varying B-splines. As exemplified herein, the approach allows us to blend together spatial, temporal, and non-linear modeling within one unified Bayesian hierarchical (disease mapping) model framework to enhance modeling of disease and health outcomes data. The method enables a variety of prior options to be considered in the fitting of the random coefficients splines (i.e., the region-specific log relative risks splines), with an objective to realistically model data of complex nature and to explore the interplay between spatial autoregressive local smoothing and temporal spline smoothing for optimal risk ensemble estimation and inference. In this AME analysis, for example, the spatial and non-spatial priors produced comparable posterior inferences for the rates and relative risks. The resulting inferences based on the spatial priors consistently suggested a tendency of neighboring regions to have similar rate and risk trends, which may have important epidemiologic or policy implications.

Since the smoothing spline and P-spline fittings produce posterior rate and risk splines under specific smoothing priors with respect to the penalties (as well as the natural boundary condition for the smoothing splines), the corresponding smoothing within the spatiotemporal disease mapping context have limited flexibility in terms of prior considerations, spatial smoothing, and 'borrowing strength'. We observed that the 3-level smoothing spline and P-spline formulations (likelihood-prior-hyperprior) may subject to considerable Bayesian sensitivity with respect to hyperprior specifications. In the literature, however, such Bayesian sensitivity and its potential influence on curve estimation and inference were often not fully explored. In this study, we illustrated the need for and the importance of examining such issues. The 4-level Bayesian smoothing spline and P-spline fittings suggested considerable Bayesian robustness with respect to rate, risk, and hyperprior estimation and inference.

The DIC assessment offers Bayesian measures of model complexity (pD, which also indicates degree of smoothness) and goodness-of-fit (Dbar or $\bar{D}$). Their sum was proposed to form a deviance information criterion which may be used for comparing models for their overall performances, in terms of balancing complexity and fit (Spiegelhalter *et al*., 2003a). It is noteworthy that in this and similar studies the DIC calculation has shown considerable stability over different MCMC runs, and we observed considerable consistency in the pD-Dbar-DIC assessments with respect to knot and model selections (MacNab, 2007; MacNab and Gustafson, 2007).

For the comparison of the three smoothing methods, the GLMM representation of the smoothing splines and P-splines (with the 2nd-order difference penalty) and the associated Bayesian estimation are presented herein. Such reparameterization also illuminates the ways the penalty splines are fitted: That is, each of the penalized regression splines is converted into a mixed-effects model such that the fixed component reflects a linear trend, a straight line with fixed intercept and slop, and the random component models the curvature of the spline; the associated penalty imposes shrinkage toward the straight line. In addition, the GLMM representation of the penalized splines adapt well within the Bayesian disease mapping context where model-based relative risks estimation and inference have been traditionally formulated within the GLMM framework. Bayesian smoothing splines and P-splines without mixed-effects reparameterization have been studied by many, see for example, Hastie and Tibshirani (2000), Lang and Brezger (2004), and Lambert and Eilers (2005).

To sum up, while the 'fixed-knot' nature of the regression splines is usually seen as a disadvantage, the very nature of 'non-regularization' enabled us to explore prior formulations for modeling spatially or randomly varying local risk trends and for facilitating 'borrowing strength', to develop *parsimonious*

DOI: 10.1002/env

models while tackling complex problems, and to make and communicate objective and interpretable judgments on balancing smoothing and fit. Models with smoothing splines and P-splines often contain large number of unknown parameters; regularized fit may results in a notably reduced 'effective number of free parameters', which is often noted as 'effectively parsimonious'. The methods have often shown to work well in the presence of ample data. However, when the data is limited, as is the case in this study and typically so in disease mapping, the methods may suffer from considerable Bayesian sensitivity and uncertainty. For data of longer time periods, say annual data of 20 years or longer, the performance of penalty splines are likely to improve. Investigations into the need for spatial smoothing (i.e., the need for CAR or MCAR priors), as well as alternative Bayesian spline models for robust risk estimation and inference, are topics of on-going research.

ACKNOWLEDGMENTS

The author thanks the referee for constructive comments which led to considerable improvement of the manuscript. The author thanks the British Columbia Ministry of Health Services and the University of British Columbia's Centre for Health Services and Policy Research for provision of the data. The work was partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC, RG-238660) and the Canadian Institute for Health Research (CIHR, MOP-64458), the British Columbia Michael Smith Foundation for Health Research, and the British Columbia Child and Family Research Institute.

REFERENCES

Böhning D, Dietz E, Schlattmann P. 2000. Space-time mixture modeling of public health data. *Statistics in Medicine* **17/18**: 2333–2344.

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**: 9–25.

Brezger A,d Lang S. 2006. Generalized structured additive regression based on Bayesian P-splines. *Computational statistics & Data Analysis* **50**: 967–991.

Brumback BA, Rice JA. 1998. Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association* **93**(443): 961–994.

Durban M, Harezlak J, Wand MP, Carroll RJ. 2004. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* **24**: 1–24.

Eilers PHC, Marx BD. 1996. Flexible smoothing using B-splines with penalized likelihood (with comments ad rejoinder). *Statistical Science* **11**(2): 89–121.

Friedman JH, Silverman BW. 1989. Flexible parsimonious smoothings and additive modeling. *Technometrics* **31**: 3–39.

Fahrmeir L, Kneib T, Lang S. 2004. Penalized structured additive regressive for space-time data: a Bayesian perspective. *Statistica Sinica* **14**: 731–761.

Green PJ, Silverman BW. 2000. *Nonparametric Regression and Generalize Linear Models*. Chapman & Hall/CRC: New York.

Hastie TJ, Tibshirani RJ. 1990. *Generalized Additive Models*. Chapman & Hall/CRC: New York.

Hastie TJ, Tibshirani RJ. 2000. Bayesian backfitting. *Statistical Science* **15**: 193–223.

Knorr-Held L, Richardson S. 2003. A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics* **52**: 169–183.

Lambert P, Eilers PH. 2005. Bayesian proportional hazards model with time-varying regression coefficients: a penalized Poisson regression approach. *Statistics in Medicine* **24**: 3977–3989.

Lang S, Brezger A. 2004. Bayesian P-splines. *Journal of Computational and Graphic Statistics* **13**: 183–212.

Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel J-F, Bertollini R (eds). 1999. *Disease Mapping and Risk Assessment for Public Health*. John Wiley & Sons: New York.

Leroux BG, Lei X, Breslow N. 1999. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, Halloran ME, Berry D (eds). Springer-Verlag: New York; 135–178.

Lin X, Zhang D. 1999. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society* **61**B: 381–400.

DOI: 10.1002/env

MacNab YC, Dean BC. 2001. Autoregressive spatial smoothing and temporal smoothing B-splines for mapping rates. *Biometrics* **57**(3): 949–956.
MacNab YC, Kmetic A, Gustafson P, Shaps S. 2006. An innovative application of Bayesian disease mapping methods to patient safety research: the Canadian iatrogenic injury study. *Statistics in Medicine* **35**(23): 3960–3980.
MacNab YC, Gustafson P. 2007. Regression B-spline smoothing in Bayesian disease mapping: with an application to patient safety surveillance. *Statistics in Medicine*, DOI: 10.1002/sim.2868.
MacNab YC. 2007. Mapping disability-adjusted life year (DALY): a Bayesian DALY methodology for regional study of disease and injury burden. *Statistics in Medicine*, DOI: 10.1002/sim.2890.
O'Sullivan F. 1986. A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science* **1**: 505–527.
Ruppert D. 2002. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**: 735–757.
Ruppert D, Carroll RJ. 2000. Spatially-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics* **42**: 205–223.
Ruppert D, Wand MP, Carroll RJ. 2003. *Semiparametric Regression*. University Press: Cambridge.
Silverman BW. 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *Journal of the Royal Statistical Society* **B**(47): 1–52.
Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2003a. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B* **64**: 583–639.
Spiegelhalter DJ, Thomas A, Best N, Lunn D. 2003b. *WinBUGS version 1.4 user manual*. MRC Biostatistics Unit, Institute of Public Health, Cambridge University: Cambridge.
Ugarte MD, Ibez B, Militino AF. 2006. Modelling Risks in Disease Mapping. *Statistical Methods in Medical Research* **15**: 21–35.
Verbyla AP, Cullis BR, Kenward MG, Welham SJ. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Applied Statistics* **48**(3): 269–311.
Waller LA, Carlin BP, Xia H, Gelfand AE. 1997. Hierarchical spatio-temporal mapping of disease rates, *Journal of the American Statistical Association* **92**: 607–617.
Wand M. 2000. A comparison of regression spline smoothing procedures. *Computational Statistics* **15**: 443–462.
White IMS, Thompson R, Brotherstone S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Diary Science* **82**: 632–638.
Zhao Y, Staudenmayer J, Coull BA, Wand MP. 2006. General design Bayesian generalized linear mixed models. *Statistical Science* **21(1)**: 35–51.