Spatial and Spatio-Temporal models with `R-INLA`

Marta Blangiardo[a,1,*], Michela Cameletti[b,1], Gianluca Baio[c,d], Håvard Rue[e]

[a]*MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, UK*
[b]*Department of Management, Economics and Quantitative Methods, University of Bergamo, Italy*
[c]*Department of Statistical Science, University College London, UK*
[d]*Department of Statistics, University of Milano Bicocca, Italy*
[e] *Department of Mathematical Sciences, Norwegian University of Science and Technology, Norway*

Abstract

During the last three decades, Bayesian methods have developed greatly in the field of epidemiology. Their main challenge focusses around computation, but the advent of Markov Chain Monte Carlo methods (MCMC) and in particular of the `WinBUGS` software has opened the doors of Bayesian modelling to the wide research community. However model complexity and database dimension still remain a constraint.

Recently the use of Gaussian random fields has become increasingly popular in epidemiology as very often epidemiological data are characterised by a spatial and/or temporal structure which needs to be taken into account in the inferential process. The Integrated Nested Laplace Approximation (INLA) approach has been developed as a computationally efficient alternative to MCMC and the availability of an `R` package (`R-INLA`) allows researchers to easily apply this method.

In this paper we review the INLA approach and present some applications on spatial and spatio-temporal data.

Keywords: Integrated Nested Laplace Approximation, Stochastic Partial Differential Equation approach, Bayesian approach, Spatial structure, Spatio-temporal structure, Area-level data, Point-level data.

[*]Tel +44(0)207 594 3309; `m.blangiardo@imperial.ac.uk`

1. Introduction

During the last three decades, Bayesian methods have developed greatly and are now widely established in many research areas, from clinical trials (Berry et al., 2011), to health economic assessment (Baio, 2012) to the social sciences (Jackman, 2009), to epidemiology (Greenland, 2006).

The basic idea behind the Bayesian approach is that effectively only one form of uncertainty exists, which is described by suitable probability distributions. Thus, there is no fundamental distinction between observable data or unobservable parameters, which are also considered as random quantities. The uncertainty about the realised value of the parameters given the current state of information (i.e. before observing any new data) is described by a *prior* distribution. The inferential process combines the prior and the (current) data model to derive the *posterior* distribution, which is typically, but not necessarily, the objective of the inference (Bernardo and Smith, 2000; Lindley, 2006).

There are several advantages to the Bayesian approach: for instance the specification of prior distributions allows the formal inclusion of information that can be obtained through previous studies or from expert opinion; the (posterior) probability that a parameter does/does not exceed a certain threshold is easily obtained from the posterior distribution, providing a more intuitive and interpretable quantity than a frequentist *p*-value. In addition, within the Bayesian approach, it is easy to specify a hierarchical structure on the data and/or parameters, which presents the added benefit of making prediction for new observations and missing data imputation relatively straightforward.

Epidemiological data, e.g. in terms of an outcome and one or more risk factors or confounders, are often characterised by a spatial and/or temporal structure which needs to be taken into account in the inferential process. Under these circumstances, the Bayesian approach is generally particularly effective (Dunson, 2001) and has been applied in several epidemiological applications, from ecology (Clark, 2005) to environmental studies (Wikle, 2003; Clark and Gelfand, 2006), to infectious disease (Jewell et al., 2009). For example, if the data consist of aggregated counts of outcomes and covariates, typically disease mapping and/or ecological regression can be specified (Lawson, 2009). Alternatively, if the outcome or risk factors data are observed at point locations, then geostatistical models are considered as suitable representations of the problem (Diggle and Ribeiro, 2007).

2

Both models can be specified in a Bayesian framework by simply extending the concept of hierarchical structure, allowing to account for similarities based on the neighborhood or on the distance, for area-level or point-reference data, respectively. However, particularly in these cases, the main challenge in Bayesian statistics resides in the computational aspects. Markov Chain Monte Carlo (MCMC) methods (Brooks et al., 2011; Robert and Casella, 2004), are normally used for Bayesian computation, arguably thanks to the wide popularity of the `BUGS` software (Lunn et al., 2009, 2012). While extremely flexible and able to deal with virtually any type of data and model, in all but trivial cases MCMC methods involve computationally- and time-intensive simulations to obtain the posterior distribution for the parameters. Consequently, the complexity of the model and the database dimension often remain fundamental issues.

The Integrated Nested Laplace Approximation (INLA; Rue et al., 2009) approach has been recently developed as a computationally efficient alternative to MCMC. INLA is designed for *latent Gaussian models*, a very wide and flexible class of models ranging from (generalized) linear mixed to spatial and spatio-temporal models. For this reason, INLA can be successfully used in a great variety of applications (e.g. Li et al., 2012; Riebler et al., 2012; Ruiz-Cárdenas et al., 2012; Martino et al., 2011; Roos and Held, 2011; Schrödle and Held, 2011a,b; Schrödle et al., 2011; Paul et al., 2010), also thanks to the availability of an `R` package named `R-INLA` (Martino and Rue, 2010). Furthermore, INLA can be combined with the Stochastic Partial Differential Equation (SPDE) approach proposed by Lindgren et al. (2011) in order to implement spatial and spatio-temporal models for point-reference data.

The objective of this paper is to present the basic features of the INLA approach as applied to spatial and spatio-temporal data. The paper is structured as follows: first in Section 2 we review the main characteristics of spatial and spatio-temporal data defined at the point and area level; then we provide an overview of the theory behind INLA in Section 3 and present two practical applications on area level data in Sections 3.2 and 3.3. After this in Section 4 we review the SPDE approach to deal with geostatistical data, and then present two practical applications on spatial and spatio-temporal point level data (Sections 4.1 and 4.2). Finally Section 5 discusses some of the issues and provides some conclusions.

2. Spatial and spatio-temporal data

Spatial data are defined as realisations of a stochastic process indexed by space

$$Y(s) \equiv \{y(s), s \in \mathcal{D}\}$$

where $\mathcal{D}$ is a (fixed) subset of $\mathbb{R}^d$ (here we consider $d = 2$). The actual data can be then represented by a collection of observations $\boldsymbol{y} = \{y(s_1), \ldots, y(s_n)\}$, where the set $(s_1, \ldots, s_n)$ indicates the spatial units at which the measurements are taken. Depending on $\mathcal{D}$ being a continuous surface or a countable collection of d-dimensional spatial units, the problem can be specified as a spatially continuous or discrete random process, respectively (Gelfand et al., 2010).

For example, we can consider a collection of air pollutant measurements obtained by monitors located in the set $(s_1, \ldots, s_n)$ of n *points*. In this case, $\boldsymbol{y}$ is a realisation of the air pollution process that changes continuously in space and we usually refer to it as geostatistical or point-reference data. Alternatively, we may be interested in studying the spatial pattern of a certain health condition observed in a set $(s_1, \ldots, s_n)$ of n *areas* (rather than points), defined for example by census tracts or counties; in this case, $\boldsymbol{y}$ may represent a suitable summary, e.g. the number of cases observed in each area.

The first step in defining a spatial model within the Bayesian framework is to identify a probability distribution for the observed data. Usually we select a distribution from the Exponential family, indexed by a set of parameters $\boldsymbol{\theta}$ accounting for the spatial correlation — note that for the sake of simplicity we slightly abuse the notation and index the generic spatial point or area by using just the subscript i, rather than the indicator s_i, in the following.

In the case of geostatistical data, the parameters are defined as a latent stationary *Gaussian field* (GF), a function of some *hyper-parameters* $\boldsymbol{\psi}$ associated with a suitable prior distribution $p(\boldsymbol{\psi})$. This is equivalent to assuming that $\boldsymbol{\theta}$ has a multivariate Normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ and spatially structured covariance matrix $\boldsymbol{\Sigma}$, whose generic element is $\boldsymbol{\Sigma}_{ij} = \mathrm{Cov}\,(\theta_i, \theta_j) = \sigma_{\mathcal{C}}^2 \mathcal{C}(\Delta_{ij})$. Here $\sigma_{\mathcal{C}}^2$ is the variance component and for $i, j = 1, \ldots, n$

$$\mathcal{C}(\Delta_{ij}) = \frac{1}{\Gamma(\lambda)2^{\lambda-1}} \left(\kappa\Delta_{ij}\right)^\lambda K_\lambda \left(\kappa\Delta_{ij}\right) \tag{1}$$

is the (isotropic) Matérn spatial covariance function[1] (Cressie, 1993) depending on the Euclidean distance between the locations $\Delta_{ij} = \|s_i - s_j\|$. The parameter K_λ denotes the modified Bessel function of second kind and order $\lambda > 0$, which measures the degree of smoothness of the process and is usually kept fixed. Conversely, $\kappa > 0$ is a scaling parameter related to the range r, i.e. the distance at which the spatial correlation becomes almost null. Typically, the empirically derived definition $r = \frac{\sqrt{8\lambda}}{\kappa}$ is used (see Section 2 in Lindgren et al., 2011), with r corresponding to the distance at which the spatial correlation is close to 0.1, for each λ.

In the case of area level data, it is possible to reformulate the problem in terms of the neighbourhood structure. Under the Markovian property that the generic element of the parameters vector θ_i is independent on any other element, given the set of its neighbours $\mathcal{N}(i)$

$$\theta_i \perp\!\!\!\perp \boldsymbol{\theta}_{-i} \mid \boldsymbol{\theta}_{\mathcal{N}(i)},$$

($\boldsymbol{\theta}_{-i}$ indicates all the elements in $\boldsymbol{\theta}$ but the $i-$th), the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ is sparse, which produces great computational benefits. In other words, for any pair of elements (i, j)

$$\theta_i \perp\!\!\!\perp \theta_j \mid \boldsymbol{\theta}_{-ij} \iff \boldsymbol{Q}_{ij} = 0$$

i.e. the non-zero pattern in the precision matrix is given by the neighbourhood structure of the process. Thus, $\boldsymbol{Q}_{ij} \neq 0$ only if $j \in \{i, \mathcal{N}(i)\}$. This specification is known as *Gaussian Markov Random Field* (GMRF, Rue and Held, 2005)

The concept of spatial process can be extended to the spatio-temporal case including a time dimension. The data are then defined by a process

$$Y(s, t) \equiv \{y(s, t), (s, t) \in \mathcal{D} \in \mathbb{R}^2 \times \mathbb{R}\}$$

and are observed at n spatial locations or areas and at T time points. When spatio-temporal geostatistical data are considered (Gelfand et al., 2010,

---

[1]Other models for the spatial covariance function are available in the geostatistical literature (see e.g. Cressie, 1993 and Banerjee et al., 2004). The fact that here we focus on the Matérn model - as required by the SPDE approach described in Section 4 - should not be considered as a restriction. In fact, as described in Guttorp and Gneiting (2006), the Matèrn family is a very flexible class of covariance functions able to cover a wide range of spatial fields.

Chapter 23), we need to define a valid spatio-temporal covariance function given by $\text{Cov}(\theta_{it}, \theta_{ju}) = \sigma_{\mathcal{C}}^2 \mathcal{C}(s_i, s_j; t, u)$. If we assume stationarity in space and time, the space-time covariance function can be written as a function of the spatial Euclidean distance Δ_{ij} and of the temporal lag $\Lambda_{tu} = |t - u|$, i.e. $\text{Cov}(\theta_{it}, \theta_{ju}) = \sigma_{\mathcal{C}}^2 \mathcal{C}(\Delta_{ij}; \Lambda_{tu})$; several examples of valid non-separable space-time covariance functions are reported in Cressie and Huang (1999) and Gneiting (2002).

In practice, to overcome the computational complexity of non-separable models, some simplifications are introduced. For example, under the separability hypothesis the space-time covariance function is decomposed into the sum (or the product) of a purely spatial and a purely temporal term (Gneiting et al., 2006), e.g. $\text{Cov}(\theta_{it}, \theta_{ju}) = \sigma_{\mathcal{C}}^2 \mathcal{C}_1(\Delta_{ij}) \mathcal{C}_2(\Lambda_{tu})$, with $\mathcal{C}_1$ and $\mathcal{C}_2$ being the spatial and temporal correlation function, respectively . Alternatively, it is possible to assume that the spatial correlation is constant in time, giving rise to a space-time covariance function that is purely spatial when $t = u$, i.e. $\text{Cov}(\theta_{it}, \theta_{ju}) = \sigma_{\mathcal{C}}^2 \mathcal{C}(\Delta_{ij})$, and is zero otherwise. In this case, the temporal evolution could be introduced assuming that the spatial process evolves in time following an autoregressive dynamics (see e.g. Harvill, 2010).

Similar reasoning can be applied to area level data; the GMRF framework can be extended to include a precision matrix defined also in terms of time, assuming again a neighborhood structure. If a space-time interaction is included, its precision can be obtained through the Kronecker product of the precision matrices for the space and time effects interacting — see Clayton (1996) and Knorr-Held (2000) for a detailed description.

3. Integrated Nested Laplace Approximation (INLA)

Often, in a statistical analysis the interest is in estimating the effect of a set of relevant covariates on some function (typically the mean) of the observed data, while accounting for the spatial or spatio-temporal correlation implied in the model.

A very general way of specifying this problem is by modelling the mean for the i-th unit by means of an additive linear predictor, defined on a suitable scale (e.g. logistic for binomial data)

$$\eta_i = \alpha + \sum_{m=1}^{M} \beta_m x_{mi} + \sum_{l=1}^{L} f_l(z_{li}). \tag{2}$$

Here α is a scalar representing the intercept; the coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)$ quantify the effect of some covariates $\boldsymbol{x} = (x_1, \ldots, x_M)$ on the response; and $\boldsymbol{f} = \{f_1(\cdot), \ldots, f_L(\cdot)\}$ is a collection of functions defined in terms of a set of covariates $\boldsymbol{z} = (z_1, \ldots, z_L)$. Upon varying the form of the functions $f_l(\cdot)$, this formulation can accommodate a wide range of models, from standard and hierarchical regression, to spatial and spatio-temporal models (Rue et al., 2009).

Given the specification in (2), the vector of parameters is represented by $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\beta}, \boldsymbol{f}\}$. In line with the discussion in Section 2, we can assume a GMRF prior on $\boldsymbol{\theta}$, with mean $\boldsymbol{0}$ and a precision matrix $\boldsymbol{Q}$. In addition, because of the conditional independence relationships implied by the GMRF, the vector of the K hyper-parameters $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)$ will typically have dimension of order $(1 + L)$ and thus will be much smaller than $\boldsymbol{\theta}$.

The objectives of the Bayesian computation are the marginal posterior distributions for each of the elements of the parameters vector

$$p(\theta_i \mid \boldsymbol{y}) = \int p(\boldsymbol{\psi} \mid \boldsymbol{y}) p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y}) \mathrm{d}\boldsymbol{\psi}$$

and (possibly) for each element of the hyper-parameters vector

$$p(\psi_k \mid \boldsymbol{y}) = \int p(\boldsymbol{\psi} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{\psi}_{-k}.$$

Thus, we need to compute: *i)* $p(\boldsymbol{\psi} \mid \boldsymbol{y})$, from which also all the relevant marginals $p(\psi_k \mid \boldsymbol{y})$ can be obtained; and *ii)* $p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y})$, which is needed to compute the marginal posterior for the parameters. The INLA approach exploits the assumptions of the model to produce a numerical approximation to the posteriors of interest, based on the *Laplace approximation* (Tierney and Kadane, 1986).

The first task *i)* consists of the computation of an approximation to the posterior marginal distribution of the hyper-parameters as

$$
\begin{aligned}
p(\boldsymbol{\psi} \mid \boldsymbol{y}) &= \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \boldsymbol{y})}{p(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})} \propto \frac{p(\boldsymbol{\psi}) p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) p(\boldsymbol{y} \mid \boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})} \\
&\approx \left. \frac{p(\boldsymbol{\psi}) p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) p(\boldsymbol{y} \mid \boldsymbol{\theta})}{\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})} =: \tilde{p}(\boldsymbol{\psi} \mid \boldsymbol{y}) \quad (3)
\end{aligned}
$$

where $\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})$ is the Gaussian approximation (Rue et al., 2009, Section 2.2) of $p(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})$ and $\boldsymbol{\theta}^*(\boldsymbol{\psi})$ is its mode. As described in Rue et al. (2009),

(3) is equivalent to the Laplace approximation of a marginal posterior density as proposed in Tierney and Kadane (1986).

The second task *ii)* is slightly more complex, because in general there will be more elements in $\boldsymbol{\theta}$ than there are in $\boldsymbol{\psi}$ and thus this computation is more expensive. One easy possibility is to approximate the posterior conditional distributions $p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y})$ directly as the marginals from $\tilde{p}(\boldsymbol{\theta} \mid \boldsymbol{\psi}, \boldsymbol{y})$, i.e. using a Normal distribution, where the precision matrix is based on the Cholesky decomposition of the precision matrix $\boldsymbol{Q}$ (Rue and Martino, 2007). While this is very fast, the approximation is generally not very good. Alternatively, it is possible to re-write the vector of parameters as $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$ and use again Laplace approximation to obtain

$$
\begin{aligned}
p(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y}) &= \frac{p\left((\theta_i, \boldsymbol{\theta}_{-i}) \mid \boldsymbol{\psi}, \boldsymbol{y}\right)}{p(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})} \\
&\approx \left. \frac{p(\boldsymbol{\theta}, \boldsymbol{\psi} \mid \boldsymbol{y})}{\tilde{p}(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})} \right|_{\boldsymbol{\theta}_{-i} = \boldsymbol{\theta}^*_{-i}(\theta_i, \boldsymbol{\psi})} =: \tilde{p}(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y}) \quad (4)
\end{aligned}
$$

where $\tilde{p}(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})$ is the Gaussian approximation of $p(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})$ and $\boldsymbol{\theta}^*_{-i}(\theta_i, \boldsymbol{\psi})$ is its mode. Because the random variables $(\boldsymbol{\theta}_{-i} \mid \theta_i, \boldsymbol{\psi}, \boldsymbol{y})$ are in general reasonably Normal, the approximation provided by (4) typically works very well. This strategy, however, can be very expensive in computational terms. Consequently, the most efficient algorithm is the "Simplified Laplace approximation", which is based on a Taylor's series expansion of the Laplace approximation $\tilde{p}(\theta_i \mid \boldsymbol{\psi}, \boldsymbol{y})$ in (4). This is usually "corrected" by including a mixing term (e.g. spline), to increase the fit to the required distribution. The accuracy of this approximation is sufficient in many applied cases and the time needed for the computations is much shorter and thus this is the standard option.

Operationally, INLA proceeds by first exploring the marginal joint posterior for the hyper-parameters $\tilde{p}(\boldsymbol{\psi} \mid \boldsymbol{y})$ in order to locate the mode; a grid search is then performed and produces a set $\mathcal{G}$ of "relevant" points $\{\psi^*\}$ together with a corresponding set of weights $\{w_{\psi^*}\}$, to give the approximation to this distribution. Each marginal posterior $\tilde{p}(\psi^* \mid \boldsymbol{y})$ can be obtained using interpolation based on the computed values and correcting for (probable) skewness, e.g. by using log-splines. For each ψ^*, the conditional posteriors $\tilde{p}(\theta_i \mid \psi^*, \boldsymbol{y})$ are then evaluated on a grid of selected values for θ_i and the

marginal posteriors $\tilde{p}(\theta_i \mid \boldsymbol{y})$ are obtained by numerical integration

$$\tilde{p}(\theta_i \mid \boldsymbol{y}) \approx \sum_{\psi^* \in \mathcal{G}} \tilde{p}(\theta_i \mid \psi^*, \boldsymbol{y}) \tilde{p}(\psi^* \mid \boldsymbol{y}) w_{\psi^*}.$$

More details on this methods can be found in Rue et al. (2009); Martins et al. (2012); Blangiardo and Cameletti (2013).

3.1. The `R-INLA` package

The INLA approach described in the previous section is implemented in the `R` package `R-INLA`, which substitutes the standalone `INLA` program built upon the `GMRFLib` library (Martino and Rue, 2010). `R-INLA` is available for Linux, Mac and Windows operating systems. The web-site `http://www.r-inla.org/` provides documentation for the package as well as many worked examples and a discussion forum.

Assuming a vector of two covariates $\boldsymbol{x} = (x_1, x_2)$ and a function $f(\cdot)$ indexed by a third covariate z_1, (2) is reproduced in `R-INLA` through the command `formula`:

```
> formula <- y ~ 1 + x1 + x2 + f(z1, ...)
```

where `y`, `x1`, `x2` and `z1` are the column names of the data frame containing the data (for simplicity, we assume throughout that the data frame name is `data`). The regression coefficients α, β_1 and β_2 are by default given independent prior Normal distributions with zero mean and small precision (or equivalently large variance).

The term $f(\cdot)$ is used to specify the structure of the function $f(\cdot)$, using the following notation:

```
> f(z1, model = "...", ...)
```

where the string associated with the option `model` specifies the type of function. The default choice is `model="iid"`, documented typing `inla.doc("iid")` and it amounts to assuming exchangeable Normal distributions on `z1`. This specification can be used to build standard hierarchical models. The list of the other alternatives is available typing `names(inla.models()$latent)`; in addition, a detailed description of all the possible choices is available at the webpage `http://www.r-inla.org/models/latent-models`.

Once the model has been specified, we can run the INLA algorithm using the `inla` function:

9

```
> mod <- inla(formula, family = "...", data)
```

where `formula` has been specified above, `data` is the data frame containing all the variables in the formula and `family` is a string that specifies the distribution of the data (likelihood). The available data distributions are retrieved typing `names(inla.models()$likelihood)` and complete descriptions with examples are provided at `http://www.r-inla.org/models/likelihoods`. The `inla` function includes many other options; see `?inla` for a complete list. Note that INLA by default estimates the posterior marginal distribution for the hyperparameters is estimated using an integration free algorithm, described in Martins et al. (2012) and which has been showed to provide reasonably accurate estimates. Nevertheless if the interest lays primarily on the hyperparameters, an alternative method based on a more computational intensive grid exploration can be employed, using the command `inla.hyperpar` after running `inla`. We advise the reader to refer to the above cited paper for more details.

We illustrate more functionalities of `R-INLA` using the real data applications described in the following sections. The complete code for running the four examples can be downloaded from *Case studies* section at `www.r-inla.org`. [2]

3.2. INLA for spatial areal data: suicides in London

Disease mapping is commonly used in small area studies to assess the pattern of a particular disease and to identify areas characterised by unusually high or low relative risk (Lawson, 2009; Pascutto et al., 2000). Here we use the example presented in Congdon (2007) to investigate suicide mortality in $n = 32$ London boroughs in the period 1989-1993.

For the i-th area, the number of suicides y_i is modelled as

$$y_i \sim \text{Poisson}(\lambda_i),$$

where the mean λ_i is defined in terms of a rate ρ_i and the expected number of suicides E_i as $\lambda_i = \rho_i E_i$. In this case, the linear predictor is defined on the logarithmic scale

$$\eta_i = \log(\rho_i) = \alpha + \upsilon_i + \nu_i, \tag{5}$$

---

[2]Note that all the code has been developed using the INLA version released on the 5th of November 2012.

α is the intercept quantifying the average suicide rate in all the 32 boroughs; $\upsilon_i = f_1(i)$ and $\nu_i = f_2(i)$ are two area specific effects; $i = \{1, \ldots, n\}$ is the indicator for each borough (spatial areas) and corresponds to the variable `ID` in the data frame.

We assume a Besag-York-Mollie (BYM) specification (Besag et al., 1991), so υ_i is the spatially structured residual, modelled using an intrinsic conditional autoregressive structure (iCAR)

$$\upsilon_i \quad | \quad \upsilon_{j \neq i} \sim \text{Normal}(m_i, s_i^2)$$
$$m_i \quad = \quad \frac{\sum_{j \in \mathcal{N}(i)} \upsilon_j}{\#\mathcal{N}(i)} \qquad \text{and} \qquad s_i^2 = \frac{\sigma_\upsilon^2}{\#\mathcal{N}(i)},$$

where $\#\mathcal{N}(i)$ is the number of areas which share boundaries with the i-th one (i.e. its neighbours), as presented in Banerjee (2004). The parameter ν_i represents the unstructured residual, modelled using an exchangeable prior: $\nu_i \sim \text{Normal}(0, \sigma_\nu^2)$.

To run this model in `R-INLA` we first specify the formula, typing

```
formula <- y ~ 1 + f(ID, model="bym",graph=LDN.adj)
```

where `LDN.adj` is the graph which assignes the set of neighbours for each borough and that can be obtained from the shape file of the study region using the `R` packages `maptools` and `spdep`. Note that `R-INLA` parametrises $\xi_i = \upsilon_i + \nu_i$ and υ_i through `f(ID, model="bym",...)`.[1]

By default, minimally informative priors are specified *i)* on the log of the unstructured effect precision[3] $\log \tau_\nu \sim \log \text{Gamma}(1, 0.0005)$ and *ii)* on the log of the structured effect precision[3] $\log \tau_\upsilon \sim \log \text{Gamma}(1, 0.0005)$. Different priors can be specified through the option `hyper` in the `formula` specification, for instance

```
formula <- y ~ 1 + f(ID, model="bym",graph=LDN.adj, hyper  =
   list(prec.unstruct  =  list(prior="loggamma",param=c(1,0.01)),
        prec.spatial = list(prior="loggamma",param=c(1,0.001))))
```

---

[1] Alternatively it is possible to specify the two BYM components separately using `f(ID, model="besag",graph=LDN.adj)` for the spatial structured one (iCAR) and `f(ID2, model="iid",graph=LDN.adj)` for the unstructured one (exchangeable). In this case the ID needs to be duplicated (`ID2=ID`) as it is not allowed to define two functions on the same variable.

[3] Recall that the precision is defined as $\tau = 1/\sigma^2$

11

Of course, as in any Bayesian analysis, the choice of the prior may have a considerable impact on the results. Thus, it is necessary to think carefully about what is being used and perform sensitivity analyses to assess how the prior influences the estimations.

The model can be run using the `inla` function:

```
mod <- inla(formula,family="poisson",data=data,E=E)
```

With respect to the discussion of Section 3, in this case the parameters estimated by INLA are represented by $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\xi}, \boldsymbol{v}\}$ and the hyper-parameters are given by the precisions $\boldsymbol{\psi} = \{\tau_v^2, \tau_\nu^2\}$.

Summary information (e.g. the posterior mean and standard deviation, together with a 95% credible interval) can be obtained for each component of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. In particular, for the so called "fixed" effects (α, in this case), this can be obtained typing `mod$summary.fixed`; similarly, the summary statistics for the "random" effects (i.e. $\boldsymbol{\xi}$ and $\boldsymbol{v}$) are produced by `mod$summary.random`. The latter is a data frame formed by $2n$ rows: the first n rows include information on the area specific residuals ξ_i, which are the primary interest in a disease mapping study, while the remaining present information on the spatially structured residual v_i only.

The posterior mean of the exponentiated intercept α implies a 4% suicide rate across London, with a 95% credibility interval ranging from 1% to 8%. Figure 1 (a) shows the map of the posterior mean for the borough-specific relative risks of suicides $\boldsymbol{\zeta} = \exp(\boldsymbol{\xi})$, compared to the whole of London. Their posterior distributions are easily obtained applying an exponential transformation to the components of $\boldsymbol{\xi}$, which are in turn produced by the command `mod$marginals.random`. The built-in functions `inla.marginal.transform` and `inla.emarginal` compute marginals of transformed variables and expected values.

The uncertainty associated with the posterior means can also be mapped and provide useful information (Richardson et al., 2004). In particular, as the interest lays in the excess risk, we can visualise $p(\zeta_i > 1 \mid \boldsymbol{y})$, using the built-in function `inla.pmarginal`; the resulting map is presented in Figure 1 (b).

Finally, it could be interesting to evaluate the proportion of variance explained by the structured spatial component. The quantity σ_v^2 is the variance of the conditional autoregressive specification, while σ_ν^2 is the variance of the marginal unstructured component. Thus, the two are not directly comparable. Nevertheless it is possible to obtain an estimate of the posterior marginal

variance for the structured effect empirically through

$$s_v^2 = \frac{\sum_{i=1}^{n} (v_i - \bar{v})^2}{n-1},$$

where $\bar{v}$ is the average of $\boldsymbol{v}$, and then compare it to the posterior marginal variance for the unstructured effect, provided by σ_ν^2

$$\text{frac}_{\text{spatial}} = s_v^2 / (s_v^2 + \sigma_\nu^2).$$

In the current example, the proportion of spatial variance is about 0.97 suggesting that almost all the variability is explained by the spatial structure.

When risk factors are available and the aim of the study is to evaluate their effect on the risk of disease (or death), ecological regression models can be specified, simply extending the procedure described above. For instance, in the present example for each of the 32 boroughs the values of an index of social deprivation and an index of social fragmentation (describing lack of social connections and of sense of community) are known and stored respectively in the variables x_1 and x_2. To evaluate their impact on the risk of suicides, the model in (5) can be reformulated as

$$\eta_i \quad = \quad \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + v_i + \nu_i,$$

which can be coded in `R-INLA` using the formula

```
formula.cov <- y ~ 1+ f(ID,model="bym", graph=LDN.adj) + x1 + x2
```

The fixed effects $(\alpha, \beta_1, \beta_2)$ estimated by INLA are presented in Table 1. If exponentiated, they can be interpreted as relative risks: an increase of 1 unit in the deprivation index and in the social fragmentation index is associated respectively with an increase of around $9\% = \exp(0.089)$ and around $20\% = \exp(0.18)$ in the risk of suicides.

The map of the borough specific relative risks $\boldsymbol{\zeta}$ and their posterior probability of exceeding 1 are shown in Figure 1 (c)-(d); note that now they are interpreted as the residual relative risk for each area (compared to the whole of London) after the risk factors x_1 and x_2 are taken into account.

3.3. INLA for spatio-temporal areal data: low birth weight in Georgia

In this section we use counts of low birth weight, defined as less than 2,500 grams, for the 159 counties in the US state of Georgia during 2000-2010 (Lawson, 2009) to build a space-time disease mapping.

13

	Mean	Sd	2.5%	50%	97.5%
α	0.059	0.016	0.028	0.059	0.091
β_1	0.089	0.023	0.042	0.089	0.133
β_2	0.180	0.021	0.138	0.180	0.222

Table 1: Summary statistics: posterior mean, posterior standard deviation (Sd) and posterior 95% credible interval for the fixed effects of the ecological regression model.

The classical parametric formulation was introduced by Bernardinelli et al. (1995), and assume that the linear predictor can be written as

$$\eta_{it} = \alpha + \upsilon_i + \nu_i + (\beta + \delta_i) \times t. \qquad (6)$$

This formulation includes the same spatial structured and unstructured components as in (5), with: $\xi_i = \upsilon_i + \nu_i$; a main linear trend β, which represents the global time effect; and a differential trend δ_i, which identifies the interaction between time and space.

Since, for identifiability purposes a sum-to-zero constraint is imposed on $\boldsymbol{\delta}$ and $\boldsymbol{\nu}$, the terms δ_i represent the difference between the global trend β and the area specific trend. If $\delta_i < 0$ then the area-specific trend is less steep than the mean trend, whilst $\delta_i > 0$ implies that the area-specific trend is steeper than the mean trend. We assume $\delta_i \sim \text{Normal}(0, \tau_\delta)$, but other specification can be used, e.g. a conditional autoregressive structure, see Bernardinelli et al. (1995), Schrödle and Held (2011a) for a detailed description.

In R-INLA the interaction term needs to be specified through the formula as follows:

```
formula1 <- y ~ 1 + f(ID.area, model="bym",graph=Georgia.adj) +
            f(ID.area1,year,model="iid") + year
```

(we save the model associated with this formula in an inla object named mod1). Note that each function f(·) can only be assigned to one covariate in R-INLA, so in this case we need to create a new variable ID.area1 which is a duplicate of ID.area. In addition, year in the f(·) term is treated as a vector of weights.

This specification assumes a linear effect of time for each area (δ_i). According to Section 3 the parameters estimated by INLA are $\boldsymbol{\theta} = \{\alpha, \beta, \boldsymbol{\xi}, \boldsymbol{\upsilon}, \boldsymbol{\delta}\}$ and the hyper-parameters are represented by $\boldsymbol{\psi} = \{\tau_\upsilon, \tau_\nu, \tau_\delta\}$.

14

The assumption of linearity in the δ_i can be released (Knorr-Held, 2000) using a dynamic non parametric formulation for the linear predictor

$$\eta_{it} = \alpha + \upsilon_i + \nu_i + \gamma_t + \phi_t. \tag{7}$$

Here α, υ_i and ν_i have the same parametrisation as in (6); however, the term γ_t represents the temporally structured effect, modelled dynamically (e.g. using a random walk) through a neighboring structure

$$
\begin{aligned}
\gamma_t \mid \boldsymbol{\gamma}_{-t} \;\; &\sim \;\; \text{Normal}\left(\gamma_{t+1}, \tau_\gamma\right) && \text{for } t = 1 \\
\gamma_t \mid \boldsymbol{\gamma}_{-t} \;\; &\sim \;\; \text{Normal}\left(\frac{\gamma_{t-1} + \gamma_{t+1}}{2}, \frac{\tau_\gamma}{2}\right) && \text{for } t = 2, \ldots, T-1 \\
\gamma_t \mid \boldsymbol{\gamma}_{-t} \;\; &\sim \;\; \text{Normal}\left(\gamma_{t-1}, \tau_\gamma\right) && \text{for } t = T.
\end{aligned}
$$

Finally ϕ_t is specified by means of a Gaussian exchangeable prior: $\phi_t \sim$ Normal$(0, \tau_\phi)$.

This model is specified in R-INLA as

```
formula2 <- y ~ 1 + f(ID.area,model="bym",graph=Georgia.adj) +
            f(ID.year,model="rw1") + f(ID.year1,model="iid")
```

We save the resulting model in the inla object mod2. In this formulation $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\xi}, \boldsymbol{\upsilon}, \boldsymbol{\gamma}, \boldsymbol{\phi}\}$ and $\boldsymbol{\psi} = \{\tau_\upsilon, \tau_\nu, \tau_\gamma, \tau_\phi\}$.

It is easy to expand this model to allow for an interaction between space and time, which would explain differences in the time trend of low birth weight for different areas, e.g. using the following specification:

$$\eta_{it} = \alpha + \upsilon_i + \nu_i + \gamma_t + \phi_t + \delta_{it}. \tag{8}$$

There are several ways to define the interaction term: here, we assume that the two unstructured effects ν_i and ϕ_t interact. We re-write the precision matrix as the product of the scalar τ_ν (or τ_ϕ) and the so called structure matrix $\boldsymbol{F}_\nu$ (or $\boldsymbol{F}_\phi$), which identifies the neighboring structure; here the structure matrix $\boldsymbol{F}_\delta$ can be factorised as the Kronecker product of the structure matrix for $\boldsymbol{\nu}$ and $\boldsymbol{\phi}$ (Clayton, 1996): $\boldsymbol{F}_\delta = \boldsymbol{F}_\nu \otimes \boldsymbol{F}_\phi = \boldsymbol{I} \otimes \boldsymbol{I} = \boldsymbol{I}$ (because both $\boldsymbol{\nu}$ and $\boldsymbol{\phi}$ are unstructured). Consequently, we assume no spatial and/or temporal structure on the interaction and therefore $\delta_{it} \sim$ Normal$(0, \tau_\delta)$ — see Knorr-Held (2000) for a detailed description of other specifications. In this model $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\xi}, \boldsymbol{\upsilon}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{\delta}\}$ and $\boldsymbol{\psi} = \{\tau_\upsilon, \tau_\nu, \tau_\gamma, \tau_\phi, \tau_\delta\}$.

The corresponding R-INLA coding is

```
formula3 <- y ~ 1 + f(ID.area,model="bym", graph=Georgia.adj) +
            f(ID.year,model="rw1")  + f(ID.year1,model="iid") +
            f(ID.area.year,model="iid")
```

and the resulting model is saved in the object `mod3`.

In the three models presented in this section, we assume the default specification of `R-INLA` for the distribution of the hyper-parameters; therefore, similarly to the disease mapping model presented earlier, $\log \tau_\upsilon \sim \log \text{Gamma}(1, 0.0005)$ and $\log \tau_\nu \sim \log \text{Gamma}(1, 0.0005)$. In addition we specify a $\text{Gamma}(1, 0.0005)$ prior on the precision of the random walk and of the two unstructured effects.

One possible tool to evaluate the fit of these three competing models is the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), which can be computed in `R-INLA`, using the option `control.compute=list(dic=TRUE)`. Table 2 presents the DIC components for the three models: the dynamic parametrisation of the time trend improves the model fit and including the interaction shows the smaller DIC suggesting that, despite the added complexity, this model has a more appropriate fit to the data in hand. For this reason we focus on the results from this model.

Model	$\bar{D}$	p_D	DIC
mod1	11698.5	173.2	11871.7
mod2	11709.9	155.9	11865.9
mod3	11509.9	306.7	11816.6

Table 2: Deviance Information Criterion (DIC) for the three spatio-temporal models considered defined by Equations (6)-(8). $\bar{D}$ is the posterior mean of the deviance, measuring model fit; p_D is effective number of parameters, representing model complexity.

The spatial trend $\zeta_i = \exp(\xi_i)$ is presented in Figure 2 (a) for the 159 counties in Georgia, while Figure 2 (b) depicts the measure of uncertainty $p(\zeta_i > 1 \mid \boldsymbol{y})$. An increased risk can be seen in some parts of the country, characterised by a spatial relative risk above 1, and a posterior probabilities above 0.8, indicating a relatively small level of associated uncertainty. The temporal trend is included in Figure 2 (c) and shows an increase in the risk of low birth weight between 2000 and 2010.

The posterior probabilities for the interactions, $p(\exp(\delta_{it}) > 1 \mid \boldsymbol{y})$, are presented in Figure 3 for three years: as expected only a handful of areas shows evidence of an interaction larger than 1, changing in different years.

16

4. The stochastic partial differential equation approach for geostatistical data

Point-reference data can be dealt with the stochastic partial differential equation (SPDE) approach proposed by Lindgren et al. (2011). This consists in representing a continuous spatial process, e.g. a GF with the Matèrn covariance function defined in (1), as a discretely indexed spatial random process (e.g. a GMRF). This in turn produces substantial computational advantages. In fact, spatial GFs are affected by the so called "big n problem" (Jona Lasinio et al., 2012; Banerjee et al., 2004), which is due to the computational costs of $\mathcal{O}(n^3)$ to perform matrix algebra operations with $n \times n$ dense covariance matrices (whose dimension is given by the number of observations at all spatial locations and time points).

In contrast, as introduced in Section 2, GMRFs are characterised by sparse precision matrices; this feature allows to implement computationally efficient numerical methods, especially for fast matrix factorization (Rue and Held, 2005). For a GMRF model in $\mathbb{R}^2$ the computational cost is typically $\mathcal{O}(n^{3/2})$, which is a significant speed up compared to $\mathcal{O}(n^3)$ of the GF. Moreover, Bayesian inference involving spatial GMRFs can be performed employing the INLA approach introduced in Section 3.

In this section we sketch the basics of the SPDE approach and we refer to Lindgren et al. (2011) for a complete description and the proofs of the results. Applications of SPDE for geostatistical data can be found in Simpson et al. (2012a,b), Bolin (2012), Cameletti et al. (2011b) and Simpson et al. (2011).

Consider a simple setting for geostatistical data where for the i-th spatial point location the observation y_i is modelled as[4]

$$y_i \sim \text{Normal}\left(\eta_i, \sigma_e^2\right) \tag{9}$$

where σ_e^2 is the variance of the zero mean measurement error e_i which is supposed to be independent on e_j with $i \neq j$. The response mean is defined as

$$\eta_i = \alpha + \sum_{m=1}^{M} \beta_m x_{mi} + \xi_i \tag{10}$$

---

[4]Here we consider the case of normally distributed data, but this is not a requirement as INLA and the SPDE approach can deal with non Gaussian responses. However, it is worth to note that in the Gaussian case, the INLA calculations are exact and the only approximation is the numerical integration required for computing $\tilde{p}(\boldsymbol{\psi} \mid \boldsymbol{y})$ in (3).

where ξ_i is the i-th realisation of the latent GF $\xi(s)$ with Matérn spatial covariance function defined in (1). In the geostatistics literature, the term $\alpha + \sum_{m=1}^{M} \beta_m x_{mi}$ is often referred to as large-scale component, while the measurement error variance σ_e^2 is known as nugget effect (see Cressie, 1993). With respect to the linear predictor introduced in (2), the function $f_i(\cdot)$ is represented by the spatially structured term ξ_i. The key idea of the SPDE approach consists in defining the continuously indexed Matérn GF $\xi(s)$ as a discrete indexed GMRF by means of a basis function representation defined on a triangulation of the domain $\mathcal{D}$

$$\xi(s) = \sum_{g=1}^{G} \varphi_g(s) \tilde{\xi}_g \ . \tag{11}$$

Here G is the total number of vertices in the triangulation, $\{\varphi_g\}$ is the set of basis functions and $\{\tilde{\xi}_g\}$ are zero-mean Gaussian distributed weights. For computational reasons as explained in Lindgren et al. (2011), the basis functions are chosen to be piecewise linear on each triangle, i.e. φ_g is 1 at vertex g and 0 elsewhere. Notice that we use the formal notation $\xi(s)$ in the left-hand side of (11) since SPDE[5] provides a representation of the whole spatial process (defined for any point s) that varies continuously in the considered domain $\mathcal{D}$.

An illustration of the SPDE approach is given in Figure 4, which displays a continuously indexed spatial random field and the corresponding finite element representation with piecewise linear basis functions over a triangulated mesh. Lindgren et al. (2011) show that the vector of basis weights $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_G)'$ is a GMRF with sparse precision matrix $\boldsymbol{Q}_{\tilde{\xi}}$ depending on the Matérn covariance function parameter κ and variance $\sigma_{\mathcal{C}}^2$, for $\alpha = 1, 2, 3, \ldots$ where $\alpha = \lambda + 1$ (recall that λ is the smoothness parameter).

Given the GF representation provided by the SPDE method, we can rewrite the linear predictor of (10) as

$$\eta_i = \alpha + \sum_{m=1}^{M} \beta_m x_{mi} + \sum_{g=1}^{G} \tilde{\boldsymbol{A}}_{ig} \tilde{\boldsymbol{\xi}} \tag{12}$$

---

[5]The terminology SPDE is related to the linear fractional stochastic partial differential equation reported in Equation (2) of Lindgren et al. (2011) whose (only stationary) exact solution is given by a GF with Matèrn covariance function. This exact solution is then approximated using the finite element representation of (11).

where $\tilde{\boldsymbol{A}}$ is the sparse $n \times G$ matrix that maps the GMRF $\tilde{\boldsymbol{\xi}}$ from the n observation locations to the G triangulation nodes. Note that in `R-INLA` this kind of model can be easily implemented specifying `model=spde` in the `f(·)` term of the `formula` definition.

The next two sections are dedicated to the implementation of a spatial and a spatio-temporal geostatistical model in `R-INLA` providing some details about the SPDE functions. At the moment `R-INLA` implements the SPDE approach for $0 < \alpha \leq 2$; in our models we have specified $\alpha = 2$ which corresponds to a smoothness parameter λ equal to 1, thus to a second-order conditional autoregressive structure. Anticipating an `R-INLA` feature for managing geostatistical data, we rewrite here (9) and (10) in matrix notation as

$$\boldsymbol{y} \quad \sim \quad \mathrm{N}\left(\boldsymbol{\eta}, \sigma_e^2 \boldsymbol{I}_n\right) \tag{13}$$

$$\boldsymbol{\eta} \quad = \quad \mathbf{1}\alpha + \boldsymbol{X}'\boldsymbol{\beta} + \tilde{\boldsymbol{A}}\tilde{\boldsymbol{\xi}} \tag{14}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is the observation vector, $\boldsymbol{I}_n$ is a n-dimensional diagonal matrix, $\mathbf{1}$ is a vector of ones and $\boldsymbol{X}$ is the $M \times n$ matrix of covariates. Moreover the term $\boldsymbol{A} = \{\mathbf{1}, \boldsymbol{X}, \tilde{\boldsymbol{A}}\}$ is called observation matrix. According to the notation introduced in Section 3, in this case the vector of parameters is defined as $\boldsymbol{\theta} = \{\tilde{\boldsymbol{\xi}}, \alpha, \boldsymbol{\beta}\}$ with hyper-parameters vector $\boldsymbol{\psi} = (\sigma_e^2, \kappa, \sigma_{\tilde{C}}^2)$.

4.1. INLA/SPDE for spatial geostatistical data: Swiss rainfall data

One of the primary objective of geostatistical modeling is the prediction of the considered phenomenon at unsampled locations conditionally on the observed data and available covariates (i.e. kriging, see Gelfand et al., 2010). To illustrate how to perform spatial prediction using INLA and the SPDE approach, we consider rainfall measurements (in 10th of mm) taken on the 8th of May 1986 at 467 locations in Switzerland. The rainfall data are part of the `sic` data set in the `geoR` library (Ribeiro and Diggle, 2001) which provides also the spatial coordinates and the elevation value (in km) for each location.

In order to make the distribution of the rainfall data approximately Normal, we use a root square transformation; the transformed values are depicted in Figure 5 (a). Moreover, following the guidelines described in Dubois (1998), we use the 100 locations marked with bullets in Figure 5 (a) for estimation purposes and we retain the remaining 367 stations (marked with triangles) for model validation, i.e. we predict rainfall in the validation sites and evaluate through indexes the model predictive performance. Finally, we

estimate the rain field on a regular grid covering Switzerland with the same resolution of the elevation surface available from the `sic97` data set in the `gstat` package (Pebesma, 2004) and depicted in Figure 5 (b). In particular, the elevation map is named `demstd` and is composed by 376×253 grid points.

In `R-INLA` the first step required to run the geostatistical spatial model introduced in Section 4 with only one covariates ($M = 1$ represented by elevation), is the triangulation of the considered spatial domain. We use the `inla.mesh.create.helper` specifying the spatial coordinates (`est.coord`) of the 100 stations used for estimation and the region borders (`sic.borders`) required to define the outer domain:

```
mesh =  inla.mesh.create.helper(
        points=est.coord, points.domain=sic.borders,
        offset=c(5, 20), max.edge=c(40,100), min.angle=c(21,21))
```

The `inla.mesh.create.helper` performs a constrained refined Delaunay triangulation for a set of spatial locations: firstly the triangle vertices are placed at the observation locations and then further vertices are added in order to satisfy triangulation quality constraints (see Lindgren et al., 2011 and references therein). To this regard the `offset` in the `inla.mesh.create.helper` function defines how much the domain should be extended in the inner and outer domains (characterised by small and large triangles, respectively), while `max.edge` and `min.angle` set the triangle structure. Depending on the values chosen for `inla.mesh.create.helper` arguments, the total number of vertices changes with a trade-off between the accuracy of the GMRF representation and the computational costs.

With the setting used above we obtain a mesh with $G = 289$ vertices, which can be accessed in the `R` terminal by typing `mesh$n` and is displayed in Figure 6. Given the mesh, we create the `spde` model object, to be used later in the specification of the `f(·)` term in the `R-INLA formula`, with

```
spde = inla.spde2.matern(mesh=mesh)
```

We exploit now the helper function `inla.stack` which takes care of building the necessary matrices required by the SPDE approach and of combining the data, the observation matrix $\boldsymbol{A}$ and the linear predictor $\boldsymbol{\eta}$, introduced in (13) and (14); some details about the usage of the `inla.stack` function can be found also in Cameletti et al. (2011b). Before employing `inla.stack`, we create the object `A.est` which corresponds to $\tilde{\boldsymbol{A}}$

```
A.est =  inla.spde.make.A(mesh, loc=est.coord)
```

20

and is a 100×289 sparse matrix that extracts the values of the latent spatial field at the observation locations. Moreover, we generate the required vectors of indices

```
field.indices = inla.spde.make.index("field", n.mesh=mesh$n)
```

with `field.indices` being a list whose first component is called `field` and contains the spatial vertex indices (i.e, the sequence of integer number from 1 to $G = 289$). Finallly, we call the `inla.stack` function that takes in input the data (`data`), an identification string (`tag`) and the components of the observation matrix (`A`) and of the linear predictor (`effects`), combined together in `list`-type objects:

```
stack.est = inla.stack(data = list(rain=est.data),
            A = list(A.est, 1),
            effects = list(c(field.indices, list(Intercept=1)),
                    list(Elevation=est.elevation)),
            tag="est")
```

Note that each term in `A` has its own linear predictor component in the `effects` object so that, for example, `A.est` is paired with the list composed by `field.indices` and `Intercept=1` (this may seem a little strange but it is due to how the SPDE related functions are internally coded). The elevation covariate is included in `A` by means of `1` - which has to be interpreted as an identity matrix - and the corresponding altitude values (`est.elevation`) are then provided as a `list` object in the `effects` term.

Similarly, we create the corresponding objects `inla.val` and `stack.val` for the 367 validation stations with the only difference that, since we are interested in prediction, we specify `data=list(rain=NA)` in the `inla.stack` function. For rainfall prediction in the $376 \times 253 = 95128$ grid points, we create the `A.pred` and `stack.pred` objects as follows

```
A.pred = inla.spde.make.A(mesh)
stack.pred = inla.stack(data = list(rain=NA),
              A = list(A.pred),
              effects = list(c(field.indices, list(Intercept=1))),
              tag="pred")
```

where, for computational reasons, we consider the mesh locations only and do not include elevation in the linear predictor. This means that later we will have to move from the mesh to the grid (with a projection) and to add back the covariate term.

Finally, we combine all the data, effects and observation matrices using the command

```
stack = inla.stack(stack.est, stack.val, stack.pred)
```

In the R-INLA formula we include the spde model object named field and defined before as well as the Elevation covariate; moreover, note that, due to the way inla.stack works, we need to specify an explicit Intercept term and remove the automatic intercept with -1.

```
formula <- rain ~ -1 + Intercept + Elevation + f(field, model=spde)
```

Finally, we can run the specified model calling the inla function as follows:

```
mod =  inla(formula, data=inla.stack.data(stack, spde=spde),
       family="gaussian",
       control.predictor=list(A=inla.stack.A(stack), compute=TRUE))
```

where the functions inla.stack.data and inla.stack.A simply extract the required data and the observation matrix from the stack object. The option compute=TRUE is required to obtain the marginal distributions for the linear predictor.

We retrieve the posterior summary statistics of the fixed effects α and β from the object mod$summary.fixed, while the posterior marginal of the precision $\tau_e = 1/\sigma_e^2$ is included in the list mod$marginals.hyperpar. If we are interested in the variance σ_e^2, we employ the function inla.emarginal for computing the expected value of the (reciprocal) transformation of the posterior marginal distribution. The results on the parameters of the Matèrn spatial covariance function can be obtained typing

```
mod.field = inla.spde2.result(mod, name="field", spde)
```

where the string name refers to the name of the spde effect used in the inla formula.

Applying the suitable transformations through the inla.emarginal function as described in Cameletti et al. (2011b), we obtain the posterior estimates for the spatial variance $\sigma_{\hat{C}}^2$ and for the range r. All the relevant posterior estimates are reported in the upper part of Table 3. As the elevation parameter β is not significant, we implement also the model without elevation and use the DIC as a model selection criterion. The DIC values reported in Table 3 are almost identical so we select the model without elevation (note that the posterior estimates for α and r do not change considerably between

the two models). With a posterior mean of 62 km for the range, we can conclude that the data are characterised by a medium spatial correlation (the maximum distance between coordinates is equal to 293 km).

	Mean	Sd	2.5%	50%	97.5%
	With elevation (DIC = -571.1897)				
α	12.084	1.577	8.801	12.134	15.085
β	0.031	0.722	-1.396	0.035	1.442
r	61.479	10.482	42.339	61.044	83.303
	Without elevation (DIC = -571.2634)				
α	12.109	1.420	9.150	12.147	14.862
r	61.673	10.384	42.709	61.240	83.299

Table 3: Posterior estimates (mean, standard deviation (sd) and quantiles) and DIC for the Swiss rainfall geostatistical model with and without elevation covariate.

We focus now on the prediction in the 367 validation stations (this case was previously identified with the string `tag="val"`). We first type

```
index.val = inla.stack.index(stack,"val")$data
```

in order to retrieve, from the full `stack` object, the indices identifying the validation stations (which are stored in the `data` component of the resulting list). Given `index.val` we extract the posterior summaries (mean and standard deviation) for the linear prediction $\boldsymbol{\eta}$ (on the square root scale) as follows

```
lp.mean.val = mod$summary.linear.predictor[index.val,"mean"]
lp.sd.val = mod$summary.linear.predictor[index.val,"sd"]
```

It is then straightforward to compare observed and predicted values (represented by the posterior mean `lp.mean.val`) and to compute predictive performance statistics. For example, the root mean square error is equal to 2.30 and the Pearson correlation coefficient is 0.86, which denotes a good correlation between observed and predicted values.

Prediction on the regular grid (here defined by a `data.frame` object named `pred.grid` with $376 \times 253 = 95128$ rows and two columns with grid coordinates) requires to create a linkage between the mesh and the grid, as we anticipated previously. This can be done using the following command:

```
proj_grid =  inla.mesh.projector(mesh, xlim=range(pred.grid[,1]),
                        ylim=range(pred.grid[,2]), dims=c(376,253))
```

Then, as done previously for the validation sites, we extract the linear predictor values on the mesh

```
index.pred = inla.stack.index(stack,"pred")$data
lp.mean.pred =  mod$summary.linear.predictor[index.pred, "mean"]
lp.sd.pred =  mod$summary.linear.predictor[index.pred, "sd"]
```

and project it from the mesh to the grid

```
lp.mean.grid = inla.mesh.project(proj_grid, lp.mean.pred)
lp.sd.grid = inla.mesh.project(proj_grid, lp.sd.pred)
```

The map of the smooth rainfall posterior mean (on the square root scale) and of the prediction standard error are shown in Figure 7. The comparison of the prediction map with the plot reported in Figure 5 (b) leads to the conclusion that the considered geostatistical model is able to reproduce quite well the spatial pattern of the rainfall data.

4.2. INLA/SPDE for for spatio-temporal geostatistical data: PM_{10} air pollution in Piemonte region

We extend the purely spatial case described in the previous section to a spatio-temporal model for particulate matter concentration (PM_{10} in $\mu g/m^3$) measured in the region of Piemonte (Northern Italy) during October 2005 - March 2006 by a monitoring network composed by 24 stations. Cameletti et al. (2011a) provide a complete description of the PM_{10} data as well as of some covariates available at the station and grid level (provided by ARPA Piemonte, Finardi et al. 2008), such as daily maximum mixing height (`HMIX`, in m), daily total precipitation (`PREC`, in mm), daily mean wind speed (`WS`, in m/s), daily mean temperature (`TEMP`, in $° K$), daily emission rates of primary aerosols (`EMI`, in g/s), altitude (`A`, in m) and spatial coordinates (`UTMX` and `UTMY` in km).

We illustrate how to predict air pollution for a given day in all the region, also where no monitoring stations are displaced. In addition, we describe how to get a map for the probability of exceeding the 50 $\mu g/m^3$ threshold fixed by the European Community for health protection. Note that this case study has already been described in Cameletti et al. (2011b), but we present it again in order to illustrate a variant in the SPDE code for producing the probability map of exceeding the fixed threshold.

Let y_{it} denote the logarithm of the PM_{10} concentration measured at station located at site s_i $(i = 1, \ldots, n)$ and day $t = 1, \ldots, T$. We assume the following distribution for the observations

$$y_{it} \sim \text{Normal}(\eta_{it}, \sigma_e^2)$$

with

$$\eta_{it} = \sum_{m=1}^{M} \beta_m x_{mi} + \omega_{it}$$

where $\sum_{m=1}^{M} \beta_m x_{mi}$ is the large-scale component including meteorological and geographical covariates, and σ_e^2 is the variance of the measurement error defined by a Gaussian white-noise process, both serially and spatially uncorrelated. The term ω_{it} is the realisation of the latent spatio-temporal process (i.e. the true unobserved level of pollution) which changes in time with first order autoregressive dynamics with coefficient a and spatially correlated innovations, given by

$$\omega_{it} = a\omega_{i(t-1)} + \xi_{it}. \tag{15}$$

In (15), we set $t = 2, \ldots, T$ and $|a| < 1$, and derive ω_{i1} from the stationary distribution $\text{Normal}(0, \sigma_{\mathcal{C}}^2/(1-a^2))$. Moreover, ξ_{it} is a zero-mean GF, is assumed to be temporally independent and is characterised by the following spatio-temporal covariance function

$$\text{Cov}\,(\xi_{it}, \xi_{ju}) = \begin{cases} 0 & \text{if} & t \neq u \\ \sigma_{\mathcal{C}}^2 \mathcal{C}(\Delta_{ij}) & \text{if} & t = u \end{cases} \tag{16}$$

for $i \neq j$, with $\mathcal{C}(\Delta_{ij})$ denoting the Matèrn spatial covariance function defined in (1). Such a model is widely used in the air quality literature thanks to its flexibility in modeling the effect of relevant covariates (i.e. meteorological and geographical variables) as well as time and space dependence (e.g. Cocchi et al., 2007; Cameletti et al., 2011a; Sahu, 2012; Fassò and Finazzi, 2011). The main drawback of this formulation is related to the computational costs required for model parameter estimation and spatial prediction when MCMC methods are used, especially in case of massive spatio-temporal datasets. Here we show how to overcome this computational challenge using the SPDE approach by representing the Màtern spatio-temporal GF as a GMRF (see Cameletti et al., 2011b for more details).

To implement this model in `R-INLA`, we need to define the triangulation of Piemonte using the `inla.mesh.create.helper` function, as described in the

previous section. After creating an object named `mesh` including $G = 142$ vertices, we define the SPDE object with `spde=inla.spde2.matern(mesh=mesh)`. The next step requires to employ the `inla.stack` function to combine the data with the observation matrix and linear predictor components; this is a slightly more complex task here, since we have to consider both spatial and temporal indexing. Let `Piemonte_data` be the data frame containing all the relevant data; for estimation purposes create the `A.est` object with

```
A.est = inla.spde.make.A(mesh,
    loc=as.matrix(coordinates[Piemonte_data$Station.ID, c("UTMX","UTMY")]),
    group=Piemonte_data$time,
    n.group=n_days)
```

where the option `group` specifies that we have 24 measurements for each of the $T = 182$ days (included as `n_days` in the code). Then we generate the spatial and temporal indexes typing

```
field.indices = inla.spde.make.index("field",n.mesh=mesh$n,n.group=n_days)
```

and then we combine all the relevant objects with

```
stack.est = inla.stack(data=list(logPM10=Piemonte_data$logPM10),
                A=list(A.est, 1),
                effects=list(c(field.indices, list(Intercept=1)),
                        list(Piemonte_data[,3:10])),
                tag="est")
```

where `Piemonte_data[,3:10]` refers to the columns containing the covariate values. In a similar way, we create `A.pred` and `stack.pred` for the $56 \times 72 = 4032$ grid points used for spatial prediction:

```
A.pred = inla.spde.make.A(mesh, loc=as.matrix(Piemonte_grid),
        group=i_day, n.group=n_days)

stack.pred = inla.stack(data=list(logPM10=NA),
            A=list(A.pred,1),
            effects=list(c(field.indices,list(Intercept=1)),
                    list(covariate_matrix_std)),
            tag="pred")
```

where `Piemonte_grid` and `covariate_matrix_std` contain the coordinates and the (standardized) covariate values for all the grid locations and the selected day (30/01/2006), respectively. Note that, differently from Section 4.1 and Cameletti et al. (2011b), we are including at this stage (and not

26

after the estimation step) the grid relevant information. This means that the output of the `inla` function will provide directly the estimate of the linear predictor (including covariates) at the grid level.

Finally we create the complete `stack` object with the following code

```
stack = inla.stack(stack.est, stack.pred)
```

and define the R-INLA formula

```
formula <- (logPM10 ~ -1 + Intercept + A + UTMX + UTMY + WS + TEMP +
        HMIX + PREC + EMI + f(field, model=spde,
        group=field.group, control.group=list(model="ar1")))
```

that includes an explicit intercept and all the meteorological and geographical covariates. Moreover, using the options `group` and `control.group`, we specify in the `f(·)` term that at each time point the spatial locations are linked by the `spde` model object, while across time, the process evolves according to an AR(1) process.

For computational reasons, it may be useful to run this model calling the `inla` function twice. We first compute only the hyper-parameters modes (se theoretical details in Section 3) only for the `stack.est` object by setting `compute=FALSE` in the `control.preditor` argument:

```
mod.mode = inla(formula,
      data=inla.stack.data(stack.est, spde=spde),
      family="gaussian",
      control.predictor=list(A=inla.stack.A(stack.est), compute=FALSE)
```

At the second step we perform the linear predictor estimation on the whole grid specifying the full object `stack` and using the mode computed previously (see the specification of the `control.mode` argument):

```
mod =  inla(formula,
      data=inla.stack.data(stack, spde=spde),
      family="gaussian",
      control.predictor=list(A=inla.stack.A(stack), compute=TRUE),
      control.mode=list(theta=mod.mode$mode$theta, restart=FALSE)
```

As shown in the previous sections, we can extract the posterior summary statistics for $\boldsymbol{\beta}$, $1/\sigma_e^2$ and a from the objects `mod$summary.fixed` and `mod$summary.hyperpar`, while posterior estimates for $\sigma_{\mathcal{C}}^2$ and r can be obtained applying the `inla.spde2.result` function — see Cameletti et al. (2011b) for more details and the relevant results.

Here we focus on the prediction of the (smooth, i.e. without the nugget effect) air pollution field for the selected day. This task is performed simply by extracting the posterior mean of the linear predictor - which is available for all the grid locations - from `mod$summary.linear.predictor` and reshaping it properly in accordance with the grid size.

```
index.pred = inla.stack.index(stack,"pred")$data
lp_grid_mean = matrix(mod$summary.linear.predictor[index.pred, "mean"],
                56, 72, byrow=T)
```

The resulting map (on the logarithmic scale) is shown in Figure 8 (left).

Analogously, we can retrieve the posterior marginal distribution of the linear predictor and, through the built-in function `inla.pmarginal` employed in Section 3.2 and 3.3, we can obtain the map of the posterior probability of exceeding the fixed threshold, presented in Figure 8 (right).

5. Discussion

In this paper we have provided a tutorial on the use of methods based on Integrated Nested Laplace Approximation for spatial and spatio-temporal models. While these models are very popular in applied research, especially in epidemiology, their general complexity remains, potentially, a fundamental issue for their implementation, particularly within the Bayesian approach. The INLA approach is in general able to provide reliable estimations in lower computational time than their corresponding MCMC-based estimations.

One of the fundamental differences between MCMC and INLA methods is that the former provide (asymptotically) exact inference, while the latter give, by definition, an approximation to the relevant posterior distributions. In many applied cases INLA performs just as well as its MCMC counterparts, especially when the latter are considered in their standard implementations. This is particularly relevant in presence of large datasets; as discussed earlier, specifically in the case of geostatistical data, the use of SPDE algorithms produce massive savings in computational times and allows the user to work with relatively complex models in an efficient way. INLA and SPDE could also help in solving the change-of-support issue, typically arising when dealing with data characterised by different spatial supports, e.g. air pollution data available at the point level combined with a health outcome available as aggregated counts of deaths/disease at the areal level — see chapter 29 in Gelfand et al. (2010). Finally, INLA (and specifically its `R` implementation) covers a wide set of problems that can be tackled with relatively

28

standard programming, which generally facilitates the practitioner's work. In fact, while most of the commands are similar to those applied in standard `R` routines (e.g. `lm` or `glm`), a wealth of options can be specified within the `R-INLA` functions, that allow the user to select different model specifications; see Martins et al. (2012) for new features.

Because of its recent inception, INLA is less established than MCMC methods (although we acknowledge a flurry of activity in the development of new MCMC algorithms, e.g. Girolami and Calderhead, 2011; and Hoffman and Gelman, 2011). Consequently, its development is still ongoing, particularly with respect to some more advanced features (e.g. the SPDE module described in Section 4). At the same time, however, it is important to notice that the increasing popularity of INLA is generating a number of contributed add-ons able to extend the built-in facilities of the `R` package. Given these characteristics, we consider INLA as a valuable addition to the Bayesian statistician's toolkit.

6. Acknowledgments

References

Baio, G., 2012. Bayesian Methods in Health Economics. CRC Chapman and Hall.

Banerjee, S., 2004. Revisiting Spherical Trigonometry with Orthogonal Projectors. The Mathematical Association of America's College Mathematics Journal. 35, 375–381.

Banerjee, S., Carlin, B., Gelfand, A., 2004. Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability. Chapman and Hall, New York.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., Songini, M., 1995. Bayesian analysis of space-time variation in disease risk. Statistics in Medicine 14 (21-22), 2433–2443.

Bernardo, J., Smith, A., 2000. Bayesian Theory. Wiley-Blackwell.

Berry, S., Carlin, B., Lee, J., M uller, P., 2011. Bayesian Adaptive Methods for Clinical Trials. CRC Chapman and Hall.

Besag, J., York, J., Mollie, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistiscal Mathematics 43, 1–59.

Blangiardo, M., Cameletti, M., 2013. Bayesian Spatio and Spatio-Temporal Models with R-INLA. Wiley.

Bolin, D., 2012. Models and methods for random fields in spatial statistics with computational efficiency from Markov properties. Ph.D. thesis, Lund University.

Brooks, S., Gelman, A., Jones, G., Meng, X. (Eds.), 2011. Handbook of Markov Chain Monte Carlo. CRC Press, Taylor & Francis Group.

Cameletti, M., Ignaccolo, R., Bande, S., 2011a. Comparing spatio-temporal models for particulate matter in Piemonte. Environmetrics 22, 985–996.

Cameletti, M., Lindgren, F., Simpson, D., Rue, H., 2011b. Spatio-temporal modeling of particulate matter concentration through the spde approach. AStA Advances in Statistical Analysis, 1–23Doi:10.1007/s10182-012-0196-3.

Clark, J., 2005. Why environmental scientists are becoming bayesians. Ecology Letters 8 (1), 2–14.

Clark, J., Gelfand, A. (Eds.), 2006. Hierarchical Modeling for the Environmental Sciences. Statistical Methods and Applications. Oxford University Press, New York.

Clayton, D., 1996. Generalised linear mized models. In: Gilks, W., Richardson, S., Spiegelhalter, D. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, pp. 275 – 301.

Cocchi, D., Greco, F., Trivisano, C., 2007. Hierarchical space-time modelling of PM_{10} pollution. Atmospheric Environment 41, 532–542.

Congdon, P., 2007. Bayesian Statistical Modelling. John Wiley and Sons, Ltd.

Cressie, N., 1993. Statistics for Spatial Data. Wiley.

Cressie, N., Huang, H., 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association 94 (448), 1330–1340.

Diggle, P., Ribeiro, J., 2007. Model-Based Geostatistics. Springer.

Dubois, G., 1998. Spatial interpolation comparison 97: foreword and introduction. Journal of Geographic Information and Decision Analysis 2, 1–10.

Dunson, D., 2001. Commentary: Practical advantages of bayesian analysis of epidemiologic data. American journal of Epidemiology 153 (12), 1222–1226.

Fassò, A., Finazzi, F., 2011. Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. Environmetrics 22 (6), 735–748.

Finardi, S., De Maria, R., D'Allura, A., Cascone, C., Calori, G., Lollobrigida, F., 2008. A deterministic air quality forecasting system for Torino urban area, Italy. Environmental Modelling and Software 23 (3), 344–355.

Gelfand, A., Diggle, P., Fuentes, M., Guttorp, P. (Eds.), 2010. Handbook of Spatial Statistics. Chapman & Hall.

Girolami, M., Calderhead, B., 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society, Series B 73 (2), 1–37.

Gneiting, T., 2002. Nonseparable, stationary covariance functions for spaceâĂŞtime data. Journal of the American Statistical Association 97 (458), 590–600.

Gneiting, T., Genton, M., Guttorp, P., 2006. Statistical Methods for Spatio-temporal systems. In: Finkenstädt, B., Held, L., Isham, V. (Eds.), Statistical Methods for Spatio-temporal systems. CRC Press, Chapmann and Hall, pp. 151–175.

Greenland, S., 2006. Bayesian perspectives for epidemiological research: I. foundations and basic methods. International journal of Epidemiology 35, 765–775.

Guttorp, P., Gneiting, T., 2006. Studies in the history of probability and statistics xlix On the matèrn correlation family. Biometrika 93 (4), 989–995.

Harvill, J., 2010. Spatio-temporal processes. Wiley Interdisciplinary Reviews: Computational Statistics 2 (3), 375–382.

Hoffman, M., Gelman, A., 2011. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. eprint arXiv:1111.4246.

Jackman, S., 2009. Bayesian Analysis for the Social Sciences. Wiley-Blackwell.

Jewell, C., Kypraios, T., Neal, P., Roberts, G., 2009. Bayesian analysis for emerging infectious diseases. Bayesian Analysis 4 (3), 465–496.

Jona Lasinio, G., Mastrantonio, G., Pollice, A., 2012. Discussing the "big n problem". Statistical Methods and Applications, 1–16.

Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine 19 (17-18), 2555–2567.

Lawson, A., 2009. Bayesian Disease Mapping. Hierarchical Modeling in Spatial Epidemiology. CRC Press.

Li, Y., Brown, P., Rue, H., al Maini, M., Fortin, P., 2012. Spatial modelling of lupus incidence over 40 years with changes in census areas. Journal of the Royal Statistical Society: Series C (Applied Statistics) 61 (1), 99–115.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). J. R. Statist. Soc. B 73 (4), 423–498.

Lindley, D., 2006. Understanding Uncertainty. Wiley-Blackwell.

Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. The BUGS Book: A Practical Introduction to Bayesian Analysis. CRC Press.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique and future directions. Statistics in Medicine 28(25), 3049–3067.

Martino, S., Aas, K., Lindqvist, O., Neef, L., Rue, H., 2011. Estimating stochastic volatility models using integrated nested laplace approximations. The European Journal of Finance 17 (7), 487–503.

Martino, S., Rue, H., 2010. Implementing Approximate Bayesian Inference using Integrated Nested Laplace Approximation: a manual for the `inla` program.
URL \url{http://www.math.ntnu.no/~hrue/GMRFsim/manual.pdf}

Martins, G., Simpson, D., Lindgren, F., Rue, H., 2012. Bayesian computation with INLA: new features. Norwegian University of Science and Technology Report.

Pascutto, C., Wakefield, J., Best, N., Richardson, S., Bernardinelli, L., Staines, A., Elliott, P., 2000. Statistical issues in the analysis of disease mapping data. Statistics in Medicine 19 (17-18), 2493–519.

Paul, M., Riebler, A., Bachmann, L. M., Rue, H., Held, L., 2010. Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested laplace approximations. Statistics in Medicine 29 (12), 1325–1339.

Pebesma, E., 2004. Multivariable geostatistics in s: the gstat package. Computers and Geosciences 30, 683–691.

Ribeiro, J., Diggle, P., 2001. geoR: A package for geostatistical analysis. R-NEWS 1 (2), http://cran.r-project.org/doc/Rnews.

Richardson, S., Thomson, A., Best, N., Elliott, P., 2004. Interpreting posterior relative risk estimates in disease-mapping studies. Environmental Health Perspectives 112 (9), 1016–1025.

Riebler, A., Held, L., Rue, H., 2012. Estimation and extrapolation of time trends in registry data – borrowing strength from related populations. Annals of Applied Statistics 6 (1), 304–333.

Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods. Springer.

Roos, M., Held, L., 2011. Sensitivity analysis in bayesian generalized linear mixed models for binary data. Bayesian Analysis 6 (2), 259–278.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields. Theory and Applications. Chapman & Hall.

Rue, H., Martino, S., 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. Journal of Statistical Planning and Inference 137, 3177–3192.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B 71 (2), 1–35.

Ruiz-Cárdenas, R., Krainski, E., Rue, H., 2012. Direct fitting of dynamic models using integrated nested laplace approximations inla. Computational Statistics & Data Analysis 56 (6), 1808 – 1828.

Sahu, S., 2012. 16 - hierarchical Bayesian models for space-time air pollution data. In: Subba Rao, T., Subba Rao, S., Rao, C. (Eds.), Time Series Analysis: Methods and Applications. Vol. 30 of Handbook of Statistics. Elsevier Publishers, Holland, pp. 477 – 495.

Schrödle, B., Held, L., 2011a. A primer on disease mapping and ecological regression using INLA. Computational Statistics 26, 241–258.

Schrödle, B., Held, L., 2011b. Spatio-temporal disease mapping using INLA. Environmetrics 22 (6), 725–734.

Schrödle, B., Held, L., Riebler, A., Danuser, J., 2011. Using integrated nested laplace approximations for the evaluation of veterinary surveillance data from switzerland: a case-study. Journal of the Royal Statistical Society: Series C (Applied Statistics) 60 (2), 261–279.

Simpson, D., Illian, J., Lindgren, F., Sørbye, S., Rue, H., 2011. Going off grid: Computationally efficient inference for log-gaussian cox processes. ArXiv e-prints.

Simpson, D., Lindgren, F., Rue, H., 2012a. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. Environmetrics 23 (1), 65–74.

Simpson, D., Lindgren, F., Rue, H., 2012b. Think continuous: Markovian Gaussian models in spatial statistics. Spatial Statistics 1, 16–29.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B Series B 64 (4), 583–639.

Tierney, L., Kadane, J., 1986. Accurate approximations for posterior moments and marginal densities. Journal of the Americal Statistical Association 393 (81), 82–86.

Wikle, C., 2003. Hierarchical models in environmental science. International Statistical Review 71 (2), 181–199.
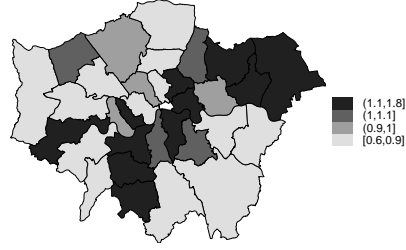
(a) Distribution of the borough specific relative risks of suicides $\zeta_i = \exp(\upsilon_i + \nu_i)$ in the disease mapping model
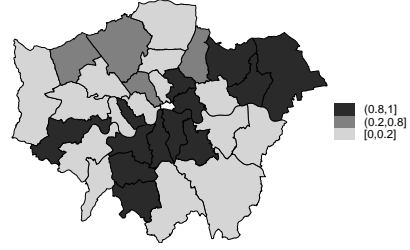
(b) Distribution of the borough specific posterior probability $p(\zeta_i > 1 \mid \boldsymbol{y})$ in the disease mapping model
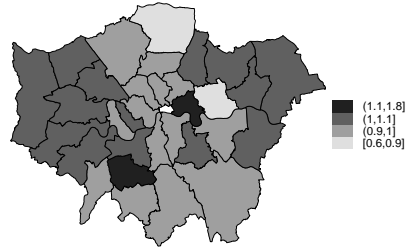
(c) Distribution of the borough specific relative risks of suicides $\zeta_i = \exp(\upsilon_i + \nu_i)$ in the ecological regression model
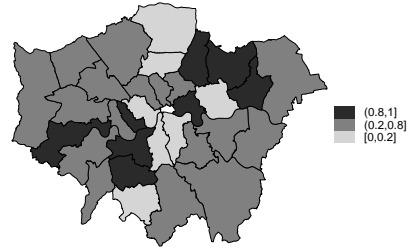
(d) Distribution of the borough specific posterior probability $p(\zeta_i > 1 \mid \boldsymbol{y})$ in the ecological regression model

Figure 1: Borough specific relative risks and posterior probabilities.
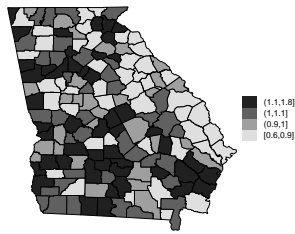
(a) Map of the spatial pattern of disease risk $\zeta_i = \exp(\upsilon_i + \nu_i)$
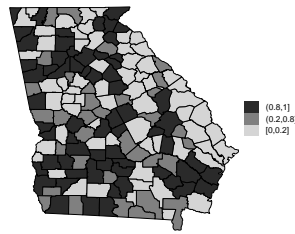
(b) Map of the uncertainty for the spatial effect $\zeta_i : p(\zeta_i > 1 \mid \boldsymbol{y})$
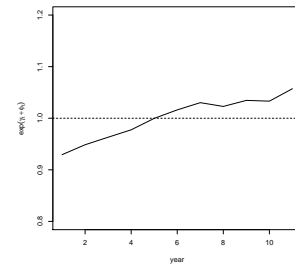
(c) Temporal trend of low birth weight risk in Georgia

Figure 2: Spatial and temporal effects.
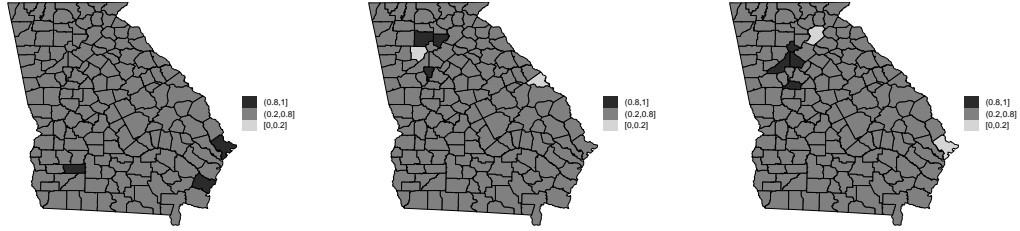
(a) Posterior probability for spatio-temporal interaction: year 2001

(b) Posterior probability for spatio-temporal interaction: year 2004

(c) Posterior probability for spatio-temporal interaction: year 2010

Figure 3: Posterior probability for the space-time interaction: years 2001, 2004 and 2010 and 159 counties of Georgia.
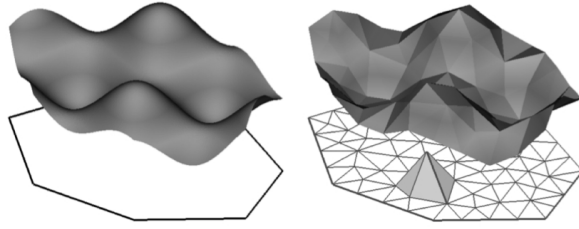
Figure 4: Example of a spatial continuous random field and the corresponding basis function representation according to (11).
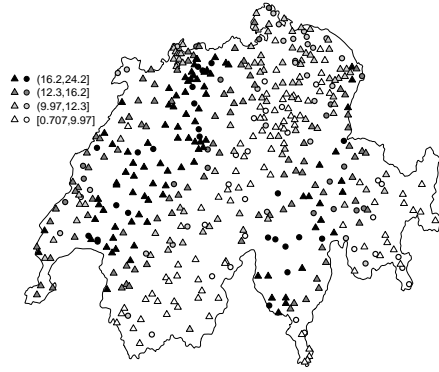
(a) Rainfall data collected on 8th May 1986 at 467 locations in Switzerland. The bullets denote the 100 estimation stations and the triangles are used for the 367 stations retained for model validation.
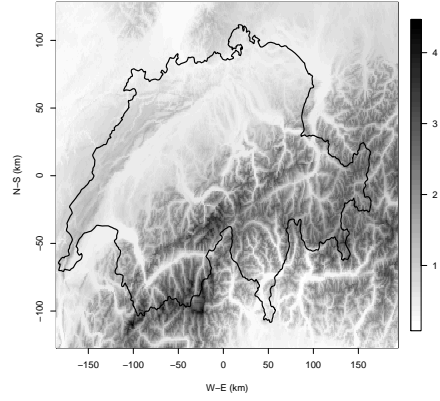
(b) Map of elevation (in km) defined on a 376×253 regular grid covering Switzerland.

Figure 5: Swiss rainfall data (on the root square scale) and elevation.

Constrained refined Delaunay triangulation
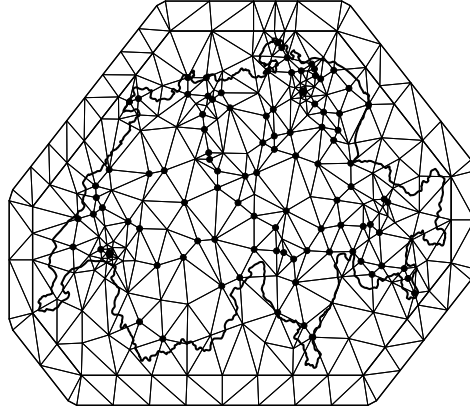
Figure 6: The Switzerland triangulation with 289 vertices and black dots denoting the 100 stations used for estimation and included in the mesh.

(a) Rainfall posterior mean (square root scale)

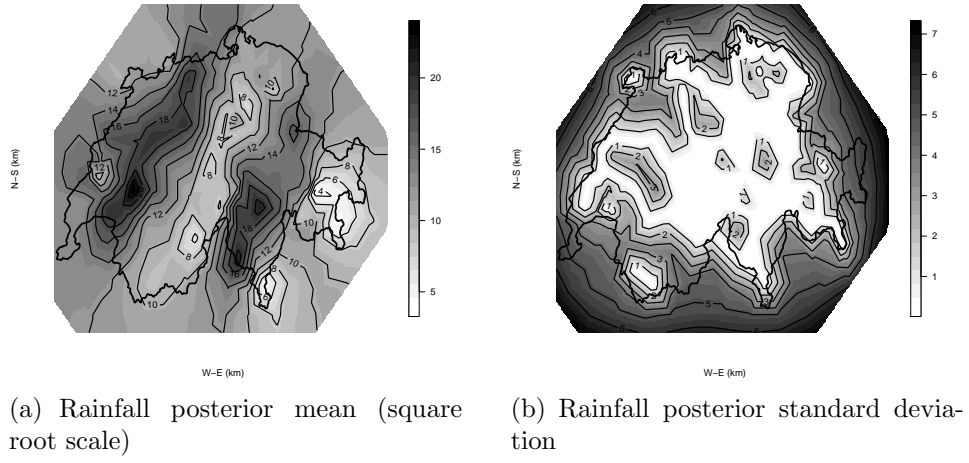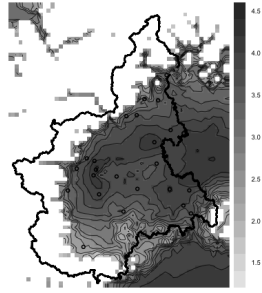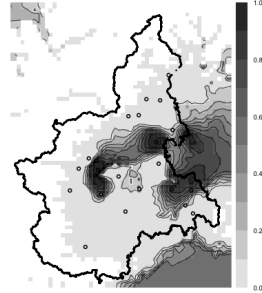(b) Rainfall posterior standard deviation

Figure 7: Map of the rainfall posterior distribution.

(a) Map of the PM$_{10}$ posterior mean on the logarithmic scale.

(b) Map of the posterior probability of exceeding the threshold of 50 $\mu g/m^3$.

Figure 8: Map of the PM$_{10}$ posterior mean and exceedance probability. Both maps refer to the selected day 30/01/2006.