Faculty of science

Spatial Statistics with Image Analysis, MASM25

GMRF:s with Non-Gaussian Data (Project 2)

Authors:
Mohamed Abrash,
Maiya Tebäck

December, 2024

Contents

1 Introduction

The number of coal mining accidents occurring in Britain are recorded and available at the The Coal Mining History Resource Centre [3]. In this project we will use the data available between year 1750 and 1980. The main idea of the project is to combine a Bayesian hierarchical model involving a Gaussian Markov Random Field (GMRF) with Laplace approximation. The model then results in a likelihood function that can be maximized using numeric optimization techniques, giving point estimates of the model parameters, which are then used to reconstruct the data and analyse it.

2 Theory

Model

In this section, we provide a Bayesian hierarchical model for the number of coal mining accidents per year. The number of accidents y_i at a given year t_i, is count data and is therefore modelled by a Poisson distributed random variable conditioned on a latent field z_i:

$$y_i \mid z_i \sim \text{Po}(e^{z_i}). \tag{1}$$

The probability density of a Poisson random variable is given by

$$p(y_i \mid z_i) = \frac{e^{y_i z_i}}{y_i!} \exp(-e^{z_i}). \tag{2}$$

This density has the property that the mean and variance are equal

$$\mathbb{E}(y_i \mid z_i) = e^{z_i}, \quad \text{Var}(y_i \mid z_i) = e^{z_i}. \tag{3}$$

e^{z_i} is the expected number of accidents each year, and the log-expectation is modelled as

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\boldsymbol{\beta} = \tilde{\mathbf{A}}\tilde{\mathbf{x}} \qquad \tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\beta} \end{bmatrix} \in \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \cdot 10^{-3} \end{bmatrix}\right). \tag{4}$$

$\mathbf{A}$ is an observation matrix and x is a Gaussian Markov Random Field (GMRF) with a suitable covariance matrix. $\mathbf{Ax}$ is then a vector of random field values at the observed locations. A Gaussian random field is a flexible model, which allows for a mean value structure, including the effect of covariates, as well as correlations in time. Choosing a Markovian structure simplifies calculations because it results in a sparse precision matrix for the random field. For the purposes of this project we limit ourselves to two structures for the precision matrix $\mathbf{Q}$, a Matérn field with $\alpha = 1$ (CAR) and $\alpha = 2$ (SAR) model:

$$\mathbf{Q}_{\text{CAR}} = \tau\left(\kappa^2\mathbf{C} + \mathbf{G}\right), \tag{5}$$

$$\mathbf{Q}_{\text{SAR}} = \tau\left(\kappa^4\mathbf{C} + 2\kappa^2\mathbf{G} + \mathbf{G}_2\right), \tag{6}$$

where τ and κ^2 are positive real numbers, $\mathbf{C}$ is a diagonal (often identity) matrix, $\mathbf{G}$ is a neighbourhood matrix, and $\mathbf{G}_2 = \mathbf{G}^T\mathbf{C}^{-1}\mathbf{G}$.

For the coal mining accident data, the Markov property is a reasonable assumption, as we expect that knowledge about the number of coal mining accidents in the neighbourhood of some year t_* gives sufficient information about the possible results of y_*.

Model Estimation

A point estimate of the model parameters can be obtained by maximizing the log posterior, obtained by using Bayes' rule, conditioning once on $\tilde{\mathbf{x}}$ and once on $\mathbf{y}$:

$$\log p(\theta|\mathbf{y}) \propto \log p(\mathbf{y}|\tilde{\mathbf{x}}, \theta) + \log p(\tilde{\mathbf{x}}|\theta) + \log p(\theta) - \log p(\tilde{\mathbf{x}}|\mathbf{y}, \theta) \tag{7}$$

which is valid for any $\tilde{\mathbf{x}}$. We now further simplify this expression by taking the second order Taylor expansion of $\log p(\tilde{\mathbf{x}}|\mathbf{y}, \theta)$ in what is known as the Laplace approximation. Below, we provide the necessary calculations needed to perform this approximation. As a slight modification of the Poisson model in (1)–(3), we consider the case where the expected number of accidents per year is instead $E_i \cdot \exp(z_i)$. Then, the observation log-density becomes

$$f_i(z_i) = \log p(y_i|z_i) = y_i \cdot \log(E_i) + y_i z_i - E_i \exp(z_i) - \log(y_i!), \tag{8}$$

which if we differentiate once and twice with respect to the latent process z_i becomes

$$f_i'(z_i) = y_i - E_i \exp(z_i) \tag{9}$$

$$f_i''(z_i) = -E_i \exp(z_i) \tag{10}$$

respectively.
We further assume that the observations are conditionally independent. This gives the joint log conditional density

$$f(z) = \sum f_i = \sum_i \log p(y_i|z_i) = \sum_i y_i \cdot \log(E_i) + y_i z_i - E_i \exp(z_i) - \log(y_i!) \tag{11}$$

Now, if we let all parameters for the Matérn field be collected in a vector as $\boldsymbol{\theta}$, we can compute the log-posterior distribution for $\tilde{\mathbf{x}}$ defined as above, using the fact that

$$p(\tilde{\mathbf{x}}|\mathbf{y}, \boldsymbol{\theta}) = \frac{p(\tilde{\mathbf{x}}, \mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{y}, \boldsymbol{\theta})} \propto p(\mathbf{y}|\tilde{\mathbf{x}}, \boldsymbol{\theta}) p(\tilde{\mathbf{x}}|\boldsymbol{\theta}). \tag{12}$$

We get, using the general template on slide 30 for lecture 7–8 [2] for the first three lines, that

$$\begin{aligned}
\log p(\tilde{\mathbf{x}}|\mathbf{y}, \boldsymbol{\theta}) &= \log p(\mathbf{y}|\tilde{\mathbf{x}}, \boldsymbol{\theta}) + \log p(\tilde{\mathbf{x}}|\boldsymbol{\theta}) + \text{const.} \\
&= \sum_i \log p(y_i|z_i, \boldsymbol{\theta}) + \log p(\tilde{\mathbf{x}}|\boldsymbol{\theta}) + \text{const.} \\
&= \sum_i \log p(y_i|z_i, \boldsymbol{\theta}) - \frac{1}{2}\tilde{\mathbf{x}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{x}} + \text{const.} \\
&= \sum_i \Big\{ y_i \cdot \log(E_i) + y_i z_i - E_i \exp(z_i) - \log(y_i!) \Big\} - \frac{1}{2}\tilde{\mathbf{x}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{x}} + \text{const.} \\
&= \sum_i \Big\{ y_i z_i - E_i \exp(z_i) \Big\} - \frac{1}{2}\tilde{\mathbf{x}}^T \tilde{\mathbf{Q}} \tilde{\mathbf{x}} + \text{const.}
\end{aligned}$$

Note here that the constant term changes between lines, most obviously in the last equality, where we have omitted terms not involving $\tilde{\mathbf{x}}$, as they will disappear when we differentiate, which we will turn to presently. By the chain rule, we note that the first and second derivatives of the observation log-density in (8) with respect to $\tilde{x}_k$ is easily obtained from the results in (9) and (10) by simply multiplying with the factor $\tilde{A}_{ik}$. As stated on slide 40 for lecture 7–8 [2], we further get that the derivative of the sum is

$$\frac{\partial \sum_i f_i(z_i)}{\partial \tilde{x}_k} = \tilde{\mathbf{A}}_{\cdot,k} \nabla_{\mathbf{z}} f = \tilde{\mathbf{A}}_{\cdot,k} [y_i - E_i \exp(z_i)]_i. \tag{13}$$

Then, it is easy to realise that

$$\frac{\partial \sum_i f_i(z_i)}{\partial \tilde{\mathbf{x}}} = \tilde{\mathbf{A}}^T \nabla_{\mathbf{z}} f = \tilde{\mathbf{A}}^T [y_i - E_i \exp(z_i)]_i, \tag{14}$$

and, as also evident from slide 41 for lecture 7–8 [2], the second derivative of the sum with respect to the full vector $\tilde{\mathbf{x}}$ will be

$$\frac{\partial^2 \sum_i f_i(z_i)}{\partial \tilde{\mathbf{x}} \partial \tilde{\mathbf{x}}^T} = \tilde{\mathbf{A}}^T \mathbf{H}_f \tilde{\mathbf{A}} = \tilde{\mathbf{A}}^T \text{diag}(-E_i \exp(z_i)) \tilde{\mathbf{A}}. \tag{15}$$

From equations (96)–(98) on page 12 of the Matrix Cookbook [1], we can observe the correct way to take the gradient of the term involving $\tilde{\mathbf{Q}}$, such that we get the final result that the first and second derivatives of the log-posterior with respect to $\tilde{\mathbf{x}}$ are

$$\frac{\partial \log p(\tilde{\mathbf{x}}|\mathbf{y}, \boldsymbol{\theta})}{\partial \tilde{\mathbf{x}}} = \tilde{\mathbf{A}}^T [y_i - E_i \exp(z_i)]_i - \frac{1}{2}(\tilde{\mathbf{Q}} + \tilde{\mathbf{Q}}^T)\mathbf{x} \tag{16}$$

$$\frac{\partial^2 \log p(\tilde{\mathbf{x}}|\mathbf{y}, \boldsymbol{\theta})}{\partial \tilde{\mathbf{x}} \partial \tilde{\mathbf{x}}^T} = \tilde{\mathbf{A}}^T \text{diag}(-E_i \exp(z_i)) \tilde{\mathbf{A}} - \frac{1}{2}(\tilde{\mathbf{Q}} + \tilde{\mathbf{Q}}^T) \tag{17}$$

respectively.
Now, we can do a Laplace approximation of the log-posterior as in slide 42 for lecture 7–78 [2], from which it is possible to identify terms to be able to make a Gaussian approximation of the same, as is done on slide 44. Finally, if we expand around $\tilde{\mathbf{z}}^{(0)} = \tilde{\mathbf{A}}\widehat{\mathbf{x}}^{(0)}$, where $\widehat{\mathbf{x}}^{(0)}$ is the mode, as on slides 46–47, we get that we can approximate the posterior distribution as Gaussian with

$$\tilde{\mathbf{x}}|\mathbf{y}, \boldsymbol{\theta} \in \mathcal{N}\Big(\widehat{\mathbf{x}}^{(0)}, \big(\tilde{\mathbf{Q}} - \tilde{\mathbf{A}}^T \mathbf{H}_f^{(0)} \tilde{\mathbf{A}}\big)^{-1}\Big) := \mathcal{N}\Big(\widehat{\mathbf{x}}^{(0)}, \mathbf{Q}_{x|y}^{-1}\Big) \tag{18}$$

Now, we can turn to the approximate log-likelihood, expressed as a function of $\boldsymbol{\theta}$, following slides 47–48 for lecture 7–8 [2]. At the mode, the log density simplifies to

$$\log p(\hat{\mathbf{x}}^{(0)}|\mathbf{y},\boldsymbol{\theta}) \propto -\frac{1}{2}\log|\mathbf{Q}_{x|y}| \tag{19}$$

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \frac{p\left(\mathbf{y} \mid \hat{\mathbf{x}}^{(0)},\boldsymbol{\theta}\right) p\left(\hat{\mathbf{x}}^{(0)} \mid \boldsymbol{\theta}\right)}{p_G\left(\hat{\mathbf{x}}^{(0)} \mid \mathbf{y},\boldsymbol{\theta}\right)} p(\theta). \tag{20}$$

Taking the log and substituting in the expressions

$$\log p(\theta|\mathbf{y}) \approx \sum_{i=1}^{n} \log(p(y_i|\mathbf{x}_i^{(0)},\theta) - \frac{1}{2}\hat{\mathbf{x}}^{(0)T}\tilde{\mathbf{Q}}\hat{\mathbf{x}}^{(0)} - \frac{1}{2}\log|\mathbf{Q}_{x|y}| + \frac{1}{2}\log|\tilde{\mathbf{Q}}| + \text{const.} := \ln p(\theta|\mathbf{y};\hat{\mathbf{x}}^{(0)}) \tag{21}$$

where the constants are independent of θ. The approximation sign is due to the fact that we approximated $p(\tilde{\mathbf{x}}|\mathbf{y},\theta)$ using the Laplace approximation.

To illustrate how the $\mathbf{C}$ and $\mathbf{G}$ matrices may be computed, we consider six regularly spaced locations on a one-dimensional grid. Let $\{\psi_i\}$ be a set of triangular basis functions on the grid $\{0,h,2h,3h,4h,5h\}$, where h is the grid-size. Then, we can calculate the elements through basic geometrical consideration. Note that the heights of the triangles are taken to be 1, and the peaks are directly above the grid points, sloping down to zero at the adjacent grid points.

$$\mathbf{C}_{i,i} \approx \langle\psi_i, 1\rangle = \int_{-\infty}^{\infty} \psi_i(t)\cdot 1 dt = \begin{cases} h, & i = 2,3,4,5 \\ h/2, & i = 1,6 \end{cases}. \tag{22}$$

The slope of each side of the triangles is given by $\pm 1/h$, and thus

$$\mathbf{G}_{i,j} = \langle\psi_i, -\psi_j^{''}\rangle = \langle\psi_i^{'}, \psi_j^{'}\rangle = \int \psi_i^{'}(t)\psi_j^{'}(t)dt = \frac{1}{h^2}\begin{cases} 2, & i = j \\ -1, & i = j \pm 1 \end{cases}, \tag{23}$$

and zero otherwise. The second equality follows because of Neumann boundary conditions.
As such, we observe that the elements of the $\mathbf{C}$ matrix scale directly with the size of the grid, and that the elements of the $\mathbf{G}$ matrix are inversely proportional to the square of the grid size. In practice, then, with increasing grid size, the contribution to the precision matrix from the $\mathbf{C}$ part will increase, and the contribution of the $\mathbf{G}$ part to the precision matrix will decrease. This suggests larger variance at the points, and smaller correlation between them — as one would expect when the points move further from one another.

3 Method

The data set is split into modelling and validation sets. The validation set consists of 5 randomly chosen decades.
Point estimates of the model parameters are computed by maximizing the log posterior (21). The maximized log posterior expression is however obtained as an approximation at the maximum of $\log p(\tilde{\mathbf{x}}|\mathbf{y}, \hat{\theta})$. The mode and the point estimates of the parameters are optimized jointly through a nested optimization, following algorithm 1.
To avoid bounded optimization, we re-parametrize the model parameters as

$$\theta_1 = \log(\tau), \quad \tau = \exp(\theta_1), \quad \tau > 0, \tag{24}$$

and

$$\theta_2 = \log(\kappa^2), \quad \kappa^2 = \exp(\theta_2), \quad \kappa^2 > 0. \tag{25}$$

Algorithm 1 Point Estimates; Nested Optimization Algorithm

Require: Outer parameters θ, inner parameters x_{mode}
Ensure: Optimized parameters θ and x_{mode}
1: Initialize θ, x_{mode}
2: **repeat** ▷ Inner Optimization
3: Update $x_{\text{mode}} \leftarrow \arg\max_x \log p(x|\mathbf{y}, \theta)$ ▷ Outer Optimization
4: **Step:** Perform one step of gradient-based optimization:

$$\theta \leftarrow \arg\max_{\theta} \log p(\mathbf{y}|\theta; x_{\text{mode}})$$

5: **until** Convergence in θ
6: **return** θ, x_{mode}

The estimated mode is used to plot the expected number of accidents for each year, denoted as $\exp(\tilde{\mathbf{A}}\mathbf{x}^{(0)})$, along with 80% Poisson prediction intervals (while ignoring uncertainty in the latent field reconstruction).

To assess the variance of the latent field components while neglecting uncertainty in the point estimates, we simulate random fields based on the following procedure:

1. Compute the Cholesky factorization of the precision matrix:

$$\mathbf{R}\mathbf{R}^T = \mathbf{Q}_{xy}.$$

2. Generate 1000 new latent fields as:

$$\tilde{\mathbf{x}}_{\text{sim}} = \mathbf{x}_{\text{mode}} + \mathbf{R}^{-1}\mathbf{e},$$

where $\mathbf{e}$ is a random vector of i.i.d. standard normal variables.

These simulated fields are then used to approximate the variances of the latent field components. The variances are computed marginally for the field components as well as jointly for the complete latent field.

To account for uncertainty in the model parameters, one would typically sample from the posterior distribution using an MCMC method. However, as this uncertainty is neglected in our analysis, we instead conduct a sensitivity analysis. By perturbing the model parameters, simulating the latent field, and analyzing how parameter changes affect the latent field variance, we can evaluate the robustness of the model under varying assumptions.

4 Results

The nested optimization described in algorithm 1 results in the maximum-likelihood point estimates shown in table 1. Note that the magnitude of the parameters is quite different. This makes intuitive sense, given how different the expressions for the $\mathbf{Q}$ matrices in which they are used look (see expressions (5) and (6) for the CAR and SAR matrices respectively). If we for instance consider the terms involving the $\mathbf{C}$ matrix; it is multiplied by a factor $\tau\kappa^2$ in the CAR case, and a factor $\tau\kappa^4$ in the SAR case, the calculation of which yields very similar results: 0.00031479 and 0.00029315 respectively.

Model	τ	κ^2
CAR	53.7737	5.8539e-06
SAR	1.3986e+04	4.5782e-04

Table 1: Point estimates of model parameters for the CAR and SAR (Matérn) models, obtained by log posterior maximization.

Next, we reconstruct the latent fields in both the CAR and the SAR cases, divided up into individual components, the intercept (mean) and random effects (spatial smooth). These correspond to the $\mathbf{B}\boldsymbol{\beta}$ and $\mathbf{Ax}$ in equation (4). One standard deviation confidence bands are also plotted, obtained through simulations as described in the Methods (section 3). The reconstruction of the components of the latent SAR field, shown in figure 2 is noticeably smoother than the reconstruction of components of the latent CAR field shown in 1. Further, we see that the intercept appears to be at around the same level in both cases, the consistency of which is promising. However, the standard deviation is quite a bit greater in the CAR case.

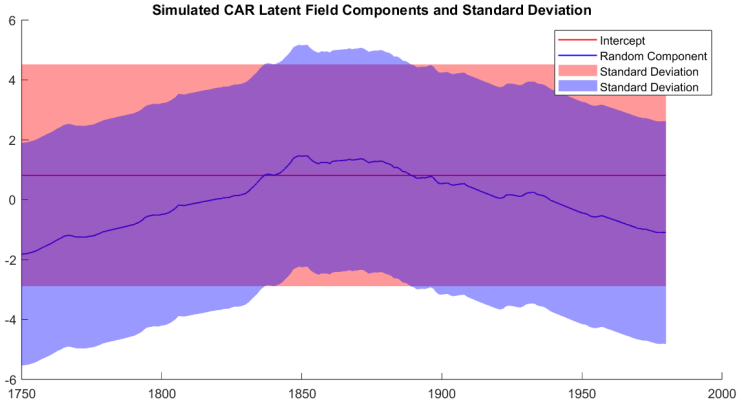

Figure 1: Latent CAR field reconstruction. The figure shows the fixed components, intercept, as well as the random effects and their standard deviations. These estimates are found by using the estimated model parameters to simulate random fields while neglecting uncertainties in the parameter estimates.

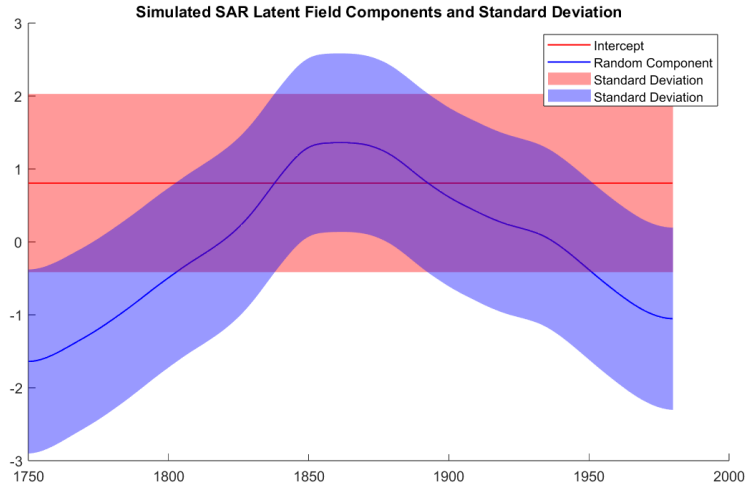

Figure 2: Latent SAR field reconstruction. The figure shows the fixed components, intercept, as well as the random effects and their standard deviations. These estimates are found by using the estimated model parameters to simulate random fields while neglecting uncertainties in the parameter estimates.

It would seem that depending on where in the data set we are, the intercept and random components explain differing amounts of the log-expectation. Toward the centre of the time period, which is also where we have the greatest number of accidents, we see that the intercept and random components are at almost the same positive magnitude, amplifying each other and sharing fairly equally in their contributions. Toward the edges of the data, the random component has switched signs, not only drowning out the intercept but going even more negative (which leads to a total expectation less than 1). In between, then, there are areas where the random effect contributes nothing to the log-expectation, as it varies continuously through the data set. Note that this of course only holds for the posterior means; the standard deviations are quite wide, although for the SAR field, there are areas in the middle and toward the edges where the standard deviation bands do not encompass 0.
We also consider the full latent fields in figures 3 and 4 for the CAR and SAR models respectively. We again clearly see the difference in smoothness, and the width of the standard deviation bands. Notice, however, that the standard deviation when taken together seems to be much less than for each of the components individually, due to them not being independent, as they were essentially treated like in the previous figures 1 and 2.

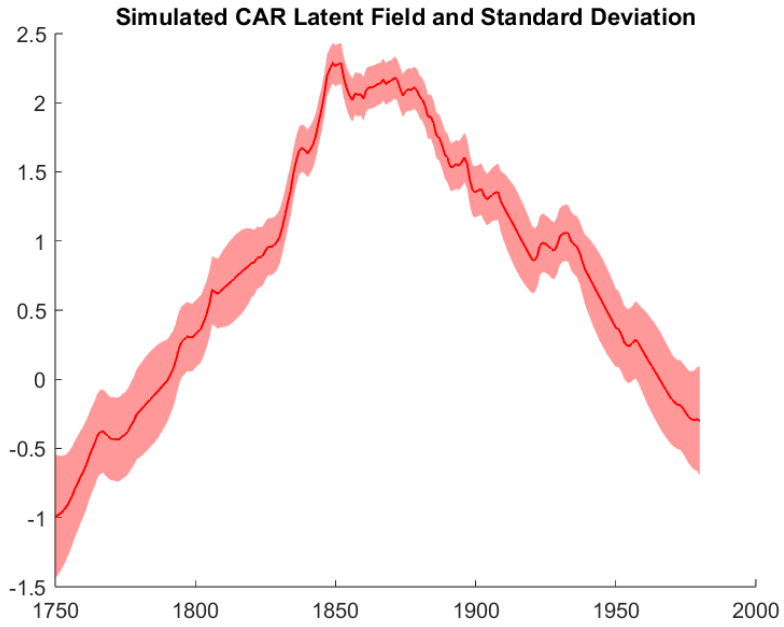

Figure 3: Latent CAR field reconstruction. This time, the complete latent field is plotted, together with its standard deviation, found through simulations taking the model parameters as fixed. Note that uncertainties in the parameters are thus ignored.

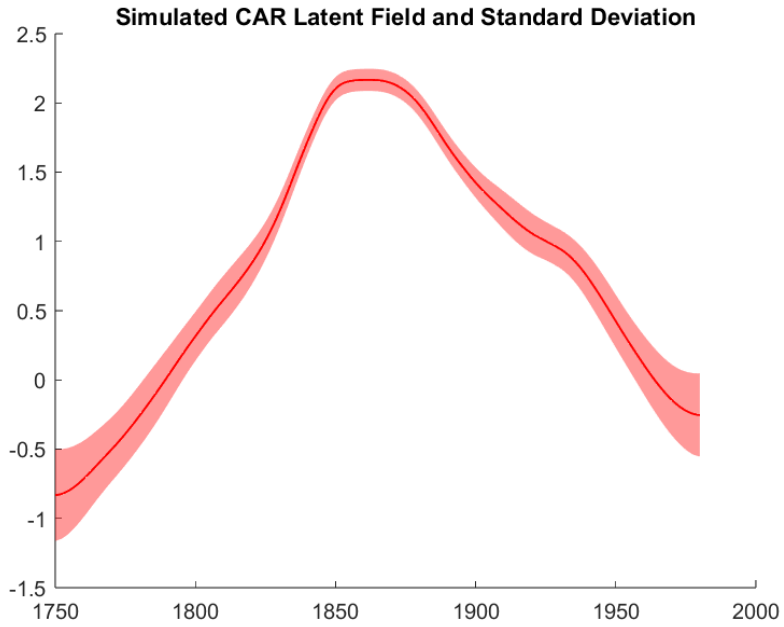

Figure 4: Latent SAR field reconstruction. This time, the complete latent field is plotted, together with its standard deviation, found through simulations taking the model parameters as fixed. Note that uncertainties in the parameters are thus ignored.

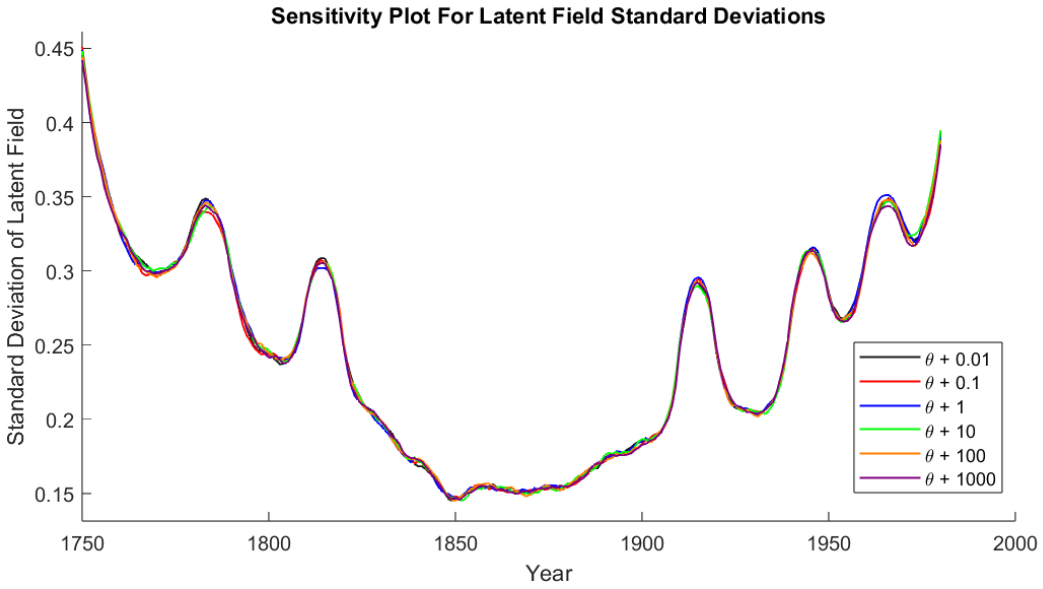

Figure 5: The model parameters are perturbed by different amounts and the latent field is sampled for each perturbed value. Standard deviations are then estimated from the samples and plotted in this figure. The plot shows that the latent field is relatively stable and that neglecting the uncertainty in the model parameters may be justified. The plot also reveals higher standard deviations towards the edges and in five bumps corresponding to the validation data.

Now, we go back to the original scale, such that we plot the expected number of accidents per year, together with the original data, 80% prediction intervals, and standard deviation as calculated through simulations. Note especially that parameter uncertainty is neglected when computing the prediction intervals — these are based on the posterior for the y_i being Poisson, given fixed z_i and parameters $\boldsymbol{\theta}$. We were also able to calculate standard deviations analytically through inversion of the posterior $\mathbf{Q}$ matrix, though the corresponding bands are not shown, as they overlap fairly perfectly with those formed from the simulations. Again, it is clear that the CAR model, shown in figure 6, is more rough than the SAR model, shown in figure 7. The SAR standard deviation bands are also more narrow than the ones for the CAR model. From just a visual inspection, the models seem to be able to capture about as much of the data within the prediction intervals, especially looking at the validation data. The same three points are outside the intervals in both cases.
We could also have simply taken as the bounds for the prediction intervals, the appropriate percentiles from the simulations of the latent fields, and then taking the exponential of these values. However, this still would not be fully appropriate, as we would ideally have had to use the simulated intensities (the latent field values), to then be used to simulate Poisson variables with those intensities, and then taking the appropriate percentiles from those simulations. Also, we note that even in this case, we would have had to let the parameters $\boldsymbol{\theta}$ be fixed. We chose to here instead conduct a sensitivity analysis, as shown in figure 5. This shows that the standard deviations remain quite stable even if the model parameters are perturbed by different values. Interestingly, and as expected, we here see clearly increases in the standard deviations in areas where we have less data — both at the edges of the timeline, as well as for the five regions where our validation data lies (these are what look like bumps in the sensitivity plot).

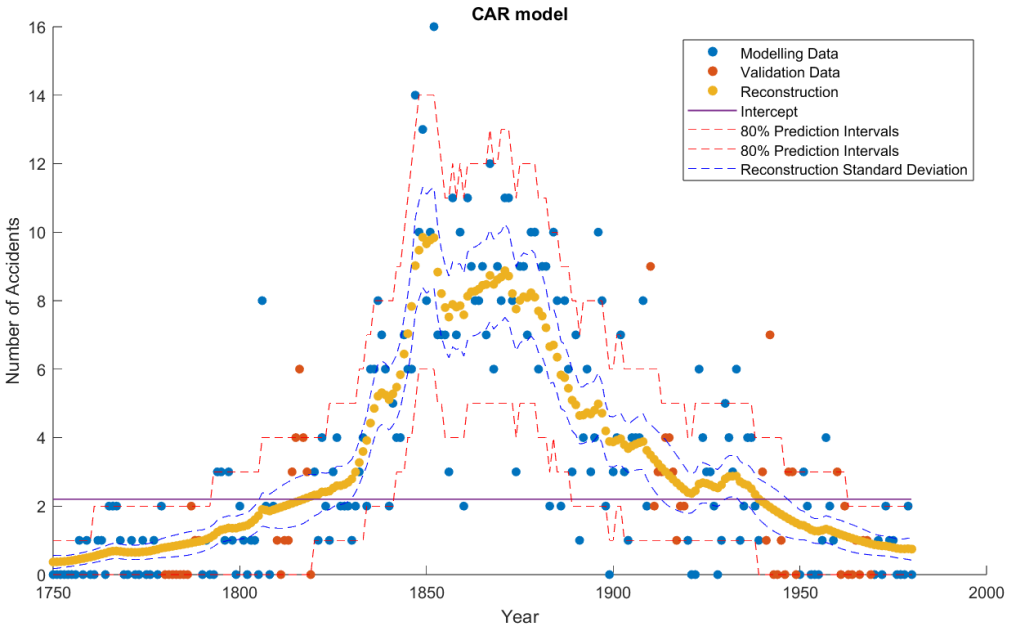

Figure 6: Reconstruction of the expected number of coal mining accidents per year using a CAR model. The expected counts are shown in yellow. The original data is also shown, in blue for the modelling data set, and in red for the validation data set. 80% prediction intervals are plotted (dashed red), calculated as for the Poisson posterior distribution of y_i in equations (1)–(3) taking z_i (and all parameters) as fixed. The reconstruction standard deviation bands (dashed blue) however, are as calculated through simulations.

We also consider the root mean square error (RMSE) of the different models on the validation set. This is shown in table 2, together with the RMSE on the modelling set just for completeness. The CAR model has a slightly smaller error on the validation set than the SAR model, though the difference is not that large. Note also that the RMSE is lower for the validation set than the modelling set. This might be because by chance, our randomly chosen validation set happened to be in areas where the amount of accidents per year was relatively small. And as will be seen toward the end of the section, related to figures 8 and 9, the variance of the simulated fields is larger for larger values of the latent fields.

Model	Modelling Set	Validation Set
CAR	1.85	1.63
SAR	2.06	1.65

Table 2: Root Mean Square Error

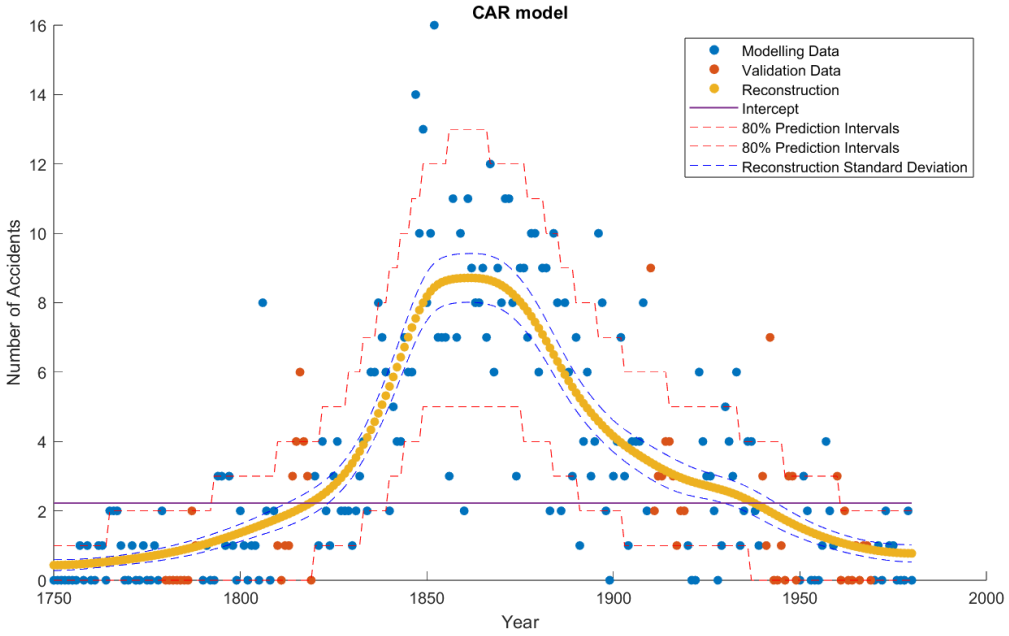

Figure 7: Reconstruction of the expected number of coal mining accidents per year using a SAR model. The expected counts are shown in yellow. The original data is also shown, in blue for the modelling data set, and in red for the validation data set. 80% prediction intervals are plotted (dashed red), calculated as for the Poisson posterior distribution of y_i in equations (1)–(3) taking z_i (and all parameters) as fixed. The reconstruction standard deviation bands (dashed blue) however, are as calculated through simulations.

Performance-wise, then, the models do not seem to differ all too much. Which one that would be preferred, then, might be more up to what one wishes to model. The CAR model seems to be more sensitive to random fluctuations, which would make sense for the data, as the number of coal mining accidents should depend on many external factors and we therefore expect random behaviour. Further, the simulated standard deviation is larger, perhaps then being more realistic in terms of depicting the true uncertainty. On the other hand, there should be dependency between years that are close together, suggesting that a smoother model, and thus SAR, would be preferable. They thus certainly have their respective benefits and disadvantages.
In figure 8, we plot the variance of the predictions when simulating the CAR field, against the observations (number of accidents). As can clearly be seen, with greater number of accidents, comes greater variance. Recalling the second derivative of $\log p(y_i|z_i)$, shown in equation (10), this makes sense — after all, when doing the Laplace approximation of the log-posterior density for the field in expression (18), the second derivative is a factor in the term contributing to the precision matrix. As it specifically contributes to the diagonal elements, corresponding with the variance, it will give a part which is, indeed, greater the greater the value of the latent field (and thus by extension the observations). The corresponding plot is shown for the SAR model in figure 9, showing a similar pattern, just with lower variance, as also expected from previous observations. Interestingly though, the SAR field also has some points with relatively low accident counts that have comparatively large variance. This might be since the variance is just lower overall, and the outliers are more noticeable; or perhaps, because of the structural difference in the precision matrices.

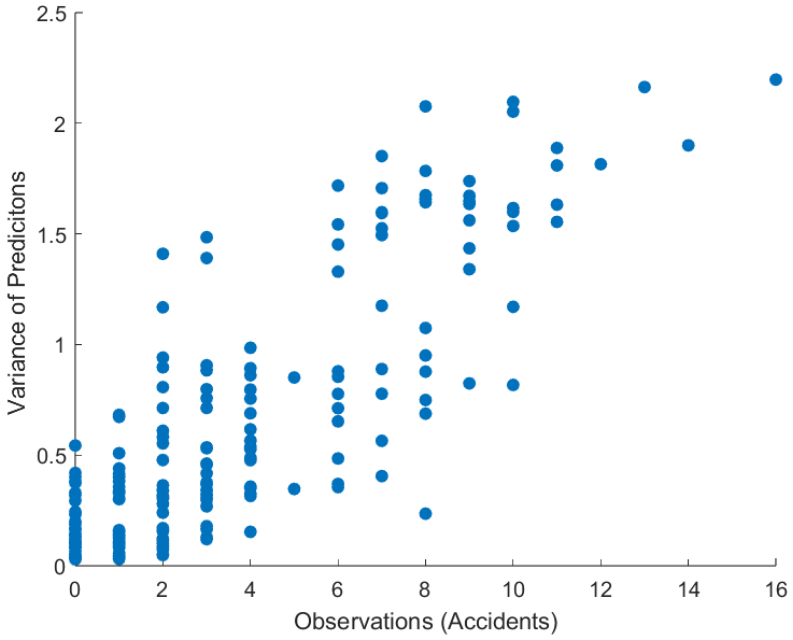

Figure 8: Variance of predictions in the CAR model was calculated by sampling latent fields, exponentiating the samples, and then computing the variance. These values were plotted against the observed data for comparison.

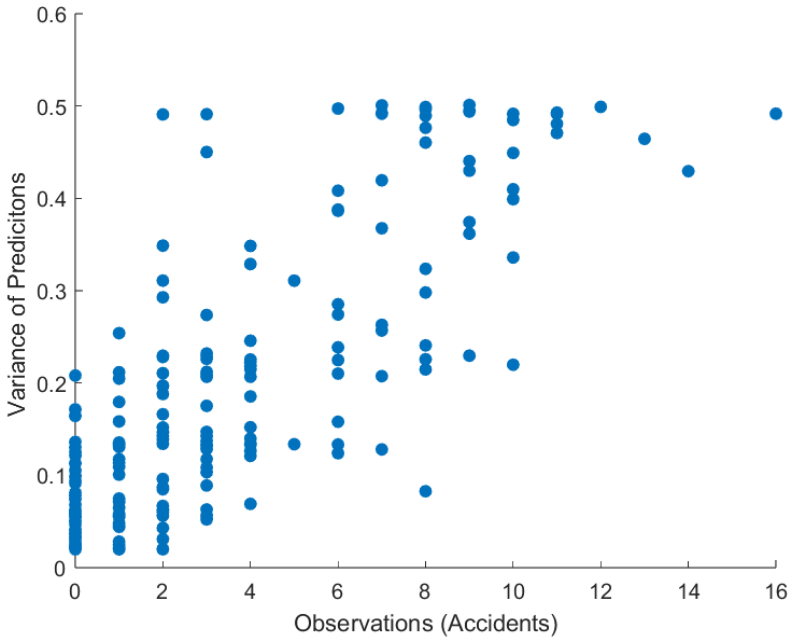

Figure 9: Variance of predictions in the SAR model was calculated by sampling latent fields, exponentiating the samples, and then computing the variance. These values were plotted against the observed data for comparison.

References

[1] Michael Syskind Pedersen Kaare Brandt Petersen. The matrix cookbook. http://matrixcookbook.com, Nov 2012.

[2] Johan Lindström. Image analysis with spatial statistics, Nov-Dec 2024. Lund University.

[3] Ian Winstanley. Coal mining history resource centre. https://www.cmhrc.co.uk.