# The Efficient Gaussian Approximation for a class of Latent Gaussian model

**Ji Won Yoon**[†] **and Simon P. Wilson**[*]

*Statistics department*
*Trinity College Dublin, Dublin*
*Ireland*

E-mail: [†]yoonj@tcd.ie [*]swilson@tcd.ie

*Abstract* — **For evaluating distributions of unknown form, a Gaussian approximation (GA) is a popular approach. The approximation is often quick and almost always computed iteratively. However, if it is used in a highly non-linear or non-Gaussian system then convergence of the approximation may be slow. In this work we propose an efficient approach to determine a good initial mode (or, equivalently for the Gaussian, a mean) for a GA, which should lead to faster convergence. The approach is illustrated through an example of Bayesian inference for disease mapping using a Gaussian (Markov) random field (GRF/GMRF) [Besag et al(1991)Besag, York, and Mollie].**

*Keywords* — **Gaussian Approximation, Markov Random Field, Initial Mode selection**

## I Gaussian Models

Gaussian Markov random fields (GMRFs) are defined as discrete Gaussian fields with a Markov property of conditional independence of a component with all others given its neighbours [Mardia(1988)]. They have seen widespread application in statistical modelling, for example in spatio-temporal models [Besag et al(1991)Besag, York, and Mollie] and dynamic linear models [West and Harrison(1997)]. Recently, [Rue et al(2009)Rue, Martino, and Chopin] introduced the Integrated Nested Laplace Approximation (INLA) for approximating the posterior distribution of both the GMRF and its parameters $\theta$ in a latent model of the form:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x},\theta) &= \prod_i p(y_i \,|\, x_i, \theta) \\ \mathbf{x} &\sim GMRF(\theta). \end{aligned} \tag{1}$$

INLA approximates the marginal posterior of $\theta$ by

$$\tilde{\pi}(\theta|\mathbf{y}) \propto \left.\frac{\pi(\mathbf{x},\theta,\mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\theta,\mathbf{y})}\right|_{\mathbf{x}=\mathbf{x}^*(\theta)} \tag{2}$$

where $\tilde{\pi}_G(\mathbf{x}|\theta,\mathbf{y})$ is the Gaussian approximation to the full conditional of $\mathbf{x}$ and $\mathbf{x}^*(\theta)$ is the mode of the full conditional of $\mathbf{x}$ for a given $\theta$. Both $\mathbf{x}^*(\theta)$ and the precision matrix of $\tilde{\pi}_G$ are found by numerical optimization of $\pi(\mathbf{x}|\theta,\mathbf{y})$.

Many of the challenges that arise with implementing Gaussian approximations are concerned with determining a more accurate precision matrix while maintaining computational speed and to suggest transformations of that matrix for efficient computation [Rue and Held(2005)]. After exploring the matrix, a Gaussian approximation starts from an initial mode which is generated either randomly or manually. In addition, better approximation is obtained by using a Gaussian approximation and its integrand; see [Rue et al(2004)Rue, Steinsland, and Erland]. However, computation time also depends on which mode is selected at the initial step. In this paper, we propose a new approach to speeding up the calculation of the Gaussian approximation by selecting a good initial mode especially for the highly nonlinear problem such as Bayesian mapping of disease.

### *a) Model for Bayesian Mapping of Disease*

Suppose that we have observations $\mathbf{y} = (\mathbf{y}_1, \cdots, \mathbf{y}_i, \cdots, \mathbf{y}_n)$ where $i$ denotes the index of spatial location. The observations are generated

from the Poisson distribution given the expected number of cases, $\lambda_i$ and the relative risk, $\mathbf{x}_i$ in the area $i$. This model is well studied in epidemiology [Mollie(1995)]. The measurement space in this model is defined by

$$p(\mathbf{y}_i|\mathbf{x}_i,\theta) = \mathcal{P}\{\mathbf{y}_i;\lambda_i\exp(\mathbf{x}_i)\} \quad (3)$$

where $\lambda_i$ is assumed known and $\mathbf{x}_i$ follows a Gaussian Markov process for $\mathbf{x}=(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n)$. Here $\mathcal{P}(\cdot)$ denotes the Poisson distribution.

*b) Gaussian Markov Random Field*

Define $\mathbf{x}$ to be an one dimensional lattice and assume that it is a Gaussian (Markov) random field. Let $\Delta_i$ be differenced values of $\mathbf{x}$ at a site $i$ on the lattice with neighbours for $i\in\{1,2,\cdots,n\}$. There are many possible ways to define a GMRF for $\mathbf{x}$ through the $\Delta_i$. For example, a second order model can be defined through assuming the differences with the four nearest neighbours $\eta_i = \{(i-1),(i+1)\}$, $\Delta_i = \sum_{k\in\eta_i}(f_k - f_i)$, to be Gaussian with mean 0 and precision $\kappa$. The distribution of $\mathbf{x}$ is of multivariate Gaussian form:

$$\begin{aligned} p(\mathbf{x}|\kappa) &\propto \exp\left\{-\frac{\kappa}{2}\sum_{i=1}^{n_1}\Delta_i^2\right\} \\ &= \exp\left\{-0.5\kappa(\mathbf{D}\mathbf{x})^T(\mathbf{D}\mathbf{x})\right\} \\ &= \exp\left\{-0.5\mathbf{x}^T\mathbf{Q}\mathbf{x}\right\}, \end{aligned} \quad (4)$$

where the precision matrix $\mathbf{Q} = \kappa\mathbf{D}^T\mathbf{D}$, and $\mathbf{D}$ is a $n\times n$ matrix with elements $D_{m,m} = -\sum_{d=1,d\neq m}^{n} D_{m,d}$, $D_{m_1,m_2} = -1$ if $m_1$th and $m_2$th components are neighbours, and 0 otherwise. The resulting precision matrix $\mathbf{Q}$ is not of full rank, in which case it is known as an intrinsic GMRF (IGMRF). This form is used often as a prior in Bayesian inference as it avoids having to specify a mean value for $\mathbf{x}$, and yet under very general conditions the posterior distribution of $\mathbf{x}$ will be a proper distribution.

*c) Gaussian Approximation*

When we have the following target distribution

$$\pi(\mathbf{x}) = \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{Q}(\mathbf{x}-\mu) + \sum_{i\in\mathbf{I}} g_i(\mathbf{x}_i)\right\}, \quad (5)$$

the Gaussian approximation of interest $\pi_G(\mathbf{x})$ is obtained by matching the modal configuration and the curvature at the mode. For disease mapping, $\pi(\mathbf{x})$ is the posterior distribution of $\mathbf{x}$ where $g_i(\mathbf{x}_i) = \log p(\mathbf{y}_i|\mathbf{x}_i) = \mathbf{y}_i\log\lambda + \mathbf{y}_i\mathbf{x}_i - \lambda e^{\mathbf{x}_i} - log(\mathbf{y}_i!)$. Eq. (5) can be transformed to the canonical form:

$$\begin{aligned} \pi(\mathbf{x}) &\propto \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{Q}(\mathbf{x}-\mu)\right. \\ &\quad \left.-\sum_i\{\lambda\exp(\mathbf{x}_i) - \mathbf{y}_i\mathbf{x}_i\}\right] = \exp\{f(\mathbf{x})\} \\ &\approx \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{c}\mathbf{x} + \mathbf{b}\mathbf{x} + \text{const.}\right\}. \end{aligned} \quad (6)$$

In order to obtain the parameters $\mathbf{c}$ and $\mathbf{b}$ of the canonical form, we use the first and the second derivatives:

$$\begin{aligned} f'(\mathbf{x}) &= -\mathbf{x}^T\mathbf{Q} - \lambda\exp(\mathbf{x}^T) + \mu^T\mathbf{Q} + \mathbf{y}^T \\ f''(\mathbf{x}) &= -\mathbf{Q} - \text{diag}(\lambda\exp(\mathbf{x})) \end{aligned} \quad (7)$$

where $\mathbf{Q}$ is the precision matrix of the prior distribution, i.e. $\Sigma^{-1}$ and $\exp(\mathbf{x}) = e^{\mathbf{x}} = (e^{\mathbf{x}_1}, e^{\mathbf{x}_2},\cdots,e^{\mathbf{x}_n})$. With an empirically chosen mode $\mathbf{m}$, we can use Taylor expansion for $f(\mathbf{x})$ as follows:

$$\begin{aligned} f(\mathbf{x}) &= (\mathbf{x}-\mathbf{m})^T\frac{f''(\mathbf{m})}{2}(\mathbf{x}-\mathbf{m}) + f'(\mathbf{m})(\mathbf{x}-\mathbf{m}) \\ &\quad +\text{constant} \\ &= -\frac{1}{2}\mathbf{x}^T\mathbf{c}\mathbf{x} + \mathbf{b}\mathbf{x} + \text{constant} \end{aligned} \quad (8)$$

Now, we obtain the $\mathbf{c}$ and $\mathbf{b}$ by

$$\begin{aligned} \mathbf{c} &= \mathbf{Q} + \text{diag}(\lambda\exp(\mathbf{m})) \quad (9) \\ \mathbf{b} &= \mathbf{m}^T\text{diag}(\lambda\exp(\mathbf{m})) - \lambda\exp(\mathbf{m}^T) + \mu^T\mathbf{Q} + \mathbf{y}^T \end{aligned}$$

Using $-\frac{1}{2}\mathbf{x}^T\mathbf{c}\mathbf{x} + \mathbf{b}\mathbf{x} + \text{constant} = -\frac{1}{2}(\mathbf{x}-\mathbf{m}^*)^T\mathbf{Q}^*(\mathbf{x}-\mathbf{m}^*)$, we can obtain

$$\begin{aligned} \mathbf{Q}^* &= \mathbf{c} = \mathbf{Q} + \text{diag}(\lambda\exp(\mathbf{m})) \quad (10) \\ \mathbf{m}^* &= \mathbf{Q}^{*-1}\mathbf{b}^T \quad (11) \end{aligned}$$

In order to obtain the optimal mode of $\mathbf{Q}^*$ and $\mathbf{m}^*$, we run Eq. (10) and (11) recursively until convergence.

## II Finding a good initial mode

Suppose that the optimal mode is $\mathbf{m}^*$ which is obtained by Eq. (10) and (11) with initial mode $\mathbf{m}_0$. The goal is to find an initial mode $\mathbf{m}_0$ which is close to $\mathbf{m}^*$. The derivative of $-\log\pi(\mathbf{x})$ in terms of $\mathbf{x}$ is given by

$$h(\mathbf{x}) = -\frac{\delta\log\pi(\mathbf{x})}{\delta\mathbf{x}} \propto \lambda e^{\mathbf{x}} - \mathbf{y} + \mathbf{Q}(\mathbf{x}-\mu). \quad (12)$$

The optimal mode satisfies that $h(\mathbf{x}) = \lambda e^{\mathbf{x}} - \mathbf{y} + \mathbf{Q}(\mathbf{x}-\mu) = 0$. Unfortunately, it is not straightforward to obtain a solution to satisfy $h(\mathbf{x}) = 0$ since there exists an $e^{\mathbf{x}}$ which is an $N\times 1$ vector. To resolve the problem of a difficult Taylor series expansion, we introduce an approximate projection scheme between vector space and matrix space. After transforming a vector $\mathbf{x}$ to a corresponding matrix $\mathbf{X}$, we can manipulate such a difficult Taylor series in a simpler way. After obtaining an approximate solution $\bar{\mathbf{X}}$ in the matrix space, we return it to the vector space $\bar{\mathbf{x}} = \mathbf{m}_0$.

For our example, $e^{\mathbf{x}}$ in a vector space is transformed to a $N \times N$ matrix as follows: $\Lambda : \mathbf{a} \rightarrow \underline{A}$ for any vector $\mathbf{a}$ where $\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_N \end{bmatrix}^T$ and $\mathbf{A} = \mathrm{diag}(\mathbf{a}) = \Lambda(\mathbf{a})$. Also, the inverse operation is written as follows: $\Lambda^{-1} : \mathbf{A} \rightarrow \tilde{\mathbf{a}}$ for any matrix $\mathbf{A}$ where $\tilde{\mathbf{a}}$ is a vector which constructed by the diagonal elements of $\mathbf{A}$. Note that $\Lambda$ is an one way function which loses the information through the projection so we use the term $\tilde{\mathbf{a}}$ rather than $\mathbf{a}$. When $\mathbf{A}$ is a diagonal matrix, $\tilde{\mathbf{a}}$ is equal to $\mathbf{a}$. For clearance, the bold capital notation stands for the transformed matrix from a vector. Therefore, $\Lambda(e^{\mathbf{x}}) = \mathbf{I} + \mathbf{X} + \frac{\mathbf{X}^T\mathbf{X}}{2} + \cdots$. Now, we have $\Lambda[h(\mathbf{x})] = \lambda \left\{ I + \mathbf{X} + \frac{\mathbf{X}^T\mathbf{X}}{2} + \cdots \right\} - \Lambda(\mathbf{y}) + \Lambda\left(\mathbf{Q}(\mathbf{x} - \mu)\right) = 0$. The gradient function $h(\mathbf{x})$ becomes

$$\begin{aligned} \Lambda[h(\mathbf{x})] &= \lambda \left\{ I + \mathbf{X} + \frac{\mathbf{X}^T\mathbf{X}}{2} + \cdots \right\} - \mathbf{Y} \\ &\quad + \Lambda\left(\mathbf{Q}\mathbf{x}\right) - \Lambda\left(\mathbf{Q}\mu\right) \\ &\approx \lambda \left\{ I + \mathbf{X} + \frac{\mathbf{X}^T\mathbf{X}}{2} \right\} - \mathbf{Y} + \Lambda\left(\mathbf{Q}\mathbf{x}\right) \\ &\quad - \Lambda\left(\mathbf{Q}\mu\right). \end{aligned}$$

Here, we assume that $\Lambda(\mathbf{Q}\mathbf{x}) \approx \mathbf{Q}\Lambda(\mathbf{x}) = \mathbf{Q}\mathbf{X}$. This assumption permits a simple approximate form:

$$\begin{aligned} \Lambda[h(\mathbf{x})] &\approx \lambda \left\{ I + \mathbf{X} + \frac{\mathbf{X}^T\mathbf{X}}{2} \right\} - \mathbf{Y} + \mathbf{Q}\mathbf{X} \\ &\quad - \Lambda\left(\mathbf{Q}\mu\right). \end{aligned}$$

This equation can be written with regard to $\mathbf{X}$ and we have

$$\Lambda[h(\mathbf{x})] \approx \frac{\lambda}{2}(\mathbf{X} - \alpha)^T(\mathbf{X} - \alpha) + \beta \tag{13}$$

where

$$\begin{aligned} \alpha &= -(\mathbf{I}_{N\times N} + \lambda^{-1}\mathbf{Q}) \\ \beta &= -\frac{\lambda}{2}\alpha^T\alpha + \lambda\mathbf{I}_{N\times N} - \mathbf{Y} - \Lambda(\mathbf{Q}\mu). \end{aligned} \tag{14}$$

Our goal is to find $\mathbf{X}$ to satisfy $\Lambda[h(\mathbf{x})] = 0$. Using Eq. (14) we can achieve it by

$$\Lambda[h(\mathbf{x})] \approx \frac{\lambda}{2}(\mathbf{X} - \alpha)^T(\mathbf{X} - \alpha) + \beta = 0. \tag{15}$$

Thus, we have $(\mathbf{X} - \alpha)^T(\mathbf{X} - \alpha) = -\frac{2}{\lambda}\beta = \mathbf{U}^T\mathbf{U}$ where $\mathbf{U}$ is obtained by Cholesky decomposition of $-\frac{2}{\lambda}\beta$. From this equation, we derive $\bar{\mathbf{X}} = \alpha + \mathbf{U}$. Eventually, we can obtain a good initial mode $\mathbf{m}_0$ via the inverse function:

$$\mathbf{m}_0 = \Lambda^{-1}(\bar{\mathbf{X}}) = \Lambda^{-1}(\mathbf{U} + \alpha). \tag{16}$$

Note that $\Lambda^{-1}(\cdot)$ is not an exact inverse function of $\Lambda(\cdot)$. This is an approximate of the inverse of $\Lambda(\cdot)$, which is pseudo-inverse, since $\Lambda(\cdot)$ is an one-way function.

**Algorithm 1** An efficient initial mode selection in Gaussian approximation: Bayesian mapping of disease

**Require:** $\mathbf{Q} > 0$
  Set up $\epsilon$ which is a threshold for convergence.
  $\alpha = -(\mathbf{I}_{n\times n} + \lambda^{-1}\mathbf{Q})$.
  $\beta = -\frac{\lambda}{2}\alpha^T\alpha + \lambda\mathbf{I} - \Lambda(\mathbf{y}) - \Lambda(\mathbf{Q}\mu)$.
  $\mathbf{U} = \mathrm{chol}\left(-\frac{2}{\lambda}\beta\right)$.
  $\mathbf{m}_0 = \Lambda^{-1}\left(U^T + \alpha\right)$.
  $\mathbf{m} = \mathbf{m}_0$.
  **while** *true* **do**
    $\mathbf{Q}^* = \mathbf{Q} + \Lambda(\lambda \exp(\mathbf{m}))$.
    $\mathbf{b} = \mathbf{m}^T\mathrm{diag}(\lambda\exp(\mathbf{m})) - \lambda\exp(\mathbf{m}^T) + \mu^T\mathbf{Q} + \mathbf{y}^T$.
    $\mathbf{R} = \mathrm{chol}(\mathbf{Q}^*)$ where $\mathbf{R}^T\mathbf{R} = \mathbf{Q}$
    $\mathbf{a} = \left(\mathbf{R}^T\right)^{-1}\mathbf{b}^T$.
    $\mathbf{m}^* = \mathbf{R}^{-1}\mathbf{a}$
    **if** $|\mathbf{m}^* - \mathbf{m}| < \epsilon$ **then**
      break
    **else**
      $\mathbf{m} = \mathbf{m}^*$ and $\mathbf{Q} = \mathbf{Q}^*$.
    **end if**
  **end while**

## III Simulation with synthetic datasets

The approach has been tested with simulated data of varying dimension. We set up a set of system parameters used as in Table 1. In order to easily monitor the performance of our approach compared to the conventional approach, a simple Gaussian Random Process is tested with $\mathbf{Q} = \kappa\mathbf{I}$ where $\kappa = 1$. As can be seen in Fig. 1, the modified GA (MGA) also converged to the desired results and it is almost overlapped to the approximation by the conventional GA (CGA).

Table 1: Parameter setting

| | | |
|---|---|---|
| $\mu$: | the mean of Gaussian process prior | $\mathbf{1}_{d_x}$ |
| $\mathbf{Q}$: | the precision of Gaussian process prior | $\mathbf{I}_{d_x \times d_x}$ |
| $\lambda$: | the expected number of cases | 1 |
| $\epsilon$: | the threshold for convergence checking | $10^{-10}$ |

Fig. 2 shows the comparison of the approximations at each iteration. In this figure, the initial mode estimated by MGA is close to the actual desired mode of the target distribution and it approximates the distribution only after a few iterations. The CGA takes more iterations to obtain reasonable approximation.

We also compared the number of iterations for both approaches with 100 randomly generated samples. For CGA, the initial modes are selected randomly. Table 2 represents means and standard deviations of the number of iterations for convergence respectively. As we can see in this table, Modified GA needs a relatively small number of

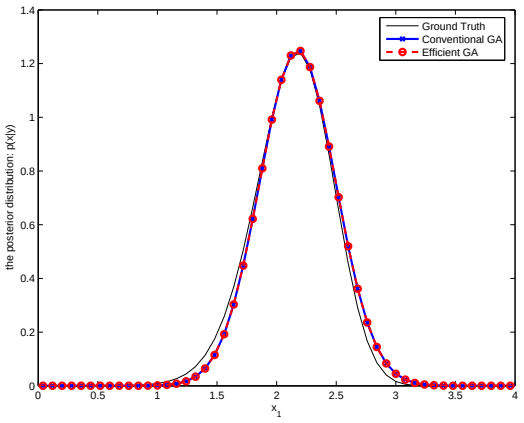


Fig. 1: 1D fitting: target Distribution $\pi(\mathbf{x})$ (solid black line), the distribution by conventional Gaussian approximation (solid red line with square marker) and the distribution by modified Gaussian approximation (dashed line with cross marker)

iterations for convergence compared to the conventional GA.

Table 2: Comparison of the number of iterations for a synthetic dataset ($D$: Dimension)

| $D$ | CGA | MGA |
|---|---|---|
| 1 | 47.59 ± 52.28 | 4.02 ± 1.93 |
| 2 | 70.51 ± 63.22 | 4.80 ± 1.47 |
| 3 | 89.80 ± 60.73 | 5.27 ± 1.19 |
| 5 | 123.73 ± 63.42 | 5.69 ± 1.18 |
| 10 | 161.46 ± 53.62 | 6.39 ± 1.48 |
| 20 | 193.18 ± 53.78 | 7.13 ± 1.32 |
| 50 | 227.61 ± 45.12 | 8.05 ± 1.34 |
| 100 | 262.04 ± 46.43 | 8.64 ± 1.47 |
| 200 | 273.55 ± 39.71 | 9.75 ± 1.68 |
| 500 | 307.72 ± 42.55 | 10.68 ± 1.45 |

Table 3 shows the time comparison. For high dimension (=500), our approach speed up almost 20 times faster than the conventional approach. Note that this time comparison is not optimized since it is estimated with Matlab in 3.16GHz, 3.25GB of RAM.

Table 3: Time comparison for a synthetic dataset

| $D$ | CGA | MGA |
|---|---|---|
| 1 | 0.00171 ± 0.0049 | 0.00016 ± 0.0016 |
| 2 | 0.00173 ± 0.0049 | 0.00000 ± 0.0000 |
| 3 | 0.00249 ± 0.0057 | 0.00016 ± 0.0016 |
| 5 | 0.00391 ± 0.0068 | 0.00015 ± 0.0015 |
| 10 | 0.00563 ± 0.0076 | 0.00015 ± 0.0015 |
| 20 | 0.00999 ± 0.0082 | 0.00032 ± 0.0023 |
| 50 | 0.02859 ± 0.0099 | 0.00110 ± 0.0040 |
| 100 | 0.13578 ± 0.0258 | 0.00672 ± 0.0078 |
| 200 | 0.69669 ± 0.1215 | 0.03627 ± 0.0077 |
| 500 | 9.58830 ± 1.3412 | 0.46592 ± 0.0470 |

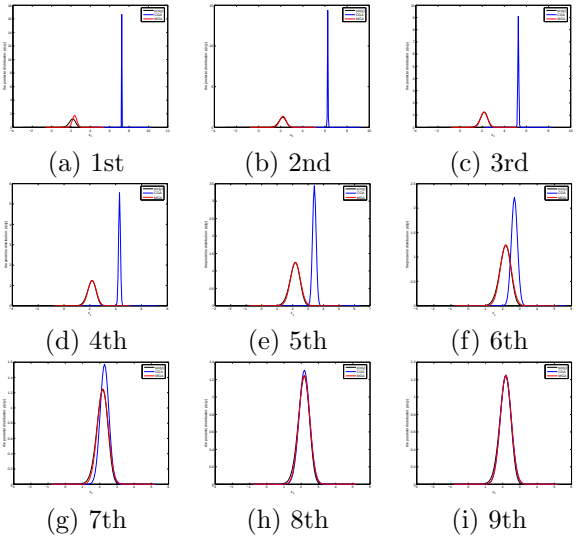


Fig. 2: Comparing the means and covariances of two different approach with 9 iterations: target distribution (solid black line), the approximated distribution by conventional Gaussian approximation (solid blue line) and the approximated distribution by modified Gaussian approximation (solid red line)

## IV Simulation with real experimental datasets

We simulate two different rainfall datasets for practical evaluation: a Swiss rainfall dataset (467 sites) and one Parana rainfall dataset (143 sites) [Diggle and Ribeiro(2007)].

Fig. 3 shows the sequential process of two datasets: (a) for Swiss rainfall data with 467 sites and (b) for Parana rainfall data with 143 sites.

Before applying our approach, we need to build a map of the relationship between sites so we used Delaunay Triangulation which is one of well known geometric approaches. The step 1 of Fig. 3 shows the Delaunay maps of two different Rainfall datasets respectively. With the map, we can build a Gaussian Markov Random Field prior. With the Bayesian Mapping of Disease model, we estimate the curvature of the rainfall datasets as shown in the step 2 of Fig. 3. Here, the raw data which are assumed to be generated from Poisson distribution according to Bayesian mapping of disease model. Then, the last step of Fig. 3 shows the pictorial views of the estimation. In order to visualize the estimated curvatures, we used voronoi diagram to draw these graphs again since voronoi diagram is equivalent to Delaunay triangulation in geometric complexity theory.

We also checked how much different the estimated values at each iterations. As we can see in Fig. 4, our propose approach reaches at the close values only with relatively few iterations. Whereas, the conventional approach with random initial mode consumes more than 100 iterations to the convergence.

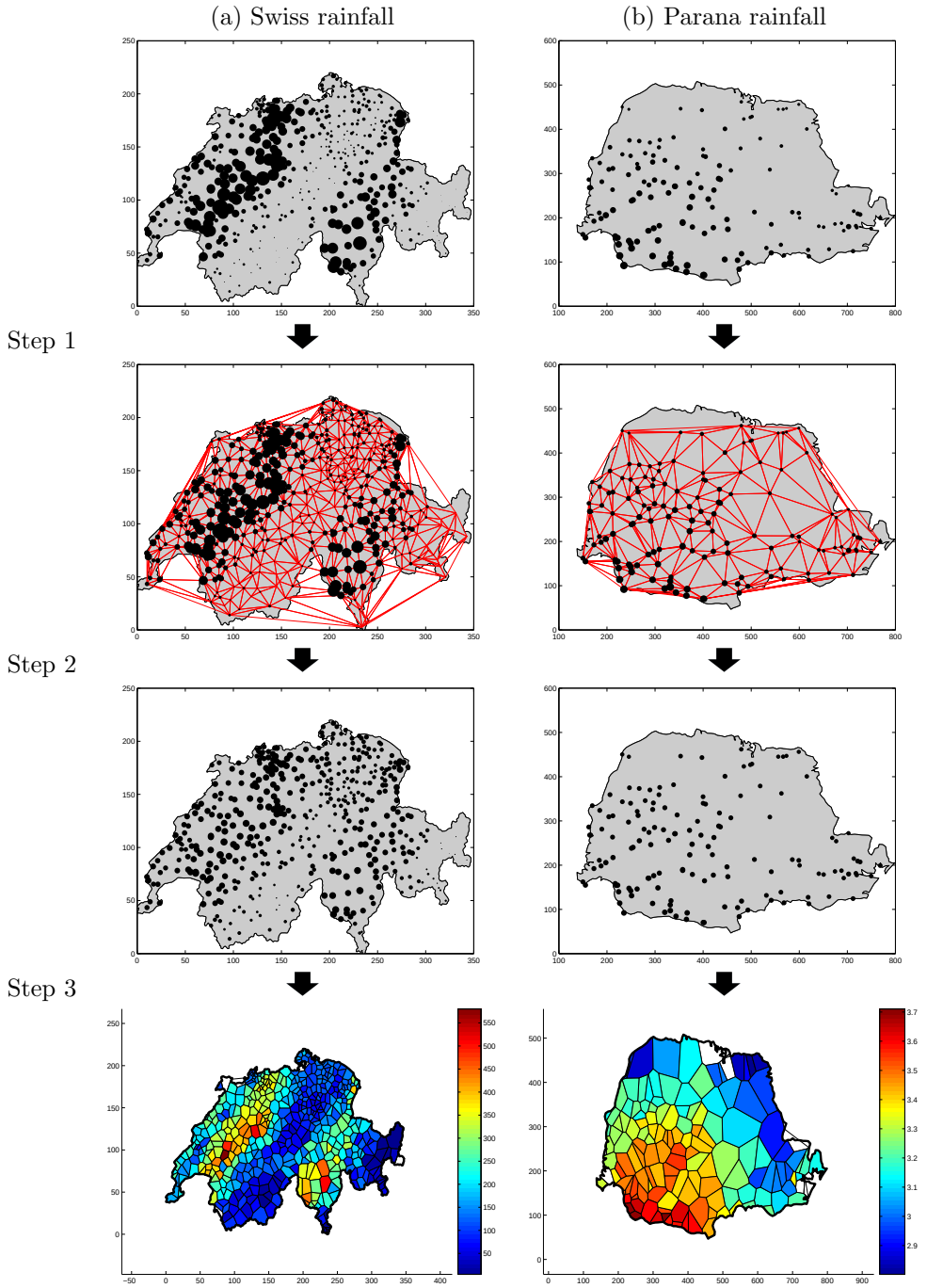


Fig. 3: Rainfall datasets

We can see more comparison of the number of iterations in Table 4. As we can see in this table, our propose approach has almost 10 times fewer iterations than conventional approaches with randomly selected initial mode.

Table 4: Comparison of the number of iterations for rainfall datasets

| Data | Sites | CGA | MGA |
|---|---|---|---|
| Swiss | 467 | $128.38 \pm 41.64$ | $43 \pm 0$ |
| Parana | 143 | $123.87 \pm 45.62$ | $25 \pm 0$ |

We also tested quantitative comparison for systematic time. While CGA takes $2.89 \pm 1.00$ and $0.14 \pm 0.05$ for Swiss and Parana rainfall datasets, MGA takes only $1.07 \pm 0.06$ and $0.03 \pm 0.01$ respectively. that is, the consumed time is almost proportional to the number of iterations and our proposed approach with good initial mode works much faster than the conventional approach.

## V Discussion

In this paper, our approach speeds up the Gaussian Approximation of Poisson model which is known

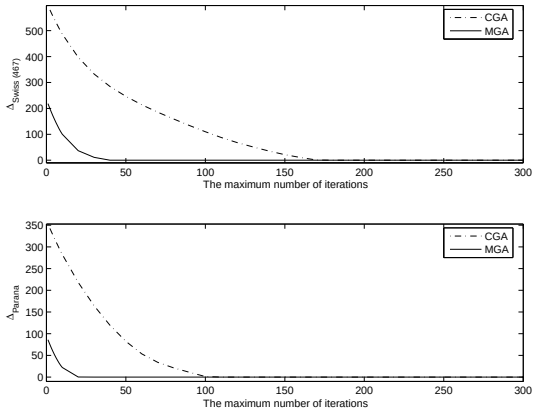


Fig. 4: Difference between approximated values of Rainfall datasets: Swiss rainfall with 100 sites (top), Swiss rainfall with 467 sites (middle) and Parana fainfall (bottom)

as Bayesian Mapping of disease by finding an good initial mode. We are currently working on how to build a generic form to estimate a good initial mode for other models in an exponential family.

The other interesting issue is to applying this to time series. This algorithm can be also applied to spatio-temporal prediction for Log-Gaussian Cox Processes [Brix and Diggle(2001)].

In addition, we can embed this algorithm to Integrated Nested Laplace Approximation (INLA) [Rue et al(2009)Rue, Martino, and Chopin]. Although INLA is well known for its fast speed, we can even far much speed up the conventional INLA by applying our proposed algorithm.

## VI Conclusion

We proposed a fast Gaussian approximation (GA) by embedding an efficient selection of the initial mode. This algorithm does not change any procedure and any results in the conventional GA. However, it saves the number of iterations of GA a lot and it results in speeding up the GA. The power of this algorithm is more effectively shown in higher dimensional problem such as Bayesian mapping of disease.

## References

[Besag et al(1991)Besag, York, and Mollie] Besag J, York J, Mollie A (1991) Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43(1):1–20

[Brix and Diggle(2001)] Brix A, Diggle PJ (2001) Spatio-temporal prediction for log-gaussian cox processes. Journal of the Royal Statistical Society Series B (Statistical Methodology) 63(4):823–841

[Diggle and Ribeiro(2007)] Diggle PJ, Ribeiro PJ (2007) Model-based Geostatistics. Springer

[Mardia(1988)] Mardia KV (1988) Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. J Multivar Anal 24(2):265–284, DOI http://dx.doi.org/10.1016/0047-259X(88)90040-1

[Mollie(1995)] Mollie A (1995) Bayesian Mapping of disease. Chapman & Hall/CRC, in Markov Chain Monte Carlo in Practice

[Rue and Held(2005)] Rue H, Held L (2005) Gaussian Markov Random Fields: Theory and Applications, Monographs on Statistics and Applied Probability, vol 104. Chapman & Hall, London

[Rue et al(2004)Rue, Steinsland, and Erland] Rue H, Steinsland I, Erland S (2004) Approximating hidden gaussian markov random fields. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66(4):877–892

[Rue et al(2009)Rue, Martino, and Chopin] 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):319–392

[West and Harrison(1997)] West M, Harrison J (1997) Bayesian forecasting and dynamic models (2nd ed.). Springer-Verlag New York, Inc., New York, NY, USA