*Journal of the Korean Data & Information Science Society*
*2015,* **26**(2), 495–504

http://dx.doi.org/10.7465/jkdi.2015.26.2.495
한국데이터정보과학회지

# Review on statistical methods for large spatial Gaussian data

Jincheol Park[1]

[1]Department of Statistics, Keimyung University

Received 30 December 2014, revised 13 January 2015, accepted 30 January 2015

**Abstract**

The Gaussian geostatistical model has been widely used for modeling spatial data. However, this model suffers from a severe difficulty in computation because inference requires to invert a large covariance matrix in evaluating log-likelihood. In addressing this computational challenge, three strategies have been employed: likelihood approximation, lower dimensional space approximation, and Markov random field approximation. In this paper, we reviewed statistical approaches attacking the computational challenge. As an illustration, we also applied integrated nested Laplace approximation (INLA) technology, one of Markov approximation approach, to real data to provide an example of its use in practice dealing with large spatial data.

*Keywords*: Gaussian field, Gaussian Markov field, integrated nested Laplace approximation.

## 1. Introduction

Suppose we have the data obtained by sampling a spatially continuous process $\{Z(s_i)\}$, $s_i \in \mathbb{R}^2$, at a discrete set of locations $\mathcal{S} = \{s_i, i = 1, \cdots, n\}$ in a spatial region of interest $\mathcal{A} \subset \mathbb{R}^2$. It is common to assume

$$\begin{aligned} Y(s_i) &= W_i^T \boldsymbol{\beta} + Z(s_i) + \varepsilon_i, \\ \varepsilon_i &\overset{iid}{\sim} N\left(0, \tau^2\right), \end{aligned} \tag{1.1}$$

where $\{Y(s_i)\}$ denotes our observations at locations, $W_i$ denotes a covariate vector associated with $Y(s_i)$, and $\boldsymbol{\beta}$ denotes a corresponding coefficient vector. It is common to model $\{Z(s_i)\}$ as a spatial Gaussian field (GF) with $E\{Z(s_i)\} = 0$, $Var\{Z(s_i)\} = \sigma^2$, and $Corr\{Z(s_i), Z(s_j)\} = \rho(\|s_i - s_j\|; \boldsymbol{\theta})$ for an appropriate correlation function with corresponding parameter $\boldsymbol{\theta}$ and Euclidean distance $\|\cdot\|$. $\tau^2$ is called the nugget variance in this context. The correlation function is chosen from some parametric families, such as the Matérn, powered exponential or spherical (Cressie, 1993). Under model (1.1), $\boldsymbol{Y} = \{Y(s_1), \ldots, Y(s_n)\}^T$ follows a multivariate Gaussian distribution,

$$\boldsymbol{Y} \sim MVN\{\boldsymbol{W}\boldsymbol{\beta}, V(\boldsymbol{\theta}, \tau^2)\}, \tag{1.2}$$

[1] Assistant professor, Department of Statistics, Keimyung University, Daegu 100-741, Korea. E-mail: park.jincheol@gw.kmu.ac.kr

where $\boldsymbol{W} = (W_1^T, \ldots, W_n^T)$ and $V(\boldsymbol{\theta}, \tau^2) = \Sigma(\boldsymbol{\theta}) + \tau^2 I$, with $I$ being the $n \times n$ identity matrix. It is evident from the equation (1.2) that the evaluation of the likelihood involves inverting an the $n \times n$ covariance matrix $V(\boldsymbol{\theta}, \tau^2)$, where the computational complexity of matrix inversion increases as $O(n^3)$ so that there have been practical limitation to employ Gaussian field model. A variety of methods for tackling this obstacle have been proposed in the literatures. These methods can be roughly grouped into three categories: likelihood approximation, lower dimensional space approximation, and sparse matrix-based approximation. In the Section 2, we will take a review on statistical methods developed to meet the large-data challenge. In the Section 3, to a real rainfall data, using integrated nested Laplace approximation (INLA) technique, we will fit the spatial model in order to illustrate its use in practice. The output of INLA will be compared with the one of Kriging (Cressie, 1993) technique.

## 2. Review on methods

### 2.1. Likelihood approximation

The methods in the first category seek to approximate the likelihood function by a product of conditional densities (see, e.g., Vecchia, 1988; Jones and Zhang, 1997; and Stein *et al.*, 2004) or in spectral domain (see e.g., Stein, 1999 and Fuentes, 2007). This approach can be demonstrated best by Vecchia (1988) which sub-grouped the observation vector $\boldsymbol{y}$ into sub-vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_b$ to obtain a sequence of vectors $\boldsymbol{y}_i^* = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i)$ and simplified the likelihood using $\boldsymbol{y}_i^*$s. Specifically, the exact likelihood of $\boldsymbol{y}$ is given by

$$p(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\theta}) = p(\boldsymbol{y}_1^*; \boldsymbol{\beta}, \boldsymbol{\theta}) \prod_{j=2}^{b} p(\boldsymbol{y}_j^* | \boldsymbol{y}_{j-1}^*; \boldsymbol{\beta}, \boldsymbol{\theta}). \tag{2.1}$$

Vecchia (1988) proposed to approximate (2.1) by replacing $\boldsymbol{y}_{j-1}^*$ in the $p(\boldsymbol{y}_j^* | \boldsymbol{y}_{j-1}^*; \boldsymbol{\beta}, \boldsymbol{\theta})$ with $\boldsymbol{u}_i$, a sub-vector of $\boldsymbol{y}^*$, where $\boldsymbol{u}_i$ is selected in a way to simplify matrix operation. Then we can approximate the exact likelihood $p(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\theta})$ by

$$p(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\theta}) \approx p(\boldsymbol{y}_1^*; \boldsymbol{\beta}, \boldsymbol{\theta}) \prod_{j=2}^{b} p(\boldsymbol{y}_j^* | \boldsymbol{u}_{j-1}; \boldsymbol{\beta}, \boldsymbol{\theta}).$$

Concerns with these methods include adequacy of the likelihood approximation and some implementation issues. In addition, expertise is required for selecting an appropriate spectral density estimate or a sequence of conditional densities.

### 2.2. Lower dimensional space approximation

The methods in the second category seek to approximate the spatial process $\{Z(s)\}$ by a lower dimensional space process $\{\tilde{Z}(s)\}$ with the use of smoothing techniques, such as kernel convolutions, moving averages, low rank splines, or basis functions, see e.g., Wikle and Cressie (1999), Lin *et al.* (2000), Kammann and Wand (2003), Paciorek (2007), and Banerjee *et al.* (2008). For example, Banerjee *et al.* (2008) considered a set of knots $\boldsymbol{s}^* = \{s_1^*, \ldots, s_m^*\}$, which may or may not form a subset of the entire collection of observed locations $\mathcal{S}$. Then Gaussian field $Z$ yield $\tilde{Z}^* = \{Z(s_1^*), \ldots, Z(s_m^*)\}^T$,

$$\tilde{\boldsymbol{Z}}^* \sim MVN(\boldsymbol{0}, \tilde{\Sigma}(\boldsymbol{\theta})), \tag{2.2}$$

where $\tilde{\Sigma}(\boldsymbol{\theta})$ is the corresponding covariance matrix of $m \times m$. Then the spatial intepolator at site $s_0$ is defined by $\tilde{Z}(s_0) \equiv E[Z(s_0)|\tilde{\boldsymbol{Z}}^*]$ which can be computed by $c_0^T(\boldsymbol{\theta})\tilde{\Sigma}(\boldsymbol{\theta})^{-1}\tilde{\boldsymbol{Z}}^*$ with $c_0(\boldsymbol{\theta}) = \{Cov(Z(s_0), Z(s_1^*)), \ldots, Cov(Z(s_0), Z(s_m^*))\}^T$. Then this interpolator defines a predictive process $\tilde{Z}(\boldsymbol{s})$ derived from parent process $Z(\boldsymbol{s})$. Benerjee *et al.* (2008) proposed to replace the equation (1.1) with

$$Y(\boldsymbol{s}) = W_i^T\boldsymbol{\beta} + \tilde{Z}(\boldsymbol{s}) + \varepsilon_i. \tag{2.3}$$

In fitting (2.3), instead of working with $n$ random effect $Z$, we now can work with $m$ random effect $\tilde{Z}$ so that we need to invert $m \times m$ covariance matrix instead of a matrix of $n \times n$ dimension.

Another approach to remark is to reduce dimension of covariance matrix by tapering method (see e.g., Furrer *et al.*, 2006; Kaufman *et al.*, 2008), which essentially set elements zero in a covariance matrix when a distance between two sites is located far in way to keep the resulting tapered matrix positive definite. Specifically, using tapering function $K_{taper}$ to the original covariance matrix $\Sigma$, we can obtain the tapered covariance matrix by $\Sigma(\boldsymbol{\theta}) \circ \mathcal{K}(\gamma)$, where the dot notation $\circ$ refers to Schur product and $\mathcal{K}(\gamma)$ is a tapering matrix with parameter $\gamma$. The $(i,j)^{th}$ element of $\mathcal{K}$ is $K_{taper}(\|s_i - s_j\|)$. Then the log-likelihood of $\boldsymbol{Z} = \{Z(s_1), \ldots, Z(s_n)\}^T$ is naturally approximated by

$$l^*(\boldsymbol{z}; \boldsymbol{\theta}, \boldsymbol{\gamma}) = -\frac{1}{2}\log|\Sigma(\boldsymbol{\theta}) \circ \mathcal{K}(\gamma)| - \frac{1}{2}\boldsymbol{z}^T\{\Sigma(\boldsymbol{\theta}) \circ \mathcal{K}(\boldsymbol{\gamma})\}^{-1}\boldsymbol{z}. \tag{2.4}$$

Even though it is intuitively appealing, this approximation is objected by the fact the corresponding *score* function is biased, that is, $E[\frac{\partial}{\partial\boldsymbol{\theta}}l^*(\boldsymbol{\theta})] \neq 0$. An alternative approximation is proposed in Kaufmann *et al.* (2008) by

$$l(\boldsymbol{z}; \boldsymbol{\theta}, \boldsymbol{\gamma}) = -\frac{1}{2}\log|\Sigma(\boldsymbol{\theta}) \circ \mathcal{K}(\boldsymbol{\gamma})| - \frac{1}{2}\boldsymbol{z}^T\left[\{\Sigma(\boldsymbol{\theta}) \circ \mathcal{K}(\boldsymbol{\gamma})\}^{-1} \circ \mathcal{K}(\boldsymbol{\gamma})\right]\boldsymbol{z}, \tag{2.5}$$

which is derived from the property of trace operation, the fact $\hat{\Sigma} = \boldsymbol{z}\boldsymbol{z}^T$, and

$$\boldsymbol{z}^T\Sigma(\boldsymbol{\theta})^{-1}\boldsymbol{z} = tr\{\boldsymbol{z}^T\Sigma^{-1}(\boldsymbol{\theta})\boldsymbol{z}\} = tr\{\boldsymbol{z}\boldsymbol{z}^T\Sigma^{-1}(\boldsymbol{\theta})\} = tr\{\hat{\Sigma}(\boldsymbol{\theta})\Sigma^{-1}(\boldsymbol{\theta})\}.$$

Then maximising $l(\boldsymbol{z}; \boldsymbol{\theta}, \boldsymbol{\gamma})$ now corresponds to solving an unbiased estimating equation for $\boldsymbol{\theta}$, that is, $E[\frac{\partial}{\partial\boldsymbol{\theta}}l(\boldsymbol{\theta})] = 0$. Although lower dimension space approximation methods can reduce the computational burden to some extent, it cannot completely avoid a matrix inversion. For a large dataset, the dimension of the approximation process $\{\tilde{Z}(s)\}$ can still be very high.

### 2.3. Markov random field approximation

The Markov random field approximation method (see e.g., Rue and Tjelmeland, 2002; Rue and Held, 2005), as suggested by its name, is to approximate the spatial process by a Markov random field of sparse covariance matrix. This method was first proposed for regularly spaced data and extended to irregularly spaced data. For example, Hartman and Hössjer (2008) approximated the spatial process by Markov random field on a lattice and interpolated the irregularly spaced data based on the estimates at the grid points of the lattice. The possibility was also discussed by Besag and Mondal (2005) of extending the *de Wijs* process,

an intrinsic and generalised Gaussian random field, approximation to irregularly spaced data. This approach typically model the data first using a Gaussian field (GF) on a set of locations $\mathcal{S}$, to construct a discretised GF with covariance matrix $\Sigma(\theta)$. Then it tries to find a Gaussian Markov field (GMRF) with neighbourhood structure and associated precision matrix $Q$, an inverse of covariance matrix, that is supposed to represent the GF. Finally using the GMRF representation, it carries out computation taking advantage of sparse matrices. Rue *et al.* (2009) proposed an efficient integrated nested Laplace approximation (INLA). The first step of INLA is an approximation $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})$ of marginal posterior of $\boldsymbol{\theta}$:

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left. \frac{p(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{p_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})},$$

where $p_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ is the Gaussian approximation to $p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ matching the true distribution at the mode $\boldsymbol{x}^*(\boldsymbol{\theta})$ for a given $\boldsymbol{\theta}$. The Gaussian approximation $p_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ is possible because $p(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ can be written in a quadratic form such as

$$\begin{aligned} p(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) &\propto p(\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta}) \prod_{i=1}^{n} p(y_i|x_i,\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\theta})Q(\boldsymbol{\theta})^{1/2} \exp\left\{ -\frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} + \sum_{i=1}^{n} \log p(y_i|x_i,\boldsymbol{\theta}) \right\}. \end{aligned}$$

Now that $p(\boldsymbol{\theta}|\boldsymbol{y})$ tend to depart from Gaussian, the approximate posterior marginals can be obtained through numerical integration using $\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y})$. The next step is to construct posterior marginals for the latent field $p(x_i|\boldsymbol{y})$ by

$$\tilde{p}(x_i|\boldsymbol{y}) = \sum_k \tilde{p}(x_i|\boldsymbol{\theta}_k,\boldsymbol{y})\tilde{p}(\boldsymbol{\theta}_k|\boldsymbol{y})\Delta_k,$$

where $\tilde{p}(x_i|\boldsymbol{\theta}_k,\boldsymbol{y})$ is an approximation to $p(x_i|\boldsymbol{\theta}_k,\boldsymbol{y})$ based upon the Laplace approximation such that

$$\tilde{p}(x_i|\boldsymbol{\theta},\boldsymbol{y}) \propto \left. \frac{p(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{p_G(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\boldsymbol{y})} \right|_{\boldsymbol{x}_i=\boldsymbol{x}_i^*(x_i,\boldsymbol{\theta})}.$$

Note that $p_G(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\boldsymbol{y})$ is the Gaussian approximation to $p(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\boldsymbol{y})$ and $\boldsymbol{x}_i^*(x_i,\boldsymbol{\theta})$ is the modal configuration.

Lindgren *et al.* (2010) showed that for some Gaussian random fields in the Matérn class, there exists an explicit link between Gaussian random fields and Gaussian Markov random fields, and suggest to model the data using Gaussian random fields but do the computations using Gaussian Markov random fields. The GMRF representation can be constructed explicitly using a certain stochastic partial differential equation (SPDE) which has Gaussian field with Matérn function as the solution when driven by Gaussian white noise. For the spatial locations on irregular gird, Lindgren *et al.* (2010) suggested to subdivide $\mathcal{A}$, the region of interest, into a set of non-intersecting triangles, where any two triangles meet in at most a

common edge or corner (see Figure 3.2). The result is a basis function representation with piecewise linear basis functions with Gaussian weights such as

$$z(\boldsymbol{s}) = \sum_{k=1}^{m} \phi_k(\boldsymbol{s}) w_k$$

where $\phi_k$ is a basis function, $w_k$ is Gaussian weights and $m$ is the number of vertices on the triangulation. Markov dependencies is determined by a general triangulation of the domain. However, as pointed out in their paper, their approach involves costs of solving stochastic partial differential equations and for irregularly spaced data, it needs additional costs for triangulation of locations of the observations.

Park and Liang (2012) proposed another hierarchical model for large irregularly spaced data by introducing an auxiliary regular lattice loaded with Gaussian Markov random field. Introducing an $M \times N$ auxiliary square lattice, $W = \{(k,l) : k = 1, \cdots, M, l = 1, \cdots, N\}$ over the $\mathcal{A}$, region of interest, we can define a zero-mean GMRF $X = \{X_{kl}, (k,l) \in W\}$ on the auxiliary lattice. Then we get log-likelihood function

$$\log p(\boldsymbol{x}|\boldsymbol{\alpha}, \sigma^2) = -\frac{MN}{2}\log(2\pi) - \frac{MN}{2}\log(\sigma^2) + \frac{1}{2}\log|Q(\boldsymbol{\alpha})| - \frac{1}{2\sigma^2}\vec{\boldsymbol{x}}^T Q(\boldsymbol{\alpha})\vec{\boldsymbol{x}},$$

where $Q(\boldsymbol{\alpha})$ is the precision matrix with associated parameter $\boldsymbol{\alpha}$ where $\vec{\boldsymbol{x}}$ is a prolonged vector of $\boldsymbol{x}$ arranged by rows, with $x_{kl}$ being its $((k-1) \times N + l)$-th element; that is,

$$\vec{\boldsymbol{x}} = (x_{11}, x_{12}, \ldots, x_{1N}, x_{21}, \ldots, x_{2N}, \ldots, x_{M1}, \ldots, x_{MN})^T.$$

Conditioned on the GMRF, we model $\{Z(s)\}$ under a regression setting such that $[Z(s_i)|X]$'s are mutually independent, where $X$ denotes the GMRF defined on the auxiliary lattice. Then we have

$$Z(s_i)|X \sim N\left\{\boldsymbol{r}_i^T Q(\boldsymbol{\alpha})\vec{X}, \sigma^2(1 - \boldsymbol{r}_i^T Q(\boldsymbol{\alpha})\boldsymbol{r}_i)\right\},$$

where $\boldsymbol{r}_i = \{Cov(Z(s_i), X_1), \ldots, Cov(Z(s_i), X_{MN})\}^T$. It is worthwhile to note that it is equivalent to assuming a regression relationship between $Z$ and $X$. This method completely avoids the problem of matrix inversion by using analytical results of GMRFs, and thus can have a high scalability. It is remarkable that the the computational complexity of the proposed method is $O(n)$, which implies that it can be applied to very large datasets with reasonable CPU times (Park and Liang, 2010).

## 3. Real data analysis

As an illustration for large data analysis, we selected INLA technology because it can be extended to non-Gaussian data, applicable in various contexts (Blangiardo *et al.*, 2013; Muff *et al.*, 2014), and it is ready to use in R (Rue *et al.*, 2014). We applied INLA method in estimating spatial field of the rainfall data available in the R-package *fields* (Nychka *et al.*, 2014). The data was collected on 1720 stations located in North America of average rainfall in tenths of millimeters for the months of June, July and August for the period of 1950 - 2010. For the analysis, we took square-root transformation to original observations. The observation field is depicted in Figure 3.1.

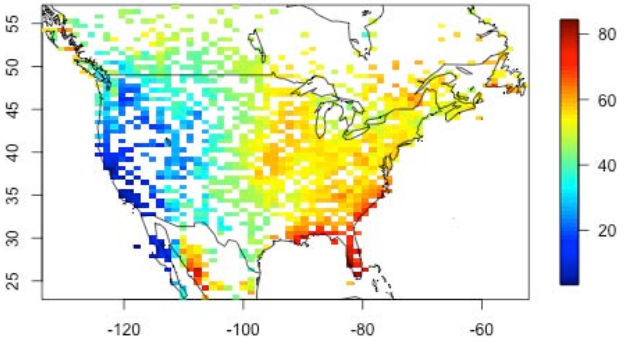


**Figure 3.1** North America rainfall data collected on 1720 stations. The data is the averaged rainfall for the months of June, July and August for the period of 1950-2010 in the unit of tenths of millimetres.

The model we considered was

$$Y(s_i) = \beta_0 + Z(s_i) + \varepsilon_i,$$
$$\varepsilon_i \overset{iid}{\sim} N\left(0, \tau^2\right),$$

with the correlation structure on $Z$ being parametrized by Matérn family with $Var(Z) = \sigma_z^2$ defined by

$$Cov\{Z(s_i), Z(s_j)\} = \frac{\sigma_z^2}{2^{\nu-1}\Gamma(\nu)}(||s_i - s_j||/\varphi)^{\nu}\mathcal{B}_{\nu}(||s_i - s_j||/\varphi), \tag{3.1}$$

where $\mathcal{B}_{\nu}$ is the modified Bessel function of the second kind, $\nu$ is called a smoothness parameter, and $\varphi$ is a parameter controlling a correlation length.

We used *INLA* R-package in fitting the spatial model. First, we subdivided the region of interest $\mathcal{A}$ using triangulation method implemented in a command of *inla.mesh.2d()* that has three primary options: *cutoff*, *offset*, and *max.edge*. The *cutoff* controls the number of triangles on regions: the smaller cutoff, the larger number of the smaller triangles are used to cover the region $\mathcal{A}$. See Figure 3.2 in which it is also worthy of noticing that the coordinates of the data are longitude and latitude so that the mesh is constructed on sphere. The *offset* controls boundary size where the larger *offset* means the wider boundary as can be seen in the Figure 3.3. The *max.edge* controls the maximum allowed triangle edge lengths in the inner domain and in the outer extension so that it is a vector length two.

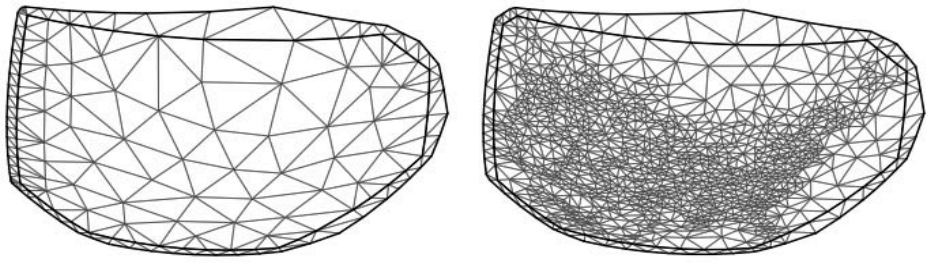

**Figure 3.2** Difference in mesh construction depending on *cutoff* option. For both figures, *offset*=(0.02, 0.02), *max.edge*=(0.1, 0.3). Left: *cutoff*=0.1. Right: *cutoff*=0.01.

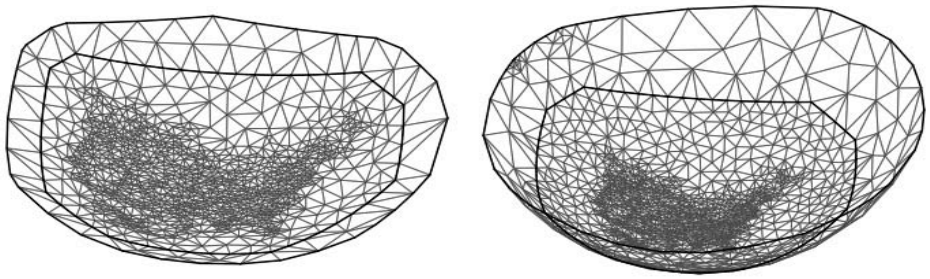

**Figure 3.3** Difference in mesh construction depending on *offset* option. For both figures *cutoff*=0.01, *max.edge*=(0.1, 0.3). Left: *offset*=(0.1, 0.1). Right: *offset*(0.3, 0.3).

To find the relatively best combination of options in meshing the region of interest, we investigated 6 combinations of options (a combination of 2 and 3 possibilities respectively of *cutoff* and *offset*) while keeping an assertional option *max.edge*=(0.1,0.3) constant. To select the best one based on the prediction performance, we selected 100 subsets consisting of 30% of the data, estimated the model without the selected dataset, and predicted the rainfall on the selected sites. Table 3.1 provides the estimated mean square prediction error(MSPE)s for 6 different combinations of options.

**Table 3.1** Prediction performance for various options on mesh construction. The numbers in the parentheses denote standard error.

| *cutoff* | *offset* | MSPE |
|---|---|---|
| 0.01 | 0.02 | 3.499 (0.019) |
| | 0.1 | 3.493 (0.017) |
| | 0.3 | 3.491 (0.017) |
| 0.001 | 0.02 | 3.460 (0.017) |
| | 0.1 | 3.456 (0.018) |
| | 0.3 | 3.455 (0.018) |

Now that (*cutoff*=0.001, *offset*=0.3) yielded the best MSPE performance, we input them in constructing mesh. Then using *inla()* command, we estimated parameters: $\hat{\beta}_0 = 43.835$, $\hat{\tau}^2 = 6.727$, $\hat{\sigma}_z^2 = 281.493$. The prediction field yielded by INLA is depicted in Figure 3.4(a) on $80 \times 200$ grid points. The R-code is provided in the following:

```
#est.x is a data location of longitude and latitude
est.loc.cartesian = inla.mesh.map(est.x, projection = "longlat")
pred.loc.cartesian = inla.mesh.map(pred.grid, projection = "longlat")

#mesh construction
mesh <- inla.mesh.2d(loc=est.loc.cartesian,cutoff=0.001,offset=c(0.3,0.3),
         max.edge=c(0.1,0.3))
spde = inla.spde2.matern(mesh)

#links the process on the mesh vertices triangles with the locations response
A.est =inla.spde.make.A(mesh, loc=est.loc.cartesian)
A.pred =inla.spde.make.A(mesh,loc=pred.loc.cartesian)
field.indices =inla.spde.make.index("field", n.spde=mesh$n)
```

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


#modeling
formula <- rain ~ -1 + Intercept +  f(field, model=spde)
mod =   inla(formula, data=inla.stack.data(stack, spde=spde), family="gaussian",
  control.predictor=list(A=inla.stack.A(stack), compute=TRUE))
```

To assess its performance, we also fitted the spatial model by Kriging (Cressie, 1993) making use of *Krig()* command of *field* of R-package with the Matérn covariance with the smoothness $\nu = 0.5$ (exponential covariance) and the correlation length $\varphi = 30000$ in (3.1). The corresponding prediction field is depicted in Figure 3.4(b), which shows similar pattern to the Figure 3.4(a).

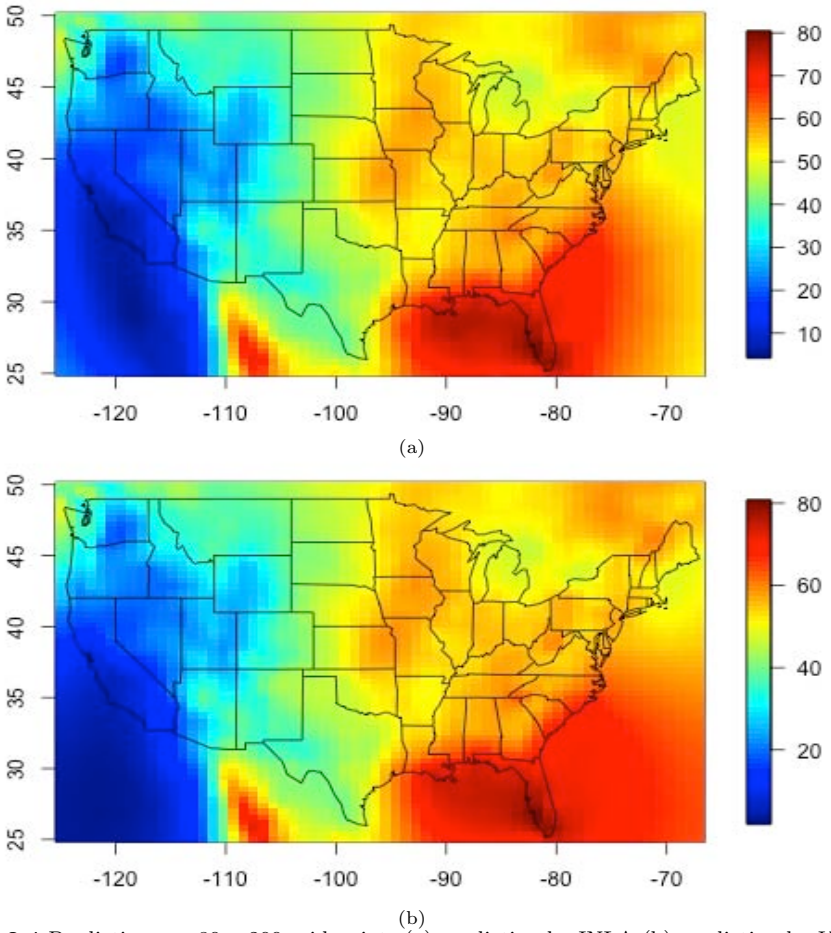


**Figure 3.4** Predictions on $80 \times 200$ grid points (a) prediction by INLA (b) prediction by Kriging

**Table 3.2** Performance of INLA compared with Kriging. The numbers in the parentheses denote standard error.

| Methods | MSPE | Time (seconds) |
|---|---|---|
| INLA | 3.455 (0.018) | 16.87 |
| Kriging | 3.365 (0.017) | 17.87 |

To obtain a better understanding of performance, we also applied Kriging to the identical dataset and summarised the comparison in the Table 3.2. A slight outperformance of INLA over Kriging can be explained by the information loss having occurred in the process of Gaussian Markov field approximation to Gaussian field in INLA. In contrast, Kriging method utilised a full covariance matrix of 1720 in running *Krig()* so that its information loss might be smaller than INLA. It is also remarkable that INLA costs comparable computing time with Kriging, even though INLA is a Bayesian approach contrary to Kriging which is based on maximum likelihood estimation. With the real dataset, we could confirm that Markov Gaussian field can approximate Gaussian field reasonably well.

## 4. Conclusion

In this paper, we have reviewed three statistical approaches geared to address the computation challenge arising in dealing with large spatial Gaussian data: likelihood approximation, lower dimensional space approximation, and Markov random field approximation. One advantage of Markov random field approximation framework is that it can be extended to non-Gaussian data without difficulty so that it is capable to deal with various data models. Selecting INLA, one of computationally efficient Markov field approximation method, we have applied it to real rainfall data as an illustration. From the real data analysis, we could confirm that the Markov Gaussian field approach can provide a reasonable approximation to Gaussian field.

## References

Banerjee, S., Gelfand, A.E., Finley, A. O. and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society B*, **70**, 825-848.

Besag, J. and Modal, D. (2005). First-order intrinsic autoregressions and the Wijs process. *Biometrika*, **92**, 909-920.

Blangiardo, M., Cameletti, M., Bairo, G. and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, **4**, 33-49

Cressie, N. A. C. (1993). *Statistics for spatial data*, 2nd edition, Wiley, New York.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association*, **102**, 321-331.

Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, **15**, 502-523.

Hartman, L. and Hössjer, O. (2008). Fast kriging of large data sets with Gaussian Markov random fields. *Computational Statistics and Data Analysis*, **52**, 2331-2349.

Jones, R. H. and Zhang, Y. (1997). Models for continuous stationary space-time processes. In *Modeling Longitudinal and Spatially Correlated Data (Lecture Notes in Statistics 122)*, edited by: Gregoire, T. G., Brillinger, D. R., Diggle, P. J., Russek-Cohen, E., Warren, W. G. and Wolfinger, R. D., Springer, New York, 289-298.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Applied Statistics*, **52**, 1-18.

Kaufman, C., Schervish, M. and Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial datasets. *Journal of the American Statistical Association*, **103**, 1156-1569.

Lin, X., Wahba, G., Xiang, D., Gao, F., Klein, R. and Klein, B. (2000). Smoothing spline ANOVA models for large datasets with Bernoulli observations and the randomized GACV. *The Annals of Statistics*, **28**, 1570-1600.

Lindgren, F., Lindström, J. and Rue, H. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society B*, **73**, 423-498.

Muff, S., Riebler, A., Held, L., Rue, H. and Saner, P. (2014). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Applied Statistics*, online version.

Nychka, D., Furrer, R. and Sain, S. (2014). fields: Tools for spatial data. R package, version 7.1.

Paciorek, C. J. (2007). Computational techniques for spatial logistic regression with large datasets. *Computational Statistical Data Analysis*, **51**, 3631-3653.

Park, J. and Liang, F. (2012). Bayesian analysis of geostatistical models with an auxiliary lattice. *Journal of Computational and Graphical Statistics.*, **21**, 453-475.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: Theory and applications*, Chapman & Hall/CRC, Boca Raton.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, **71**, 319-392.

Rue, H., Martino, S., Finn, L., Simpson, D., Riebler, A. and Krainski, E. T. (2014). INLA: Functions which allow to perform full Bayesian analysis of latent Gaussian models using integrated nested Laplace approximation. R package, version 0.0-1404466478.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian field. *Scandinavian Journal of Statistics*, **29**, 31-49.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory of Kriging*, Springer, New York.

Stein, M. L. Chi, Z. and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society B*, **66**, 275-296.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society B*, **50**, 297-312.

Wikle, C. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815-829.