# Smoothing of land use maps for trend and change detection in urbanization

**Massimo Ventrucci · Daniela Cocchi · Marian Scott**

Received: date / Accepted: date

**Abstract** Urban sprawl and its evolution over relatively short periods of time demands that we develop statistical tools to make best use of the routinely produced land use data from satellites. An efficient smoothing framework to estimate spatial patterns in binary raster maps derived from land use datasets is developed and presented in this paper. The framework is motivated by the need to model urbanization, specifically urban sprawl, and also its temporal evolution. We frame the problem as estimation of a *probability of urbanization* surface and use Bayesian P-splines as the tool of choice. Once such a probability map is produced, with associated uncertainty, we develop exploratory tools to identify regions of significant change across space and time. The proposal is used to study urbanisation and its development around the city of Bologna, Emilia Romagna, Italy, using land use data from the Cartography Archive of Emilia Romagna Region for the period 1976-2008.

**Keywords** binary raster · contour uncertainty · landscape metrics · smoothing · urban sprawl

Massimo Ventrucci
Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy
Tel.: +39 051 2098201
Fax: +39 051 232153
E-mail: massimo.ventrucci@unibo.it

Daniela Cocchi
Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy
Tel.: +39 051 2098201
Fax: +39 051 232153
E-mail: daniela.cocchi@unibo.it

Marian Scott
School of Mathematics and Statistics, College of Science and Engineering, University of Glasgow, Glasgow G12 8QQ, UK
Tel.: +44 01413305125
E-mail: marian.scott@glasgow.ac.uk

## 1 Introduction

Remotely sensed land use data form a powerful resource to study the spatial pattern of many environmental and urban systems as well as monitoring their evolution over the years. Urban planners are interested in investigating patterns of urban development for a number of purposes, including the definition of areas suitable for new urban settlements, the detection of compactly urbanized regions in contrast to sparsely urbanized areas, while ecologists are often interested in fragmentations of natural habitats. In the urban geography literature, the situation when an urban agglomerate develops sparsely is denoted as *urban sprawl* (EEA, 2006). This phenomenon is linked to inefficient urban growth, often characterized by low building and population density over rural areas, and causes increased environmental and infrastructural costs (Borrego et al., 2006; Kelly-Schwartz et al., 2004; Wilson and Chakraborty, 2013). Urban sprawl is also a main driver of landscape fragmentation, land use changes, increase in built-up areas and rapid urban growth (Wei and Ye, 2014). These situations require methods to quantify urban sprawl and to detect changes in the land use pattern across time.

Regarding methods to quantify urban sprawl, research has mostly been focused on indicators of urban intensity and morphology, computed from land use *raster* data (i.e. a map of pixels), at a spatially aggregated level (Angel et al., 2010; Dong and Pengyu, 2014; Jaeger et al., 2010; Torrens, 2008; Tsai, 2005). Altieri et al. (2014) proposed valid indicators to compare urban sprawl levels in different geographical regions. However, indicators offer a spatially aggregated view of the urban sprawl phenomena, missing a fine-scale representation of it. To our knowledge there is no attempt to construct maps showing estimated urban sprawl levels as a continuous surface over space. In this paper, we present a statistical modelling framework to develop this surface and use it to monitor urban sprawl at fine spatial scale and across different times. Our first objective is to efficiently estimate urban intensity as a *probability of urbanization surface*, applying spatial smoothing to land use maps at given time points. A smooth surface aids visualization of large scale trends over space, while surface uncertainty quantification provides inferential tools to detect regions of pixels with increased urbanization over time. Thus, the second goal is to develop suitable exploratory tools to investigate changes across space and time.

There is a vast literature on detecting changes in land use maps, mostly focusing on analyzing remote sensing images across multiple times (Coppin et al., 2004). These methods are good at identifying changes at the pixel scale, which is the scale defined by the image resolution. This fine detail might be computationally demanding in large images, and undesirable when the interest is in detecting changes at a large spatial scale. For instance, Pasanen and Holmström (2015) proposed smoothing of remote sensing images as a more flexible way to detect changes at a larger than a pixel scale. A similar idea is proposed in this paper, where a general smoothing framework based on Bayesian P-splines is developed to estimate large scale trends and changes in land use data. In contrast to traditional methods we use classified land use maps, motivated by the large availability of these types of data which are routinely produced by environmental agencies. Data on land use are released in the form of large vectorial maps, i.e. collections of polygons, produced using remotely sensed images as the primary information source. Vector to raster conversion allows a grid structured format which is easier to handle by modellers and computers. The estimation of spatial trends in raster land use maps calls for efficient smoothing methods for grid structured data to be developed.

Literature on bivariate smoothing offers various proposals, from thin-plate splines to penalized splines including the kriging algorithm used in geostatistics. In general, thin-plate spline is a natural approach for smoothing over a multi-dimensional (e.g. spatial, or spatiotemporal) domain. The disadvantage is in terms of the high computational cost implied by calculating its full-rank smoother matrix. Full rank smoothers involve as many basis functions as data and can be demanding even for moderately large rasters, because of the need to invert a square matrix of dimension given by the number of pixels. The kriging algorithm used in geostatistics also falls in this class; for a discussion of the connections between spline based methods and kriging see Ruppert et al. (2003) ch. 13. This smoother derives from a model assuming a Gaussian Random Field (GRF) for the spatial field underlying the data, which implies again inverting large and dense covariance matrices. In contrast, low-rank smoothers are cheaper in terms of computation, since they use much less basis functions than data with a sensible reduction of the number of parameters to estimate; examples are: penalized splines with truncated power basis functions (Ruppert et al., 2003), thin-plate regression splines (Wood, 2003) and low-rank thin-plate splines built on a radial basis (Crainiceanu et al., 2005). All these low-rank methods imply a non sparse smoother matrix, which may still be quite computationally demanding in cases where a large number of spline coefficients is needed to describe the surface variability. For all these reasons, in this work we focus on a computationally more efficient approach based on a Bayesian version of the P-splines method by Eilers and Marx (1996). This uses a low-rank basis of local (i.e. non zero over a limited domain) B-splines and a random walk prior for the spline coefficients (Lang and Brezger, 2004). A key aspect of this approach is that the posterior distribution of the spline coefficients has a sparse precision (i.e. inverse covariance) matrix, that allows efficient sparse matrix computations and relatively fast Markov Chain Monte Carlo (MCMC) algorithms.

In practice, the proposed framework develops in three steps. The starting point is converting a land use map from vector to raster, which produces a binary grid dataset, with black pixels representing the land use category under study (e.g. urban) and white pixels indicating all the other land use classes. At the second stage, a *smooth* map representing the probability of urbanization surface is obtained by fitting a Bayesian P-spline model to the raster of binary realizations. At the third step, a posterior sample from this probability surface is obtained via Markov Chain Monte Carlo (MCMC) and used to detect relevant changes in the urban process across space and time. In particular, two objectives are addressed in this paper: detecting regions where the probability of urbanization is significantly higher than a threshold; detecting regions where the probability of urbanisation has changed (e.g. increased) over time.

The plan of the paper is as follows. In section 2, the P-spline method is briefly revised with a proposal for modelling binary rasters; details on the MCMC algorithm are left as supplemental material. In section 3, the exploratory tools performing pixel-wise analysis on the estimated surfaces are presented. An application is given in section 4, using rasters of urban residential use over the metropolitan area around Bologna, Italy. The paper closes with a discussion in section 5.

## 2 Smoothing raster data

### 2.1 Rasters

Vectorial land use maps are derived by classifying images collected via remote sensing or aerial photos and consist of a collection of categorical valued polygons, each polygon being assigned to a land use class. A further operation, called *rasterization* is usually undertaken to convert polygons into pixels. The result is a raster map, i.e. a grid structured dataset of categorical response pixels, where each pixel is assigned to a land use class. Land use raster maps need much less memory storage than vectorial data: even though these maps are sometimes large, with thousands of response pixels, the regular grid structure is particularly suitable for quantitative analysis and spatial statistical modelling. Throughout the paper, the focus will be on modelling binary rasters on urbanization, where each pixel is either urban (black) or non urban (white). Nevertheless, the models presented in section 2.4 can be easily adapted to the more general case of binomial response rasters, where, for instance, the proportion of land covered by urbanization is observed at each pixel.

### 2.2 B-spline basis for rasters

Let us assume that we have $n = n_1 n_2$ pixels stored in a raster, i.e. a matrix $\boldsymbol{Y}$ with $n_1$ rows and $n_2$ columns. In the following, the P-spline approach is presented and extended to smoothing of binary raster data in a Bayesian hierarchical modelling framework. The basic P-spline approach for raster data performs non parametric regression on row and column indices of the raster, respectively $\boldsymbol{r} = [1, ..., r, ..., n_1]^\top$ and $\boldsymbol{c} = [1, ..., c, ..., n_2]^\top$, which are considered as covariate vectors. We indicate with $y_{rc}$ the observation at row $r$ and column $c$ (i.e. at pixel $(r, c)$), and with $\mu_{rc}$ its expected value. This expected value can be seen as a latent value to be estimated. When $y_{rc}$ is binary then $\mu_{rc} \in (0, 1)$ is a probability value. The surface is obtained by collecting $\mu_{rc}$ over all pixels in a vector of length $n$ denoted as $\boldsymbol{\mu}$.

Following Eilers et al. (2006), $\boldsymbol{\mu}$ can be modelled as a surface varying smoothly over the raster region by constructing two marginal basis matrices composed by local cubic B-splines functions: $\boldsymbol{R} = [b_1(\boldsymbol{r}), ..., b_{q_r}(\boldsymbol{r})]$, of dimension $n_1 \times q_r$, containing B-splines evaluated at row indices and $\boldsymbol{C} = [b_1(\boldsymbol{c}), ..., b_{q_c}(\boldsymbol{c})]$, of dimension $n_2 \times q_c$, with B-splines evaluated at column indices. The full basis matrix is built by the Kronecker product of the marginal bases, $\boldsymbol{B} = \boldsymbol{C} \otimes \boldsymbol{R}$ of dimension $n \times q$, with $q = q_r q_c$. Columns of $\boldsymbol{B}$ contain cubic bivariate B-splines, centred at knots lying on a regular *knot-grid* which, ideally, underlies the whole raster map. This generates a set of equally spaced bivariate B-splines evaluated at each pixel over the raster map.

The surface is built as a weighted sum of bivariate B-splines,

$$\mu_{rc} = \boldsymbol{B}_{rc}\boldsymbol{\theta} \qquad r = 1, ..., n_1; \quad c = 1, ..., n_2, \tag{1}$$

where notation $\boldsymbol{B}_{rc} = [b_1(c), ..., b_{q_c}(c)] \otimes [b_1(r), ..., b_{q_r}(r)]$ indicates the row entry of matrix $\boldsymbol{B}$ containing the bivariate B-spline basis functions evaluated at pixel $(r, c)$, while $\boldsymbol{\theta}$ is the associated vector (of length $q$) of spline coefficients.

## 2.3 Knot-grid resolution

The choice of $q$, i.e. how fine to choose the knot-grid is critical. Eilers and Marx (1996) suggest the use of a relatively large number of knots such that the surface overfits the data, since surface smoothness is then imposed by a penalty on second order differences between neighbouring spline coefficients. In our large raster dataset a sensible approach seems to take the knot-grid resolution to be much lower than the data resolution. This is useful for two reasons: first, to meet our objective of estimating the large scale spatial pattern removing small scale features and second, to reduce the number of parameters to estimate and speed up computations which otherwise, for very large raster datasets, might even be infeasible. On the other side, if the number of basis functions adopted is too low this will result in a poor representation of the surface variability, i.e. a very smooth probability surface which does not allow features of interest at the desired spatial detail to be detected.

Our suggestion is to set $q$ according to the required spatial detail, by following a geographic criterion, i.e. selecting knots separated by a pre-defined spatial distance. In the application of Section 4, we tried different choices of $q$ by using several knot spacings (1 $km$, 500 $m$, 350 $m$) and display results for the case where 1 knot each 500 $m$ is used; this choice offers a good compromise between computation feasibility and informativeness of the estimated surface in terms of spatial variability of the urban pattern and returned useful maps for visualizing/quantifying urban sprawl.

## 2.4 Bayesian P-splines

The Bayesian P-spline approach proposed by Lang and Brezger (2004) assumes an Intrinsic Gaussian Markov Random Field (IGMRF) prior for the spline coefficients $\boldsymbol{\theta}$, conditional on a precision parameter $\lambda$,

$$\pi(\boldsymbol{\theta}|\lambda) = (2\pi)^{-\mathtt{rank}(\boldsymbol{K})/2}(|\lambda\boldsymbol{K}|^*)^{1/2}\exp\left\{-\frac{\lambda}{2}\boldsymbol{\theta}^\top\boldsymbol{K}\boldsymbol{\theta}\right\} \tag{2}$$

where $|\lambda\boldsymbol{K}|^*$ is the generalized determinant. Equation (2) specifies a multivariate Gaussian distribution for $\boldsymbol{\theta}$, with zero mean vector and rank deficient precision matrix $\boldsymbol{Q} = \lambda\boldsymbol{K}$. Basically, an IGMRF prior induces *smoothness* on the modelled surface by forcing correlation between adjacent spline coefficients through its *structure* matrix $\boldsymbol{K}$. The latter is a sparse known matrix specifying conditional dependencies among spline coefficients. The sparse nature of $\boldsymbol{K}$ is particularly useful to speed up computations and model fitting (Rue, 2001).

In general, conditional dependencies in $\boldsymbol{K}$ are defined on the basis of some pre-defined neighbouring relationship. There are several ways to define the structure of an IGMRF on a regular or irregular lattice; see Rue and Held (2005), Chapter 3. A suitable and computationally efficient way to define an IGMRF for our set of spline coefficients laying on a regular knot-grid is to assume the following Kronecker product form for the structure matrix,

$$\boldsymbol{K} = (\boldsymbol{I}_{q_c} \otimes \boldsymbol{D}_r^\top\boldsymbol{D}_r + \boldsymbol{D}_c^\top\boldsymbol{D}_c \otimes \boldsymbol{I}_{q_r}). \tag{3}$$

In (3), $\boldsymbol{I}_{q_r}$ ($\boldsymbol{I}_{q_c}$) is the identity matrix of size $q_r$ ($q_c$), and $\boldsymbol{D}_r$ ($\boldsymbol{D}_c$) is a matrix which realizes $d$ order differences between neighbouring coefficients along rows (columns) of the knot-grid. Typically, $d$ equal to 1 or 2 is chosen, to penalize first or second order differences, respectively. In the application presented in Section 4, we will use second order differences $d = 2$. The IGMRF structure specified in (3) corresponds to the penalty matrix used in Eilers et al. (2006) for smoothing data on a regular grid via penalized maximum likelihood.

One advantage of using a fully Bayesian approach is that the posterior distribution for the surface, $\pi(\boldsymbol{\mu}|\boldsymbol{y})$, properly incorporates uncertainty about $\lambda$, which is assumed as a random term in the model. As a prior for $\lambda$, Lang and Brezger (2004) suggested a $Gamma(a, b)$, with shape $a = 1$ and rate $b$ taken to be small, as an attempt of non informativeness on the variance $\lambda^{-1}$.

### 2.5 Smoothing binary raster data

We apply Bayesian P-splines to our binary raster data case. The first stage of our model specifies a Binomial likelihood for the data,

$$y_{rc}|\delta, \boldsymbol{\gamma}, \boldsymbol{\theta} \sim Ber(\mu_{rc}) \tag{4}$$

$$g(\mu_{rc}) = \eta_{rc} = \delta + \boldsymbol{x}_{rc}^\top\boldsymbol{\gamma} + \boldsymbol{B}_{rc}\boldsymbol{\theta} \tag{5}$$

In (4) it is assumed that observations $y_{rc}$ are conditionally independent Bernoulli variables with parameter $\mu_{rc}$, given the parameters specified in the linear predictor (5). The latter is the sum of some fixed effects and a P-spline component $\boldsymbol{B}_{rc}\boldsymbol{\theta}$, specified as in (1). Vector $\boldsymbol{x}_{rc} = [x_{rc}, ..., x_{p,rc}]^\top$ contains $p$ covariates observed at pixel $(r, c)$, $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_p]^\top$ is the vector of the associated slopes and

$\delta$ is an overall intercept. The link function $g$ is assumed as the inverse cumulative distribution function (cdf) of the standard normal distribution, giving a probit regression model. Due to the binary nature of the data, the latent value $\mu_{rc}$ expresses the probability of urbanization evaluated at pixel $(r, c)$; the collection of these values over all pixels gives the smooth probability surface $\boldsymbol{\mu}$. The probability surface expressed in the scale of the linear predictor is $\boldsymbol{\eta}$.

Note that the P-spline component $\boldsymbol{B\theta}$ in (5) captures large scale spatial variability. This is suitable for our purpose of detecting large scale patterns. The small scale variability present in the data is absorbed in the residuals. Ideally, the latter should be spatially unstructured, even though in some datasets residuals at neighbouring pixels may be correlated. Accounting for this extra variation is important, especially when the goal is estimation of the fixed effects $\boldsymbol{\gamma}$ or predictions at new spatial locations and time. One way to model small scale extra variability is to add a set of spatial effects in (5), one for each pixel, with an IGMRF prior for them. A similar approach has been proposed in Lee and Durbán (2009) in a mixed model setting, using restricted maximum likelihood inference. As pointed out by Lee and Durbán (2009), models of this type may present identifiability issues: in some situations, the large scale and small scale sources of variation may be poorly identifiable based on the observed data. The Bayesian paradigm may offer a convenient workaround to the identifiability issue, through the use of informative priors that constrain the degrees of freedom assigned to each component (Ventrucci and Rue, 2015). Future extension of the framework presented here for modelling land use raster will investigate suitable priors for cases where large and small scale spatial effects are needed.

When land use raster data are available at different time points $t = 1, ..., T$, (e.g., different years) one interest is to highlight regions of the probability surface where a significant change over time is noticed. To detect spatial regions where a temporal change occurred, we modify model (5) by allowing a set of spline coefficients for each time point, $\boldsymbol{\theta}_t$. Our model for temporal raster data is:

$$\begin{aligned} y_{rct}|\delta_t, \boldsymbol{\gamma}, \boldsymbol{\theta}_t &\sim Ber(\mu_{rct}) \\ g(\mu_{rct}) &= \delta_t + \boldsymbol{x}_{rct}^{\mathsf{T}}\boldsymbol{\gamma} + \boldsymbol{B}_{rct}\boldsymbol{\theta}_t \qquad t = 1, ..., T; \end{aligned} \tag{6}$$

where $\mu_{rct}$ is the probability surface at pixel $(r, c)$ and time $t$, $\boldsymbol{x}_{rct}$ is a vector of covariates observed at pixel $(r, c)$ and time $t$, $\boldsymbol{B}_{rct} = [b_{t,1}(c), ..., b_{t,q_c}(c)] \otimes [b_{t,1}(r), ..., b_{t,q_r}(r)]$ is the row entry of the (time-specific) basis matrix $\boldsymbol{B}_t = \boldsymbol{C} \otimes \boldsymbol{R}$, containing the B-splines evaluated at pixel $(r, c)$ and time $t$. Regarding the unknown parameters in the linear predictor (6), $\delta_t$ is a time specific intercept which capture variations in the average level of urbanization at different times, $\boldsymbol{\gamma}$ is a vector of covariate effects and $\boldsymbol{\theta}_t$ is a vector of length $q$ containing the spline coefficients that determine the surface at time $t$. Note that, for simplicity, we assume $\boldsymbol{\gamma}$ to be constant over time, though extension to time-specific slopes is straightforward. At the second stage, we specify an IGMRF prior as in (2), with precision $\boldsymbol{Q}_t = \lambda_t \boldsymbol{K}$ for each set of coefficients $\boldsymbol{\theta}_t$, $t = 1, ..., T$. Note that

$\lambda_t$ depends on time, giving a flexible model where the degree of smoothness of the fitted surface at a certain time can be different from the smoothness of the surface at another time. At the third stage of the hierarchy, the model is completed by specifying independent diffuse normal priors with mean zero and small precision (e.g. $10^{-5}$) for the fixed effects, i.e. $\delta_t$, $t = 1, ..., T$, and $\boldsymbol{\gamma}$, and a $Gamma(a = 1, b = 5 \cdot 10^{-5})$ for each IGMRF precision parameter $\lambda_t$, $t = 1, ..., T$.

### 2.6 Model fitting

The posterior distribution for the probability surface $\pi(\boldsymbol{\mu}|\boldsymbol{y})$ in models (5) or (6) is intractable. We use an MCMC Gibbs sampler based on the augmented approach by Albert and Chib (1993) to build a sample from the posterior; for details see the supplemental material. Though MCMC typically requires time consuming iterative computations, there are some practical advantages for using simulation based methods in our raster data case. First, we only need to store in memory an MCMC sample (at convergence) from the joint posterior of the spline coefficients $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and fixed effects $\pi(\boldsymbol{\gamma}|\boldsymbol{y})$, then by combining them, a sample from the posterior surface $\pi(\boldsymbol{\mu}|\boldsymbol{y})$, or $\pi(\boldsymbol{\eta}|\boldsymbol{y})$, is easily obtained for further analysis. Second, the posterior surface distribution properly incorporates uncertainty about $\lambda$. Finally, the detection of significant features across the probability surface can be performed on the basis of a large MCMC sample from $\pi(\boldsymbol{\mu}|\boldsymbol{y})$, which is discussed next.

## 3 Detecting changes across space and time

Formal tests of hypotheses for comparing nonparametric surfaces were introduced in Bowman (2006), where two types of procedures are described: a global test to check the assumption of nonlinearity, based on an F-statistic (i.e. a generalization of an anova-type test) and a local point-wise test to detect the pixels where the evidence for non linearity is strongest. The procedures proposed in sections 3.1 and 3.2 are close in spirit to the local test in Bowman (2006). The latter is based on a t-statistic of the type $(\hat{\mu}_{rc,t_1} - \hat{\mu}_{rc,t_0})/st.dev.(\hat{\mu}_{rc,t_1} - \hat{\mu}_{rc,t_0})$, where $\hat{\mu}_{rc,t_i}$ is the estimated surface at pixel $(rc)$ and time $t_i$. This t-statistic quantifies, in units of standard error, the difference between estimates at $t_0$ and $t_1$, in a given pixel $(r, c)$. Note that, similarly one could test the difference between the surface at a given time and a constant surface at a threshold value, say $th$, using a t-statistics like $(\hat{\mu}_{rc,t_i} - th)/st.dev.(\hat{\mu}_{rc,t_i})$. Similar tests have been used in the analysis of brain imaging data via smoothing techniques (Ventrucci et al., 2011). For the local t-statistic, Bowman (2006) describes computation of a p-value using quadratic forms; in some cases, a p-value can be derived from the standard normal distribution under the assumption of asymptotic normality for $\hat{\mu}_{rc,t_i}$.

Following our Bayesian analysis, procedures for pixel-wise surface comparisons can be developed by analysing the marginal posterior distribution at each pixel $(r, c)$ and time $t$. A sample from these marginals can be obtained for free as a by-product of the MCMC algorithm adopted to fit the model. After convergence of the MCMC, we collect a sample of 1000 realizations from $\pi(\mu_{rct}|\boldsymbol{y})$, $r = 1, ..., n_1$, $c = 1, ..., n_2$, $t = 1, ..., T$ and compute empirical summaries, such as:

- the sample mean, denoted as $\hat{\mu}_{rct}$, which gives the fitted value (in the response scale) from our model at a given pixel and time;
- the $\alpha$ sample quantile of the empirical distribution for the probability surface, denoted as $\hat{\mu}_{rct,\alpha}$; the quantile of $\pi(\mu_{rct})$, at probability $\alpha$, is defined as the minimum value of $\mu_{rct}$ that realizes $F(\mu_{rct}) \geq \alpha$, with $F(\cdot)$ the cdf of $\pi(\mu_{rct})$.

Empirical quantiles allows calculation of a pixel-wise credible interval, at level $100(1-\alpha)\%$, as $(\hat{\mu}_{rct,\alpha/2}, \hat{\mu}_{rct,1-\alpha/2})$. An intuitive rule to decide whether or not a pixel falls inside an uncertainty region (i.e. a region likely affected by sprawl) on the basis of credible intervals for $\hat{\mu}_{rct}$ will be described in section 3.1. A rule to decide whether or not a pixel falls inside an increased probability region (i.e. an area characterized by significantly growing urbanization) on the basis of credible intervals for $\hat{\eta}_{rct}$ will be proposed in section 3.2.

We would like to point out that the procedures outlined in the following two sections do not represent a Bayesian formal testing procedure. For this, one would need calculation of the Bayes factor at each pixel, to compare the marginal likelihood under the null and alternative models, which is a computationally intensive task for non Gaussian likelihoods (Frühwirth-Schnatter and Wagner, 2008). However, we believe that the methods we introduce below provide intuitive means of quantifying the information present in the data about the underlying spatial patterns. This will assist in monitoring of urban sprawl at a given time, and changes in urbanization across time.

### 3.1 Monitoring urban sprawl at a given time

In a situation where a detailed definition of urban sprawl is lacking and sprawl is measured in terms of urban size and morphology (Jaeger et al., 2010), the development of statistical methods for the identification of compactly urbanized areas as opposed to *sprawling* regions is important for urban planning purposes. For instance, urban planners may be interested in exploratory tools to identify regions with a probability of urbanization exceeding a threshold, say $th \in (0, 1)$. The user may choose the most appropriate set of thresholds to explore patterns at several urban intensity levels. This can help in identifying homogeneous areas within a city characterized by different levels of urbanization. To this aim, we propose drawing contour lines at level $th$ and quantifying their uncertainty; we denote this an *uncertainty region* at level $th$. From an urban planning point of view, locating uncertainty regions helps in detecting areas characterized by non compact patterns, i.e. urban sprawl.

**Table 1** Rule to define the uncertainty region for a contour line at level $th$, using a credible level equal to $(1-\alpha)\%$.

| Pixel at location $(r, c)$ and time $t$ lays inside: | Criterion |
|---|---|
| highly urbanized region (at level $th$) | $\hat{\mu}_{rct,\alpha/2} \geq th$ |
| limited urbanization region (at level $th$) | $\hat{\mu}_{rct,1-\alpha/2} \leq th$ |
| uncertainty region (at level $th$) | $\hat{\mu}_{rct,\alpha/2} < th$ and $\hat{\mu}_{rct,1-\alpha/2} > th$ |

Given a threshold specifying an urbanization level $th$, let an uncertainty region be a collection of pixels where the probability of urbanization is neither significantly higher nor lower than $th$. Pixel-wise credible intervals allow practical and computationally efficient rules for selecting uncertainty regions. Given a credible level $100(1-\alpha)\%$, say equal to 95% (i.e. $\alpha = 0.05$), an equal-tails credible interval for $\hat{\mu}_{rct}$ is constructed by taking the quantiles $\hat{\mu}_{rct,0.025}$ (i.e. $\hat{\mu}_{rct,\alpha/2}$) and $\hat{\mu}_{rct,0.975}$ (i.e. $\hat{\mu}_{rct,1-\alpha/2}$) as the lower and upper limits, respectively. A rule to assign pixels to *highly urbanized*, *limited urbanization* or *uncertainty region* at levelt $th$ is outlined in Table 1. According to this, a given pixel is assigned to the highly urbanized region when the lower credible limit is above $th$, i.e. $\hat{\mu}_{rct,\alpha/2} \geq th$. Analogously, a pixel is assigned to the limited urbanized area when the upper credible limit is below $th$, i.e. $\hat{\mu}_{rct,1-\alpha/2} \leq th$. Finally, when none of the aforementioned options is the case, a pixel is assigned to the uncertainty region. In this way, the statistical detection of urban sprawl is obtained by the joint exploration of contour lines and the definition of uncertainty regions.

As an alternative rule one could assign a pixel to the highly urbanized area when $Pr(\mu_{rct} \geq th|\boldsymbol{y})$ is at least $1-\alpha/2$. Choosing $\alpha = 0.05$ may result in a overly restrictive criteria, very conservative w.r.t. the null model, indicating that the posterior mean $\hat{\mu}_{rct}$ corresponds to $th$. Such a restrictive rule requires at least 95% (posterior) probability mass beyond $th$. However, note that the simulation based approach presented here is very flexible, because based on an MCMC sample one can easily recompute the selection criteria setting a different $\alpha$ to achieve the desired level of conservativeness.

## 3.2 Monitoring changes in urbanization across time

The rationale behind assuming a separate smooth probability surface at each time in model (6) is to investigate smooth regions characterized by a change in the probability of urbanization, between two arbitrary time points. We denote this area as *changed*, or *increased* probability region. For instance, an urban planner may want to investigate the location of increased probability regions between a current time $t_1$ w.r.t. a past time $t_0$, to track the urban areas which have developed more during that period of time. In order to track these changes at a high spatial detail, a relevant number of basis functions needs to be specified when building the basis matrix $\boldsymbol{B}$. However, when the interest is in detecting changes occurring at a fairly large spatial scale, a moderate number of basis functions is sufficient, which also helps in reducing the computational cost of model fitting.

For monitoring large scale changes in urbanization between $t_0$ and $t_1$, our proposal is to compare the two marginal posterior for the surface at $t_0$ and $t_1$, in a pixel-wise manner and work out rules to select changed, or increased, pixels. The increased probability region is defined as the collection of pixels showing a significant increase. We firstly present a rule that compares marginals for the surface expressed in the scale of the linear predictor, where the distributions are more symmetric. We let the user specify a desired credible level $100(1-\alpha)\%$, then a pixel is assigned to the increased probability region when $\hat{\eta}_{rct_1,\alpha/2} \geq \hat{\eta}_{rct_0,1-\alpha/2}$, i.e. when the (equal-tails) credible intervals at level $100(1-\alpha)\%$ for $\hat{\eta}_{rct_0}$ and $\hat{\eta}_{rct_1}$ do not overlap. In Figure 1, see an example where this criteria is applied to two empirical marginals, $\pi(\eta_{rct_1}|\boldsymbol{y})$ (blue) and $\pi(\eta_{rct_0}|\boldsymbol{y})$ (red), referred to times $t_1$ and $t_0$, respectively. The sample quantiles involved in making the decision are also displayed in Figure 1, these are: $\hat{\eta}_{rct_1,\alpha/2}$ (blue solid line) and $\hat{\eta}_{rct_0,1-\alpha/2}$ (red dashed line). Note that, in this particular case the decision obtained on the basis of credible intervals at level 95% (i.e. $\alpha = 0.05$, left panel) is different from that obtained at 90% (i.e. $\alpha = 0.10$, right panel). In principle, several rules with arbitrary level of conservativeness can be created by changing $\alpha$. For instance, choosing a credible level of 80% (i.e. $\alpha = 0.2$) will return a less restrictive criteria and a larger increased probability region, as we will se in the application in section 4.

Another intuitive rule to define the increased probability region may select pixels such that $Pr(\eta_{rct_1} > \eta_{rct_0}|\boldsymbol{y}) \geq 1-\alpha$. This rule does not focus on when credible intervals do not overlap, but only on the probability than the surface at time $t_1$ is higher than the one at time $t_0$. For a given $\alpha$, this rule is less conservative (w.r.t. the null model indicating no change between $t_0$ and $t_1$) than the criterion presented above. However, analogously to the rule presented first, choosing $\alpha = 0.05$ may be overly conservative, because only the pixels showing a 95% increase in the probability of urbanization will be selected; the user may then set $\alpha$ to larger values than 0.05, to select pixels with 90% (i.e. $\alpha = 0.10$) or 80% (i.e. $\alpha = 0.20$) increase.

We have seen that several selection rules with different level of conservativeness may be designed to the purposes of monitoring urban sprawl at a given time and monitoring changes across time. Importantly, all these criteria are built on suitable summaries from the marginal posteriors, either in the response or linear predictor scales, which can be computed at no additional cost, as a by-product of the MCMC methods adopted to fit the model.

## 4 Application

### 4.1 Data description and goals

The proposed framework is illustrated on land use maps taken from the city of Bologna, in the Emilia Romagna Region of Italy. The aim is to study the

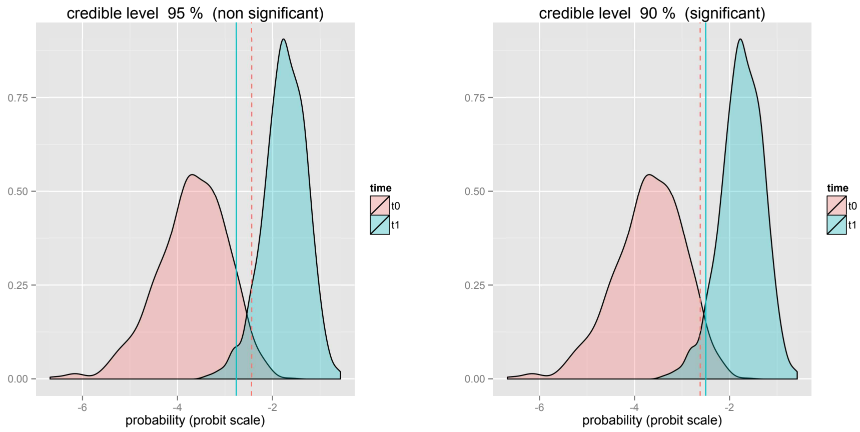


**Fig. 1** Graphical representation of the criterion described in section 3.2 to select increased probability pixels, using a credible level equal to 95% (left panel) and 90% (right panel). In each panel the empirical posterior distribution (i.e. a histogram from a large MCMC sample) of the probability surface (expressed in the probit scale), evaluated at a given pixel, for $t_1$ (red) and $t_0$ (blue) are displayed. The vertical solid blue line and the vertical dashed red line indicate the sample quantile $\hat{\eta}_{rc,t_1,\alpha/2}$ and $\hat{\eta}_{rc,t_0,1-\alpha/2}$. Note that, when 90% credible level is set, the pixel is selected and, hence, assigned to the increased probability region (i.e. $\hat{\eta}_{rc,t_1,\alpha/2} > \hat{\eta}_{rc,t_0,1-\alpha/2}$), whereas, using 95% credibel level, the pixel is not selected.

pattern of residential urban use in a subregion of Bologna province. In Figure 2, the urban residential pattern observed in 2008 for the sixty municipalities (identified by grey lines) included in the province of Bologna is shown as black pixels superimposed on terrain elevation data, displayed on a colour scale. The red box in Figure 2 shows the selected study region which includes the *metropolitan belt* region, an administrative area given by the union of all municipalities sharing borders with Bologna city, which is of particular interest for urban planning purposes and the focus of our application.

Vectorial land use maps referring to four different time points (years 1976, 1994, 2003 and 2008) have been taken from the Cartography Archive of the Emilia Romagna Region. They consist of a collection of polygons to which a category of land use has been assigned on the basis of the standard protocol defined by CORINE Land Cover programme (EEA, 1994). Data were converted from polygons to raster using the `R` package `raster` (Hijmans, 2013), to produce the residential use binary pattern. In terms of resolution, each pixel in the raster map has side length of around 170 $m$ and area of around 3 hectares, similarly to rasters produced by the Environmental European Agency displaying the Urban Morphological Zones (UMZ) over Europe and recommended for studying urban sprawl (EEA, 2011): each UMZ pixel area is 1 hectare, in the highest resolution case and 6.25 hectares, in the lowest. The study region considered has a total area of around 1380 $Km^2$, resulting in a raster matrix with $n_1 = n_2 = 216$ at each time.

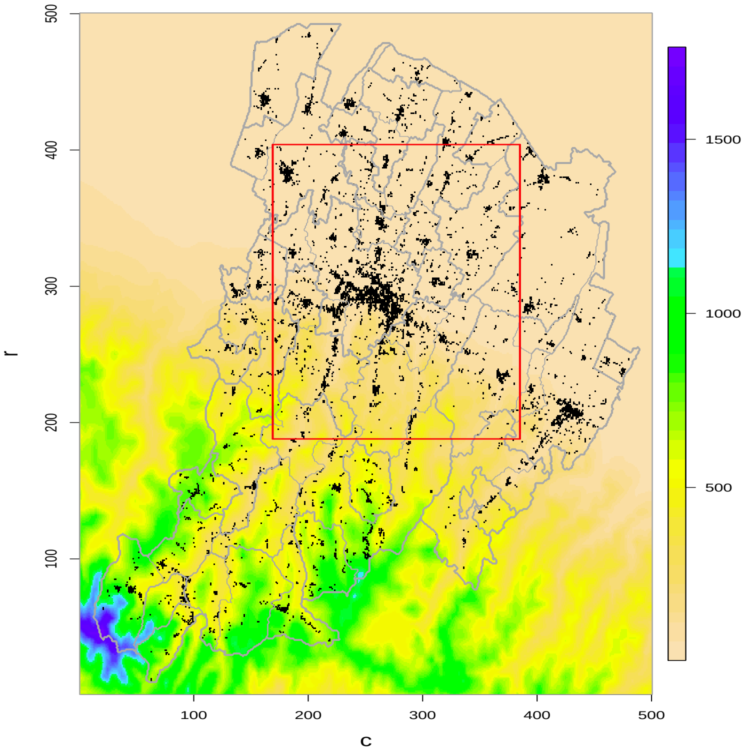


**Fig. 2** A map of the Bologna province consisting of sixty municipalities (boundaries as grey lines). The urban residential pattern (black pixels) observed in 2008 is superimposed on terrain elevation data (expressed in a colour scale). The red box shows the selected study region, which includes the metropolitan area around Bologna city.

The analysis of this dataset has to deal with issues about the classification method, since the standard adopted for assigning polygons to land use classes has slightly changed between $\{1976, 1994\}$ and $\{2003, 2008\}$; polygonal data for 2003 and 2008 have been created using more than 80 land use categories, while data from 1976 and 1994 are based on a less detailed classification. The framework proposed in this paper is able to overcome these problems by estimating the large scale pattern of urbanization, removing small scale structures which can be due, first, to land use misclassification incurred in the rasterization process and second, to heterogeneities in the classification standard adopted.

The binary raster maps referred to the assumed study region at different years are shown in Figure 3. From visually inspecting these maps, we see that changes in size and fragmentation might have taken place in the residential pattern of Bologna during the last four decades. The most prominent feature in the spatial pattern is the polycentric shape of the metropolitan area: the main black patch in the middle represents residential urbanization inside the city of Bologna, with the surrounding smaller agglomerates denoting the centres of neighbouring municipalities. The patterns referred to 2003 and 2008 are quite similar, but there seems to be some evidence of an increase in the intensity of urbanization in some regions, between 1976 and 2008. Also, a general increase in the level of urban sprawl and fragmentation seems to have occurred over time.

## 4.2 Results

In order to investigate the spatial extent of urban sprawl and its changes across time, we fit model (6) to our raster dataset and illustrate the methods proposed in Section 3. Covariates include a time dependent intercept, capturing the overall-space probability of urbanization at each time, and terrain elevation. The probability of urbanization surface was modelled with Bayesian P-splines as described in Section 2.3. To check how well MCMC computational time scales to changing knot-grid resolution, we ran model (6) choosing knot-spacing equal to 1 $km$ ($q = 1089$), 500 $m$ ($q = 3969$) and 350 $m$ ($q = 10609$), approximately, along both rows and columns. The Gibbs sampler took around 3, 4 and 6.5 minutes to run one hundred iterations for $q = 1089$, $q = 3969$ and $q = 10609$, respectively, using an Intel(R) Core(TM) i7 CPU 2.00GHz. Below, results are reported for $q = 3969$, thus the focus is on changes operating at a spatial scale not lower than 500 $m$.

Next we show an application of the tools described in Sections 3.1 and 3.2 to analyse changes across space and time. Figure 4 shows contour analysis maps for years 1976 (left) and 2008 (right), with red contour lines at level $th$ equal to 0.7 (top panels), 0.5 (central panels) and 0.2 (bottom panels). Contour uncertainty regions (blue shadowed areas) have been calculated applying the rule reported in Table 1 at credible level equal to 95%. In each panel of Figure 4, contours and uncertainty regions are superimposed to the estimated probability surface, indicated in a grey color scale. Looking at both 1976 and 2008 estimates, we note that uncertainty regions are typically located at the boundary or in proximity of the core of urban agglomerates, where urban sprawl is usually expected.

Different levels of the threshold $th$ are used in Figure 4 in an exploratory analysis aimed to highlight several urban sprawl patterns, occurring at different urban intensity levels. In the top panels, for instance, areas with estimated probability higher than $th = 0.7$ are displayed depicting quite clearly the historical residential pattern of the city, which is a large scale feature of the urban pattern. In the bottom panel the contour lines at level $th = 0.2$ can highlight multiple residential urban agglomerates of smaller extension w.r.t. the historical residential area. By comparing the left and right hand panels of Figure 4, we see that uncertainty regions are sprawling and fragmenting more in 2008 than in 1976, for any intensity level $th$. This shows that the *leap frog* type of sprawling in the metropolitan area around Bologna has increased in the last four decades.

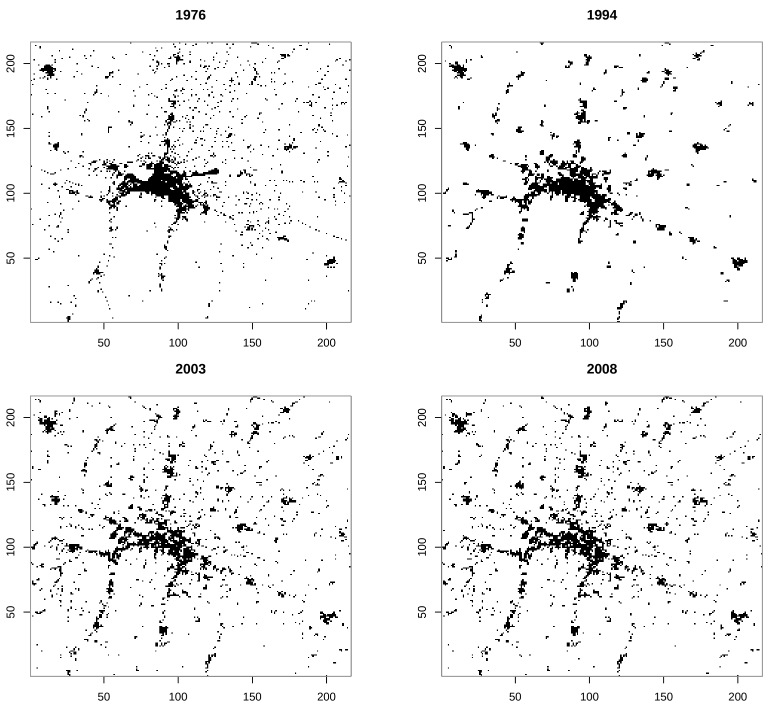


**Fig. 3** The urban residential land use pattern and its evolution over years $\{1976, 1994, 2003, 2008\}$ in the study region identified by the red box in Figure 2, i.e. the metropolitan area around Bologna city. The polycentric nature of the metropolitan area is evident from the maps. The central urban agglomerate shows the residential pattern in the city of Bologna, while the smaller urban patches placed around it represent neighouring municipalities.

Figure 5 focuses on the detection of increased probability regions to monitor changes between $t_0 = 1976$ and $t_1 = 2008$. Each panel displays the residential urban pixels, for both 1976 (black) and 2008 (red), together with the increased probability regions (grey shadowed areas). The increased probability regions are identified using the first rule presented in section 3.2, which compares pixel-wise credible intervals in the probit scale. Again, we use this tool for exploratory purposes, considering credible intervals at different levels, namely 95% (i.e. $\alpha = 0.05$), 80% (i.e. $\alpha = 0.2$) and 60% (i.e. $\alpha = 0.4$), respectively, from left to right. As expected, the higher $\alpha$ the larger the increased probability region selected, as a result of applying a less restrictive rule. We also applied the rule given at the end of Section 3.2 looking at the posterior probability for the surface at time $t_1$ being higher than the one at time $t_0$ and obtained similar results.

In conclusion, it is worth noting that regardless the level of conservativeness specified, the detected increased probability regions match well the areas with

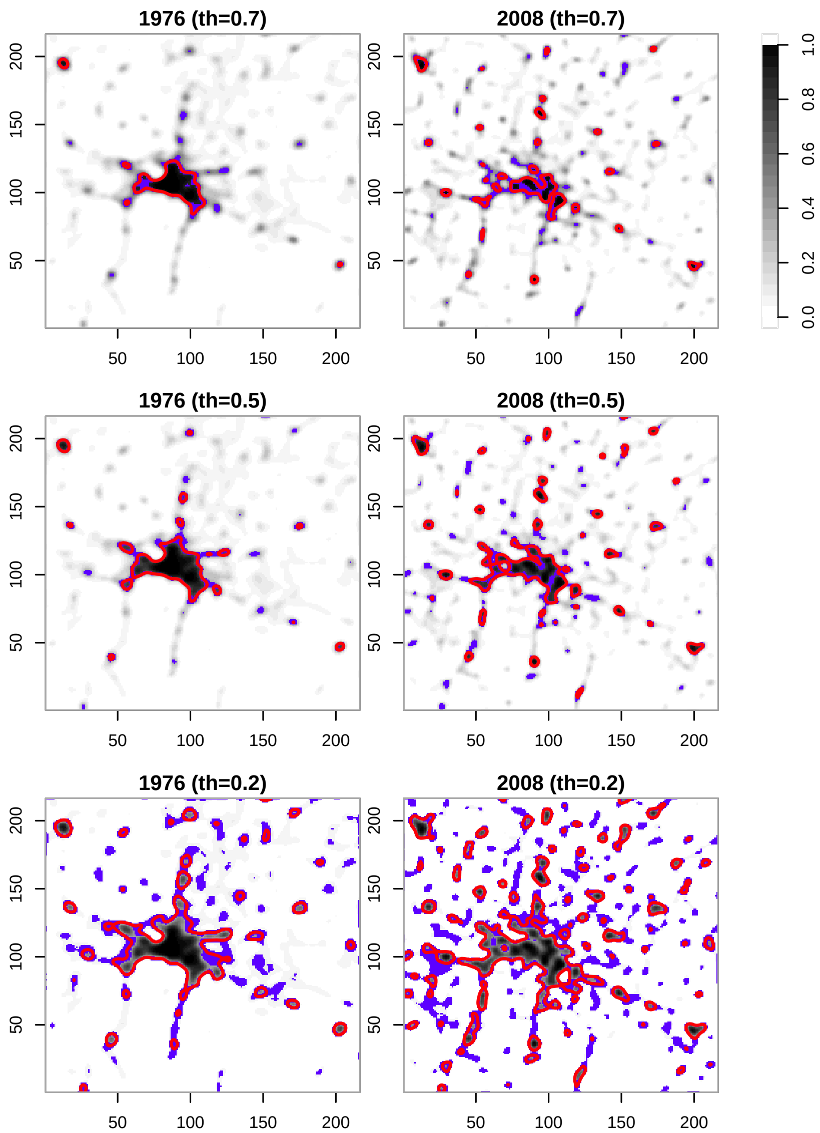


**Fig. 4** Contour analysis maps. Each panel displays the estimated probability surface in grey colors, for $t = 1976$ (left hand panels) and $t = 2008$ (right hand panels), with red contour lines at levels $th = 0.7$ (top panels), $th = 0.5$ (central panels) and $th = 0.2$ (bottom panels). Uncertainty regions for the contour lines are displayed as blue shadowed areas, at 95% credible level. Uncertainty regions are typically located at the boundary of urban agglomerates, where urban sprawl is usually expected. At any $th$ level, we see that uncertainty regions are more extended and fragmented in 2008 than in 1976, as an indication that urban sprawl in the metropolitan area around Bologna has increased in the last four decades.

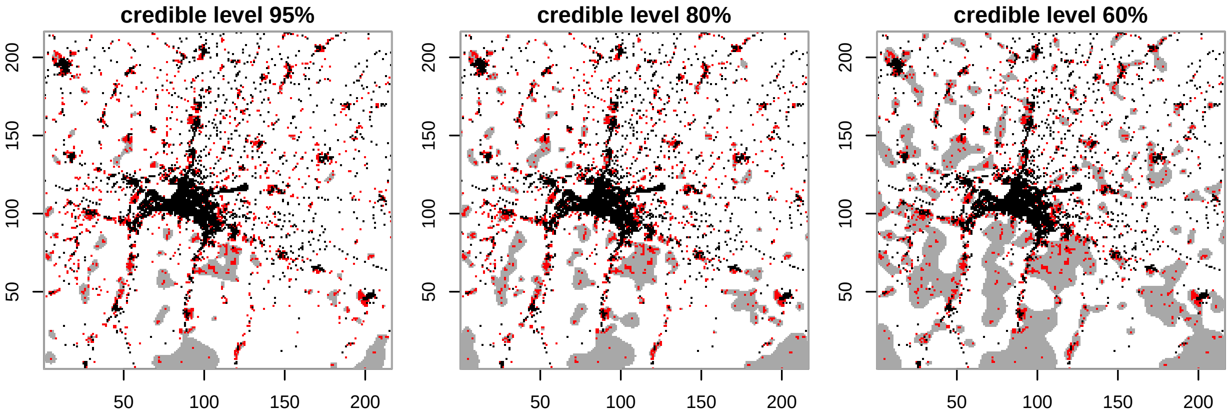


**Fig. 5** Surface comparison maps: 2008 *versus* 1976. Grey shadowed areas indicate increased probability regions in 2008 with respect to 1976, with an estimated increase of at least 95% (left), 80% (central) and 60% (right). In each panel, urban pixel referred to year 1976 (2008) are shown in black (red). Note, increased probability regions detect regions of change, rather than simply identifying the location of new urbanized pixel.

*new* urbanization. The pixel-wise procedure proposed in Section 3.2 seems effective in identifying regions within the metropolitan area around Bologna where land use exploitation for residential purposes has been more intensive over the last forty years.

## 5 Discussion

Relevant changes in the urban phenomena across space are not easily identifiable by visually inspecting raster maps, as the large scale spatial pattern is typically masked by both small scale structures and random noise. There is a vast literature on statistical detection of significant patterns in spatial data, such as spatial hot-spots and clusters (Duczmal et al., 2010; Lawson, 2010; Patil et al., 2010). These methods often use different techniques to achieve similar goals to those pursued in this paper, and are applied in several fields from ecology to epidemiology. In the applied context of this work, a modelled representation of the urban spatial pattern helps in detecting significant changes over a wide urban agglomerate, such as a metropolitan area, discounting changes occurring at a small scale which are more likely attributable to local features. Working on a smooth representation of the raster map, it is easier to detect spatially structured changes over time. This has been done by comparing surfaces at two different times, which is a practical solution to the problem of identifying large spatial regions changing across time. Note that this is different from the problem of detecting changes at the pixel level, i.e. the *new* urbanized pixels. In this sense, the methods proposed in this work may be seen as alternative, or complementary, to traditional change detection methods.

On the computational side, P-spline is a stable and efficient method for smoothing, which is a crucial aspect when analyzing large rasters. This arises for basically two reasons. First, P-spline smoothing implements low-rank bases of spatial B-spline functions, hence the number of parameters to estimate is much lower than the number of pixels composing the surface. Second, the B-splines are local functions, i.e. non zero in a limited spatial domain, thus sparse matrix computation can be adopted which speeds up sampling from the full conditional distribution of the surface coefficients, needed at each MCMC iteration.

Surface smoothness depends to some extent on the number of basis functions adopted. In cases where the observed pattern is the result of several processes going on simultaneously at different spatial levels, a possible strategy is to focus on a scale of interest and utilize smoothing as a mean of removing variation at smaller scales; this is the approach used here, where a geographic criterion to define the resolution of the knot-grid is adopted. Alternatively, several smoothness levels can be applied with the aim of detecting features at different spatial resolutions.

Working on a raster representation of widely available land use maps allowed us to build a general framework, applicable by practitioners of environmental agencies, for instance. The advantage of working with publicly available data comes to a price in terms of inability to account for errors in data preprocessing, e.g. classification of land use polygons (Foody, 2002) and polygon-to-raster conversion (Lechner et al., 2009). Accuracy of land use classifications algorithms is very important when the target is to detect land cover changes at a very fine spatial scale. For our purpose of modelling large scale spatial trends, the choice of the polygon-to-raster conversion criterion seems a much more critical issue. Errors due to polygon-to-raster conversion might be sensibly reduced by using more detailed rasterization criteria at the first stage of our framework. Increasing the grid resolution does not give a practical solution, because of the trade-off between high raster resolution and computational efficiency. However, given a "feasible" raster resolution, one may use a vector-to-raster conversion algorithm producing binomial proportion data, i.e. the percentage of the pixel covered by the land use category under study. In this way, the raster will appear as a grey-coloured intensity map, instead of black and white, yielding a more precise representation of the urban pattern. A first attempt in this direction showed that the rasterization algorithm is slower, but loss of information is substantially reduced w.r.t. the binary rasterization. This option is worth to investigation in the future, since the smoothing models and fitting procedures proposed in this work apply straightforwardly to the case of Binomial responses.

As a second issue, the pixel-wise procedures proposed in Sections 3.1 and 3.2 do not account for multiple testing. Methods to build simultaneous credible intervals for penalized splines have been proposed by Krivobokova et al. (2010) for Gaussian data. Extension to the spatial case and to non Gaussian data can be computationally demanding and is currently an open research line in spatial statistics. In a recent paper, Bolin and Lindgren (2013) proposed methods based on excursion sets; this approach could be applied in our context to identify pixels exceeding a certain threshold, ensuring that the statement holds for all of them simultaneously. Another possibility to control for multiple testing is to estimate the false discovery rate associated to any set of selected pixels; see Ventrucci et al. (2010) for an application in spatial epidemiology. Both strategies are worthwhile to be investigated in the future for building inferential tools dealing with simultaneous inferences over the smooth surface.

Finally, methods presented in this work can be adapted to the analysis of spatial point patterns, when points over a continuous space are summarized into grid counts, and generally to remotely sensed data available in raster format, such as land cover maps adopted in landscape fragmentation, deforestation and plant ecology studies.

## Acknowledgements

Massimo Ventrucci is funded by a FIRB 2012 grant (project nr. RBFR12URQJ, title: Statistical modeling of environmental phenomena: pollution, meteorology, health and their interactions), for research projects of national interest provided by the Italian Ministry of Education, Universities and Research.

## References

Albert, J., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88, 669–679.

Altieri, L., Cocchi, D., Pezzi, G., Scott, E. M., Ventrucci, M., 2014. Urban sprawl scatterplots for urban morphological zones data. Ecological Indicators 36 (0), 315 – 323.
URL http://www.sciencedirect.com/science/article/pii/S1470160X13002732

Angel, S., Parent, J., Civco, D., 2010. The fragmentation of urban footprints: Global evidence of sprawl, 1990-2000. Tech. rep., Lincoln Institute of Land Policy.

Bolin, D., Lindgren, F., 2013. Excursion and contour uncertainty regions for latent gaussian models. Journal of the Royal Statistical Society, Series B, in press.

Borrego, C., Martins, H., Tchepel, O., Salmim, L., Monteiro, A., Miranda, A., 2006. How urban structure can affect city sustainability from an air quality perspective. Environmental Modelling & Software 21 (4), 461–467.

Bowman, A. W., 2006. Comparing nonparametric surfaces. Statistical Modelling 6 (4), 279–299.
URL http://smj.sagepub.com/content/6/4/279.abstract

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Review ArticleDigital change detection methods in ecosystem monitoring: a review. International Journal of Remote Sensing 25 (9), 1565–1596.

Crainiceanu, C., Ruppert, D., Wand, M., 2005. Bayesian analysis for penalized spline regression using winbugs. Journal of Statistical Software 14 (1), 1–24.

Dong, H., Pengyu, Z., 2014. Smart growth in two contrastive metropolitan areas: A comparison between Portland and Los Angeles. Urban Studies.

Duczmal, L., Tavares, R., Patil, G., Cançado, A. L. F., 2010. Testing spatial cluster occurrence in maps equipped with environmentally defined structures. Environmental and Ecological Statistics 17 (2), 183–202. URL http://dx.doi.org/10.1007/s10651-010-0141-0

EEA, 1994. Corine land cover. Tech. rep., Commission of the European Communities. URL http://www.eea.europa.eu/publications/COR0-landcover

EEA, 2006. Urban sprawl in Europe. The ignored challenge. Tech. Rep. 10, Environmental European Agency.

EEA, 2011. Analysing and managing urban growth. Tech. rep., Environmental European Agency. URL http://www.eea.europa.eu/articles/analysing-and-managing-urban-growth

Eilers, P., Currie, I., Durbán, M., 2006. Fast and compact smoothing on large multidimensional grids. Computational Statistics & Data Analysis 5, 61–76.

Eilers, P., Marx, B., 1996. Flexible smoothing with b-splines and penalties. Statistical Science 11, 89–121.

Foody, G. M., 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80 (1), 185 – 201. URL http://www.sciencedirect.com/science/article/pii/S0034425701002954

Frühwirth-Schnatter, S., Wagner, H., 2008. Marginal likelihoods for non-gaussian models using auxiliary mixture sampling. Computational Statistics & Data Analysis 52 (10), 4608 – 4624. URL http://www.sciencedirect.com/science/article/pii/S016794730800176X

Furrer, R., Sain, S. R., 2010. spam: A sparse matrix R package with emphasis on mcmc methods for Gaussian Markov Random Fields. Journal of Statistical Software 36 (10), 1–25. URL http://www.jstatsoft.org/v36/i10/

Hijmans, R. J., 2013. raster: Geographic data analysis and modeling. R package version 2.1-49. URL http://CRAN.R-project.org/package=raster

Jaeger, J., Bertiller, R., Schwick, C., Kienast, F., 2010. Suitability criteria for measures of urban sprawl. Ecological Indicators 10, 397–406.

Kelly-Schwartz, A., Stockard, J., Doyle, S., Schlossberg, M., 2004. Is sprawl unhealty? a multilevel analysis of the relationship of metropolitan sprawl to the health of individuals. Journal of Planning Education and Research 24, 184–196.

Krivobokova, T., Kneib, T., Claeskens, G., 2010. Simultaneous confidence bands for penalized spline estimators. Journal of the American Statistical Association 105 (490), 852–863.

URL http://dx.doi.org/10.1198/jasa.2010.tm09165
Lang, S., Brezger, A., 2004. Bayesian p-splines. Journal of Computational and Graphical Statistics 13, 183–212.
Lawson, A. B., 2010. Hotspot detection and clustering: ways and means. Environmental and Ecological Statistics 17 (2), 231–245.
URL http://dx.doi.org/10.1007/s10651-010-0142-z
Lechner, A., Stein, A., Jones, S., Ferwerda, J., 2009. Remote sensing of small and linear features: Quantifying the effects of patch size and length, grid position and detectability on land cover mapping. Remote Sensing of Environment 113, 2194–2204.
Lee, D., Durbán, M., 2009. Smooth-car mixed models for spatial count data. Computational Statistics and data Analysis 53, 2968–2977.
Pasanen, L., Holmström, L., 2015. Bayesian scale space analysis of temporal changes in satellite images. Journal of Applied Statistics 42 (1), 50–70.
Patil, G. P., Joshi, S. W., Koli, R. E., 2010. Pulse, progressive upper level set scan statistic for geospatial hotspot detection. Environmental and Ecological Statistics 17 (2), 149–182.
URL http://dx.doi.org/10.1007/s10651-010-0140-1
Rue, H., 2001. Fast Sampling of Gaussian Markov Random Fields. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 63 (2), pp. 325–338.
Rue, H., Held, L., 2005. Gaussian Markov Random Fields. Chapman and Hall/CRC.
Ruppert, D., Wand, P., Carroll, R., 2003. Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.
Torrens, P., 2008. A toolkit for measuring sprawl. Applied Spatial Analysis and Policy 1, 5–36.
Trautmann, H., Steuer, D., Mersmann, O., Bornkamp, B., 2012. truncnorm: Truncated normal distribution. R package version 1.0-6.
URL http://CRAN.R-project.org/package=truncnorm
Tsai, Y., 2005. Quantifying urban form: Compactness versus sprawl. Urban Studies 42(1), 141–161.
Ventrucci, M., (ne Ferguson), C. M., Gross, J., Schoffelen, J.-M., Bowman, A. W., 2011. Spatiotemporal smoothing of single trial {MEG} data. Journal of Neuroscience Methods 200 (2), 219 – 228.
URL http://www.sciencedirect.com/science/article/pii/S0165027011003396
Ventrucci, M., Rue, H., Nov. 2015. Penalized complexity priors for degrees of freedom in Bayesian P-splines. ArXiv e-prints.
URL http://arxiv.org/pdf/1511.05748v2.pdf
Ventrucci, M., Scott, E. M., Cocchi, D., 2010. Multiple testing on standardized mortality ratios: a bayesian hierarchical model for fdr estimation. Biostatistics.
URL http://biostatistics.oxfordjournals.org/content/early/2010/06/24/biostatistics.kxq040.abstract
Wei, Y., Ye, X., 2014. Urbanization, urban land expansion and environmental change in China. Stochastic Environmental Research and Risk Assessment

28 (4), 757–765.
Wilson, B., Chakraborty, A., 2013. The environmental impacts of sprawl: Emergent themes from the past decade of planning research. Sustainability 5, 3302–3327.
Wood, S. N., 2003. Thin plate regression splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65 (1), 95–114. URL http://dx.doi.org/10.1111/1467-9868.00374

## Supplemental material

Augmented model representation

Let us recall the spatiotemporal model described in equation (6), section 2.5 of our paper.

$$\begin{aligned} y_{rct}|\delta_t, \boldsymbol{\gamma}, \boldsymbol{\theta}_t &\sim Ber(\mu_{rct}) \\ g(\mu_{rct}) &= \delta_t + \boldsymbol{x}_{rct}^{\mathsf{T}}\boldsymbol{\gamma} + \boldsymbol{B}_{rct}\boldsymbol{\theta}_t \qquad t = 1, ..., T; \end{aligned} \tag{7}$$

The posterior distribution of model (7) is intractable, thus MCMC methods based on Metropolis Hasting (M-H) are needed to draw a sample from the posterior distribution of the probability surface. A simpler approach which allows to avoid complicated M-H algorithms is to use the popular alternative representation of a probit model proposed by Albert and Chib (1993). Under their approach, model (7) is equivalent to the augmented model:

$$y_{rc} = \begin{cases} 1 & \text{if } s_{rct} > 0 \\ 0 & \text{otherwise} \end{cases}$$

where,

$$\begin{aligned} s_{rct} &= \delta_t + \boldsymbol{x}_{rct}^{\mathsf{T}}\boldsymbol{\gamma} + \boldsymbol{B}_{rct}\boldsymbol{\theta}_t + \epsilon_{rct} \\ \epsilon_{rct} &\sim N(0, 1) \end{aligned} \tag{8}$$

In the first level of the hierarchy, a set of $nT$ auxiliary variables, one at each pixel and time, is introduced by adding standard normal random variables $\epsilon_{rct}$ to the linear predictor, as shown in equation (8). These auxiliary variables can be collected in vector $\boldsymbol{s}$ which represents a set of pseudo-data. Note that the binary response $y_{rc}$ is now determined by the sign of $s_{rct}$.

At the second level of the hierarchy the model is completed by priors for the intercepts $\delta_1, ..., \delta_T$, the slopes $\boldsymbol{\gamma}$ and the vectors of spline coefficients $\boldsymbol{\theta}_t$, $t = 1, ..., T$, (as described in section 2.5 of our paper).

$$\begin{aligned} \delta_t &\sim N\left(0, \tau^{-1}\right) \qquad t = 1, ..., T, \\ \boldsymbol{\gamma} &\sim N\left(0, \tau^{-1}\boldsymbol{I}_p\right) \\ \boldsymbol{\theta_t} &\sim N\left(0, \boldsymbol{Q}_t^{-1}\right) \qquad t = 1, ..., T, \end{aligned} \tag{9}$$

where $\boldsymbol{Q}_t = \lambda_t \boldsymbol{K}$ is the IGMRF prior precision matrix, while $\tau$ is the prior precision for the fixed effects which we take equal to $10^{-5}$. At the third level of the hierarchy, a prior uninformative $Gamma(a, b)$, with shape $a = 1$ and rate $b = 5 \cdot 10^{-5}$ is assumed for the precision parameters $\lambda_t$, $t = 1, ..., T$.

Model fitting details

Below find some more detail about model fitting via MCMC. For simplicity of notation we collect the fixed effects (i.e. the intercept terms and the slopes) in a unique vector $\boldsymbol{\beta} = \left[\delta_1, ..., \delta_T, \boldsymbol{\gamma}^\top\right]^\top$ (extension to time-specific slopes is straightforward). In addition, we specify $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_1^\top, ..., \boldsymbol{\theta}_T^\top\right]^\top$ as the full vector of spline coefficients (where $\boldsymbol{\theta}_t$ is the vector of $q$ spline coefficients representing the surface at time $t$) and $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_T]^\top$ the associated precision parameters. The fixed effect design matrix $\boldsymbol{X}$ includes covariates and additional dummy variables for the time-specific intercepts (or simply a row of ones if an overall intercept $\delta$ is assumed in the model). The full basis matrix is given by $\boldsymbol{B} = \boldsymbol{I}_T \otimes \boldsymbol{B}_t$, with $\boldsymbol{B}_t = \boldsymbol{C} \otimes \boldsymbol{R}$ (see section 2.5 of our paper).

The joint posterior of our model is

$$\pi(\boldsymbol{s}, \boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{s})\pi(\boldsymbol{s}|\boldsymbol{\beta}, \boldsymbol{\theta})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda}), \tag{10}$$

Note that $\pi(\boldsymbol{y}|\boldsymbol{s})$ is equal to 1 by assumption, as the observed data are not random in an augmented model approach. Thus, conditionally on the auxiliary variables $\boldsymbol{s}$, the binary observations $\boldsymbol{y}$ and parameters $(\boldsymbol{\gamma}, \boldsymbol{\theta})$ are independent. The full conditional distribution for the set of pseudo-data $\boldsymbol{s}$ is a truncated multivariate normal (Albert and Chib, 1993),

$$s_{rct}|all \sim \begin{cases} N\left(\delta_t + \boldsymbol{x}_{rct}^\top \boldsymbol{\gamma} + \boldsymbol{B}_{rct}\boldsymbol{\theta}_t, 1\right) I\left(s_{rct} > 0\right) & \text{if } y_{rct} = 1; \\ N\left(\delta_t + \boldsymbol{x}_{rct}^\top \boldsymbol{\gamma} + \boldsymbol{B}_{rct}\boldsymbol{\theta}_t, 1\right) I\left(s_{rct} \le 0\right) & \text{otherwise,} \end{cases} \tag{11}$$

where, as specified in section 2 of our paper, recall that notation $\boldsymbol{B}_{rct}$ indicates the specific row entry of $\boldsymbol{B}_t$ with B-splines evaluated at pixel $(r, c)$ and time $t$. From (11) it follows that the full conditional distributions for both fixed effects $\boldsymbol{\beta}$ and spline coefficients $\boldsymbol{\theta}$ are Gaussian Markov Random Fields (GMRFs, Rue and Held (2005)). These full conditionals are reported below.

$$\begin{aligned} \boldsymbol{\beta}|all &\sim N\left(\boldsymbol{Q}_\beta^{-1} b_\beta, \boldsymbol{Q}_\beta^{-1}\right) \\ \boldsymbol{Q}_\beta &= \boldsymbol{X}^\top \boldsymbol{X} + \tau \\ b_\beta &= \boldsymbol{X}^\top \left(\boldsymbol{s} - \boldsymbol{B}\boldsymbol{\theta}\right) \end{aligned} \tag{12}$$

$$\begin{aligned} \boldsymbol{\theta}|all &\sim N\left(\boldsymbol{Q}_\theta^{-1} b_\theta, \boldsymbol{Q}_\theta^{-1}\right) \\ \boldsymbol{Q}_\theta &= \boldsymbol{B}^\top \boldsymbol{B} + diag(\boldsymbol{\lambda}) \otimes \boldsymbol{K} \\ b_\theta &= \boldsymbol{B}^\top \left(\boldsymbol{s} - \boldsymbol{X}\boldsymbol{\beta}\right) \end{aligned} \tag{13}$$

$$\lambda_t|all \sim \mathrm{G}\left(a + \frac{rank(\boldsymbol{K})}{2}, b + \frac{\boldsymbol{\theta}_t^\top \boldsymbol{K} \boldsymbol{\theta}_t}{2}\right) \quad \forall t = 1, ..., T \tag{14}$$

A Gibbs algorithm can be implemented by sampling in turn from the full conditionals (11), (12), (13) and (14). Sampling from full conditionals in (13) can be done efficiently in one block, using the algorithms proposed by Rue and Held (2005), which perform solve operations on the Cholesky factor of the sparse precision matrix $\boldsymbol{Q}_\theta$. A computationally intensive Cholesky update must be done at each MCMC iteration when sampling a new $\boldsymbol{\theta}$, which contains several thousand elements. However, the computational cost of each Cholesky update can be substantially reduced when the sparse structure of the Cholesky triangle is known (Furrer and Sain, 2010). The algorithm has been implemented in R using the package `spam` (Furrer and Sain, 2010) which includes fast routines for sampling GMRFs based on sparse Cholesky decomposition. Sampling from truncated normal distributions (11) is efficiently done using the package `truncnorm` (Trautmann et al., 2012). For identifiability of the P-spline components and the intercept terms, suitable sum-to-zero constraints must be applied to the spline coefficients sampled at each MCMC iteration. If the model includes time dependent intercepts (as the model used in the application in section 4 of our paper), we need to center $\boldsymbol{\theta}_t$ such that $(\boldsymbol{C} \otimes \boldsymbol{R})\boldsymbol{\theta}_t = \mathbf{0}$, at each MCMC iteration. (If the model contained only an overall intercept $\delta$, it would suffice to center $\boldsymbol{\theta}$ such that $\boldsymbol{B}\boldsymbol{\theta} = \mathbf{0}$).

In the application in section 4 of our paper, results are based on an MCMC sample obtained by thinning a total of 30000 Gibbs iterations, after removal of 10000 burn-in iterations. We choose to collect only one sample every 30 in order to remove chain autocorrelation and guarantee a large effective sample size (ESS). As an alternative to thinning, to guarantee large ESS one could store a very large MCMC sample (perhaps much larger than 1000) of the probability surface (of $n_1 n_2$ pixels), but this can require huge memory storage even with rasters of moderate size. Finally, as regards computational time, the Gibbs algorithm takes around 3, 4 and 6.5 minutes to run one hundred iterations of model (7) when using knot spacing equal to 1 $km$ ($q = 1089$), 500 $m$ ($q = 3969$) and 350 $m$ ($q = 10609$), respectively, using an Intel(R) Core(TM) i7 CPU 2.00GHz. `R` code is available on request.