

Spatial Statistics - Exercise 3

Karine Hagesæther Foss, Martin Outzen Berild, Sindre Henriksen

April 2019

Introduction

Problem 1

a) The '*seismic.dat*' data contains seismic data on a regular grid (75×75), L_D , denoted as $\{d(\mathbf{x}); \mathbf{x} \in L_D\}$; $d(\mathbf{x}) \in \mathbb{R}$, and visually represented by the map in Figure 1. The seismic data has the likelihood model given by the equation

$$[d_i | \mathbf{l}] = \begin{cases} 0.02 + U_i, & l_i = 0 \Rightarrow \text{sand} \\ 0.08 + U_i, & l_i = 1 \Rightarrow \text{shale} \end{cases}, i = 1, 2, \dots, n, \quad (1)$$

where $U_i \stackrel{\text{iid}}{\sim} \mathcal{N}\{0, 0.06^2\}$. This means that we can rewrite Equation 1 to

$$[d_i | \mathbf{l}] = \begin{cases} A_i \stackrel{\text{iid}}{\sim} \mathcal{N}\{0.02, 0.06^2\}, & l_i = 0 \Rightarrow \text{sand} \\ B_i \stackrel{\text{iid}}{\sim} \mathcal{N}\{0.08, 0.06^2\}, & l_i = 1 \Rightarrow \text{shale} \end{cases}, i = 1, 2, \dots, n. \quad (2)$$

The likelihood model $p(\mathbf{d} | \mathbf{l})$ is then given by

$$\begin{aligned} [\mathbf{d} | \mathbf{l}] \sim p(\mathbf{d} | \mathbf{l}) &= \prod_{i=1}^n p(d_i | l_i) \\ &= (0.0072\pi)^{-n/2} \prod_{i=1}^n l_i \exp \left\{ \frac{(d_i - 0.08)^2}{0.0072} \right\} + (1 - l_i) \exp \left\{ \frac{(d_i - 0.02)^2}{0.0072} \right\} \\ &= (0.0072\pi)^{-n/2} \exp \left\{ \frac{1}{0.0072} \left[\sum_{i: l_i=0} (d_i - 0.02)^2 + \sum_{i: l_i=1} (d_i - 0.08)^2 \right] \right\}. \end{aligned} \quad (3)$$

b) We consider a uniform form prior $p(\mathbf{l}) = \text{const.}$ We have $p(\mathbf{l}) = \prod_{i=1}^n p(l_i)$ and $p(l_i) = 1/2$. Assuming the data are collected independently, i.e. $p(\mathbf{d}) =$

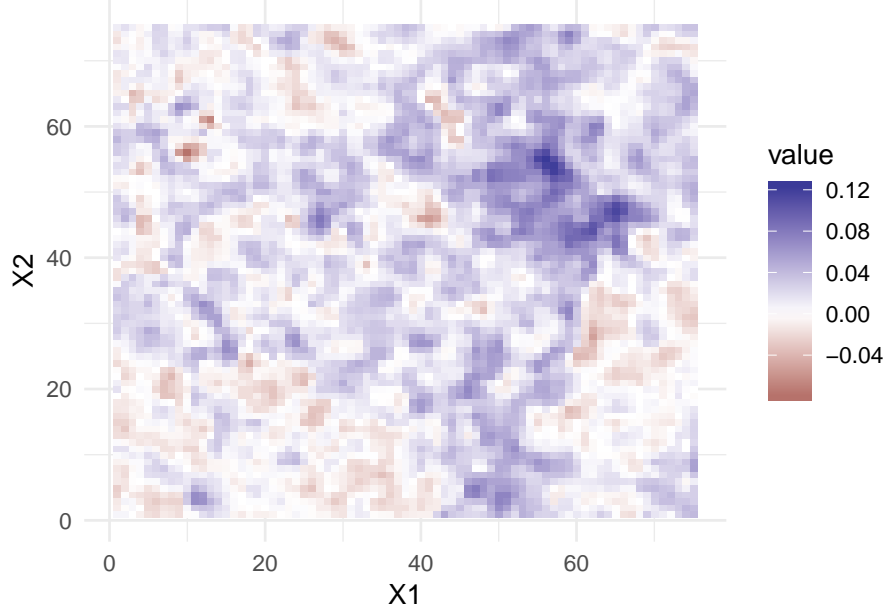


Figure 1: Map of the seismic data $d(\mathbf{x})$ on a grid \mathbf{L}_D

$\prod_{i=1}^n p(d_i)$, we get the posterior

$$\begin{aligned} p(\mathbf{l} | \mathbf{d}) &= \frac{p(\mathbf{d} | \mathbf{l})p(\mathbf{l})}{p(\mathbf{d})} \\ &= \prod_{i=1}^n \frac{p(d_i | l_i)p(l_i)}{p(d_i)} \\ &= \prod_{i=1}^n p(l_i | d_i), \end{aligned}$$

where $p(d_i) = 1/2 \cdot [p(d_i | l_i = 0) + p(d_i | l_i = 1)]$. Thus the values of l_i , $i = 1, \dots, n$ are independent Bernoulli random variables with parameters $p_i = p(l_i | d_i)$. The expectation of the i th element is p_i and the variance is $p_i(1 - p_i)$.

Figure 2, Figure 3 and Figure 4 show 10 simulations from the posterior distribution, the posterior mean and the posterior variance, respectively. We see that the mean is high where the measurements are high and low where the measurements are low. The variance is low both where the measurements are high and low, while it is larger in areas where the measurements are not clearly indicating whether the area contains shale or sand. The posterior mean $\mathbf{E}[\mathbf{l} | \mathbf{d}]$ is the probability that an area contains shale (1).

The maximum marginal posterior predictor MMAP $\{\mathbf{l} | \mathbf{d}\}$ is the vector of values maximizing $p(l_i | d_i)$, $i = 1, \dots, n$, which are 1 where $p(l_i | d_i) > 0.5$ and 0 else. A plot of the MMAP is shown in Figure 5.

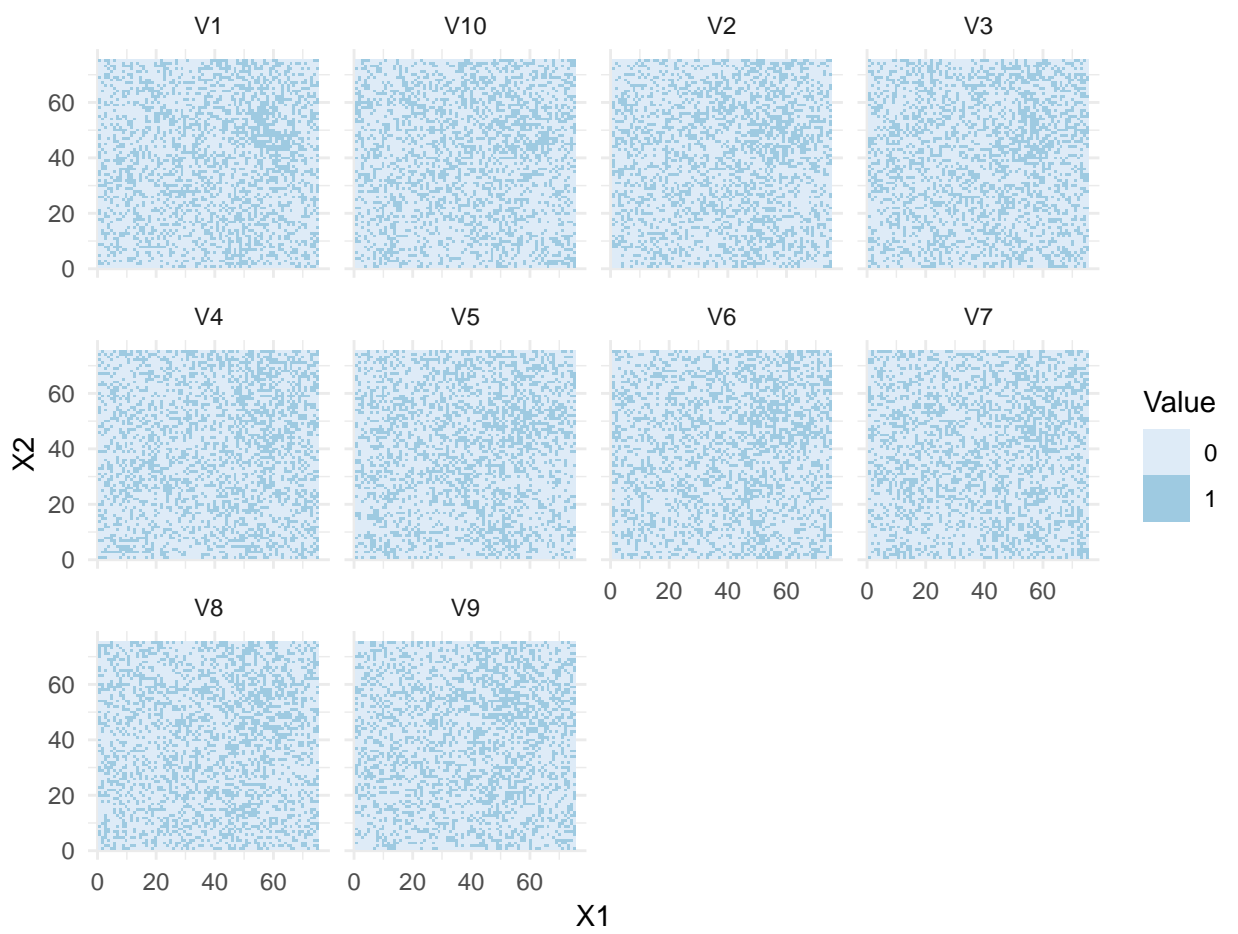


Figure 2: 10 simulations from the posterior given in b).

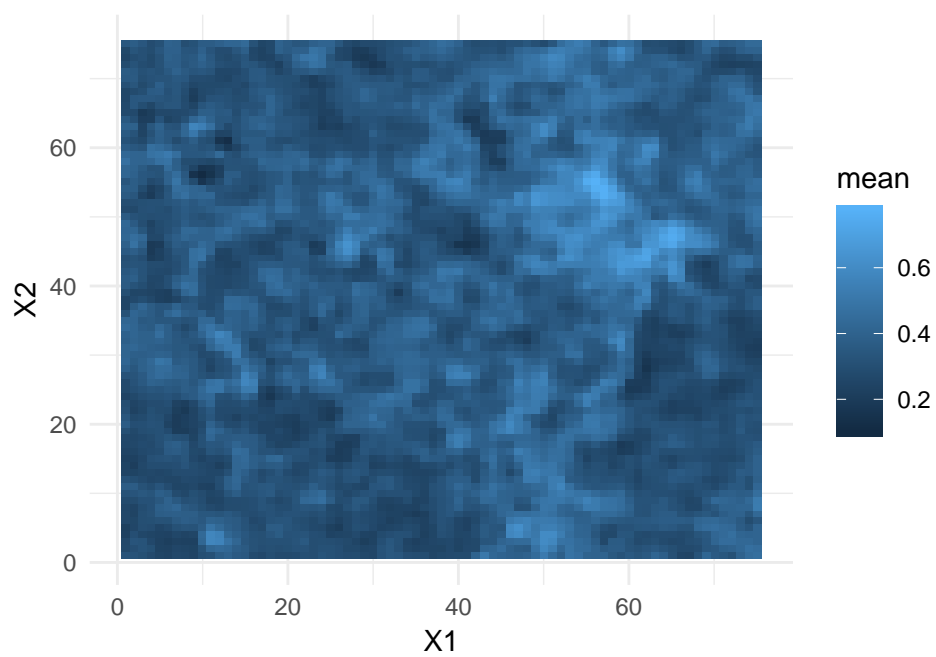


Figure 3: Posterior mean as derived in b).

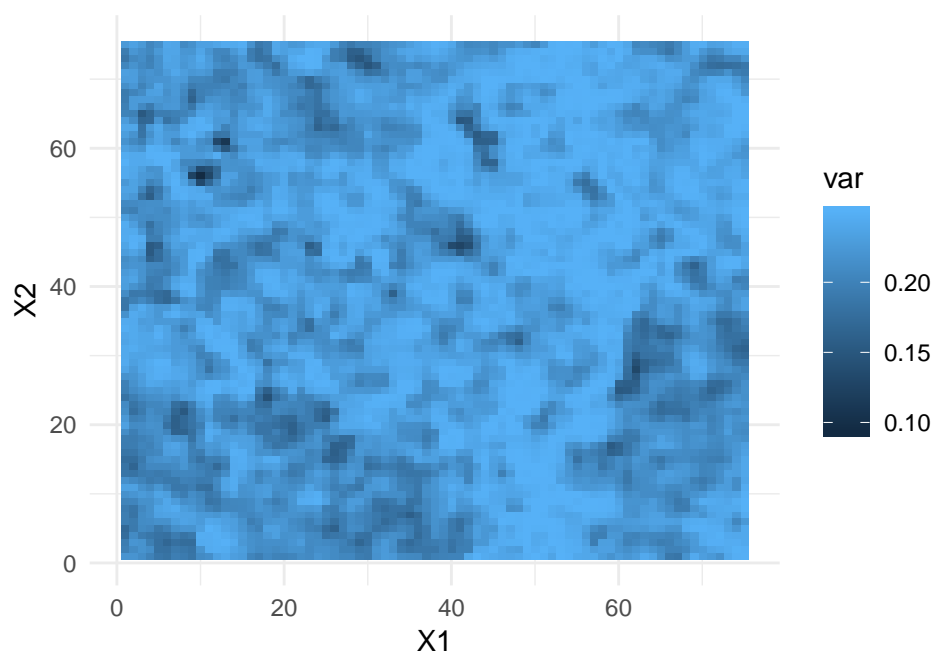


Figure 4: Posterior variance as derived in b).

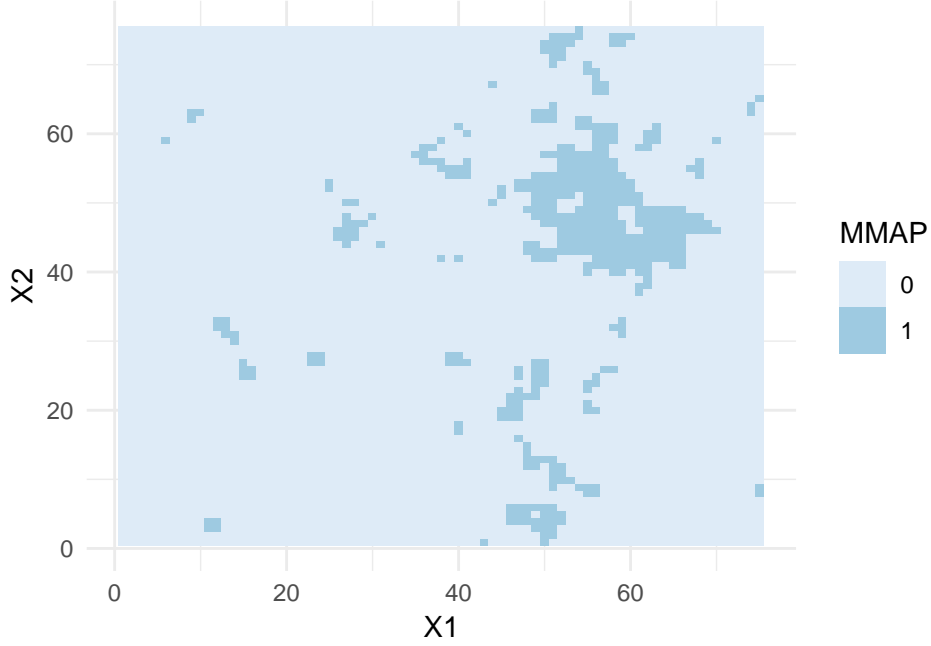


Figure 5: Maximum marginal posterior predictor as derived in b).

c) We will now consider a Markov RF prior model for $\{l(\mathbf{x}; \mathbf{x} \in L_D)\}$, represented by the vector \mathbf{l} . This can be specified by the Markov formulation or Gibbs formulation. The Markov formulation for the prior is

$$p(l_i | l_j; j \in \mathbf{n}_i) = \text{const}_M \times \exp \left\{ \beta \sum_{j \in \mathbf{n}_i} I(l_i = l_j) \right\}; \quad i = 1, 2, \dots, n, \quad (4)$$

where \mathbf{n}_i is the neighborhood of grid node i , consisting of the four closest grid nodes and $I(\cdot)$ is the indicator function. This is the Ising model.

This neighboring system corresponds to a clique system where each clique consists of two neighboring nodes, and we write it

$$\mathbf{C}_L = \{ \langle i, j \rangle; i, j \in L_D \} \quad (5)$$

Then the associated Gibbs formulation of the prior is

$$p(\mathbf{l}) = \text{const}_G \times \prod_{\langle i, j \rangle} \exp \{ \beta I(l_i = l_j) \} \quad (6)$$

$$= \text{const}_G \times \exp \left\{ \beta \sum_{\langle i, j \rangle} I(l_i = l_j) \right\}. \quad (7)$$

Further, we can find expressions for the posterior distributions $p(\mathbf{l} | \mathbf{d})$ and $p(l_i | \mathbf{d}, \mathbf{l}_{-i})$, $i = 1, \dots, n$ and index $-i$ meaning all nodes except i .

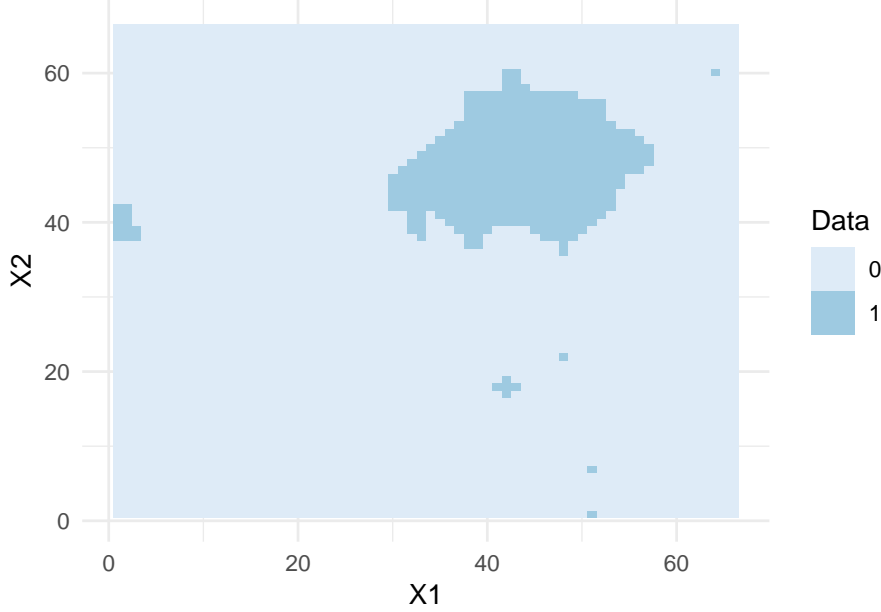


Figure 6: Data from comparable domain D_c

$$\begin{aligned}
 p(l_i|\mathbf{d}, \mathbf{l}_{-i}) &= \frac{p(\mathbf{d}|\mathbf{l})p(l_i|\mathbf{l}_{-i})}{p(\mathbf{d}|\mathbf{l}_{-i})} = \frac{\prod_k p(d_k|l_k) \times p(l_i|l_j; j \in \mathbf{n}_i)}{p(d_i|\mathbf{l}_{-i}) \prod_{k \neq i} p(d_k|l_k)} \\
 &= \frac{p(d_i|l_i)p(l_i|l_j; j \in \mathbf{n}_i)}{p(d_i|\mathbf{l}_{-i})},
 \end{aligned} \tag{8}$$

where $p(d_i|\mathbf{l}_{-i}) = \sum_{l_i^* \in \{0,1\}} p(d_i|l_i^*)p(l_i^*|l_j; j \in \mathbf{n}_i)$ and $p(d_i|l_i)$ and $p(l_i|l_j; j \in \mathbf{n}_i)$ have been specified previously.

$$p(\mathbf{l}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{l})p(\mathbf{l})}{p(\mathbf{d})} = \frac{p(\mathbf{d}|\mathbf{l})p(\mathbf{l})}{\sum_{\mathbf{l}^*} p(\mathbf{d}|\mathbf{l}^*)p(\mathbf{l}^*)},$$

so that $p(\mathbf{d})$ is a sum of all 2^n possible configurations of the n -vector \mathbf{l} .

The aim is to produce realizations from the posterior distribution $p(\mathbf{l}|\mathbf{d})$ and find the posterior mean $\mathbf{E}[\mathbf{l}|\mathbf{d}]$, variance $\text{Var}\{\mathbf{l}|\mathbf{d}\}$ and alternative prediction $\text{MMAP}\{\mathbf{l}|\mathbf{d}\}$. We will have to use an MCMC sampler to generate realizations from the posterior, and then $\mathbf{E}[\mathbf{l}|\mathbf{d}]$, $\text{Var}\{\mathbf{l}|\mathbf{d}\}$ and $\text{MMAP}\{\mathbf{l}|\mathbf{d}\} = \text{argmax}_{l_i} \{p(l_i|d_i)\}$ will be estimated numerically. The latter prediction is calculated simply by counting the frequency of 0 and 1 in each position over the posterior realizations, and assigning the one with the highest frequency.

We want to use observations from a comparable domain $D_c \subset \mathbb{R}^2$ to train the prior model, i.e estimate β . The data from that area is displayed in Figure 6. Here we have exact observations, and call these \mathbf{l}° .

We use a maximum psedo-likelihood procedure to estimate β . We would want to find Maximal Marginal Likelihood (MML) estimate by maximizing $p(\mathbf{d}; \beta)$, but

we have seen that $p(\mathbf{d})$ is very computer demanding. Thus, we instead maximize the Marginal pseudo-Likelihood (MpL),

$$\begin{aligned} & \prod_{i=1}^n p(d_i, d_j; j \in \mathbf{n}_i) \\ & \approx \prod_{i=1}^n \sum_{[l_i^*, l_j^*; j \in \mathbf{n}_i] \in \{0,1\}} \prod_{j=i, j \in \mathbf{n}_i} p(d_i | l_i^*) p(l_i^* | l_j^*, j \in \mathbf{n}_i; \beta) \\ & := \hat{p}(\mathbf{d}; \beta). \end{aligned}$$

Since we have exact observations, this simplifies to

$$\hat{p}(\mathbf{d}; \beta) = \hat{p}(\mathbf{l}^o; \beta) = \prod_{i=1}^n p(l_i^o | l_j^o, j \in \mathbf{n}_i; \beta),$$

with

$$p(l_i^o | l_j^o, j \in \mathbf{n}_i; \beta) = \frac{\exp \{ \beta \sum_{j \in \mathbf{n}_i} I(l_i = l_j) \}}{\sum_{l_i^* \in \{0,1\}} \exp \{ \beta \sum_{j \in \mathbf{n}_i} I(l_i^* = l_j) \}}.$$

Thus,

$$\hat{\beta} = \operatorname{argmax}_{\beta} \{ \log(\hat{p}(\mathbf{l}^o; \beta)) \},$$

which we find using the `optim()` function in R to be $\hat{\beta} = 1.72$.

We now want to generate realizations from the posterior distribution $p(\mathbf{l} | \mathbf{d}; \hat{\beta})$. We do this by using MCMC/Gibbs sampler. To avoid boundary problems, we use wrapping boundary conditions, which means that we let the system vertically and horizontally periodic, so that when we need a value that is located outside the grid, we use the value on the other side of the grid, vertically or horizontally.

In the Gibbs sampler, we propose a value (either 0 or 1) for one location at a time, from $p(l_i | \mathbf{d}, \mathbf{l}_{-i})$ as given in equation (8). Because of the properties of this full conditional distribution, the acceptance probability is always 1, so the step is always accepted. It is obvious that we will have to propose to change quite a few locations before the the chain can converge and between realizations that we can assume are independent. Therefore, realizations of the chain is only saved for every $n = 75 \times 75$ proposals. This will be called one iteration. That is, when we run the MCMC 2000 iterations, $n \cdot 2000$ locations have been proposed to change.

To assess the question of convergence, we display the fraction of 0 and 1 in the Markov Chain. This is shown in Figure 7. By this trace plot, it at least seems that the fractions have stabilized. This is of course not a proof of convergence. In particular, it should be noted that it is possible to reach the situation were all tiles become 0 (sand), and it is very hard to get out of this stage. This has not happened in this iteration, and we assume that the chain gives us samples from the posterior distribution after just a short burn-in period. To be sure, we disregard the first 500 samples.

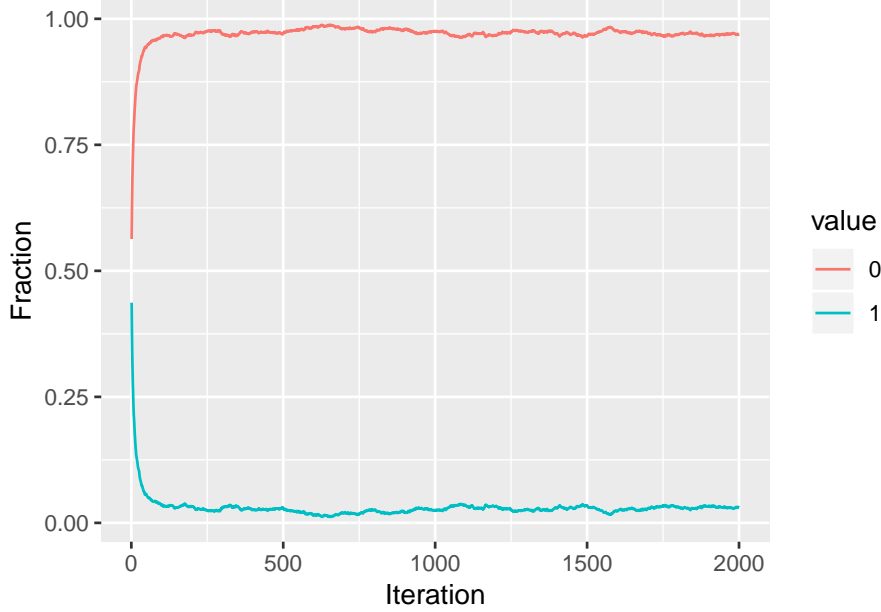


Figure 7: MCMC trace plot of the fraction of grid points with value 0 (sand) and 1 (shale).

In Figure 8, 10 realizations of the posterior distribution are shown. These are drawn from the MCMC samples, evenly spaced out between sample 600 and 2000. This means that a lot of changes are proposed between each realization, so that we can assume they are independent. The realizations are similar, but have individual differences, just as we would expect if they are drawn from the same distribution.

We can now compute and plot the posterior mean, variance and MMAP for each grid node. These are displayed in Figures 9, 10 and 11, respectively. We see that in the MMAP, and also indicated in the posterior mean, we get a clear area of shale in the upper right corner ($l_i = 1$), and sand everywhere else. The realizations have the same trend, though the shape of shale area differ somewhat and they get some small areas or single grid points with shale in between the sand. The variance is only somewhat large for a very small area, as the difference in shape is not very large between the realizations. We also observe that the degree of clustering is very similar to the data from the comparable domain D_c displayed in Figure 6, as this is the data we have used to estimate β . The posterior distribution is affected by this parameter from the prior model, as well as by the nature of the observed data, shown in Figure 1.

d) We see that the results obtained in b) and c) are quite different. In b) we assume independence between the the values $l_i, i = 1, \dots, n$, which means that we get very limited information, and we have a lot of mixing in the posterior realizations. In both cases, we get somewhat more shale ($l_i = 1$) in the upper



Figure 8: 10 independent realizations from the posterior distribution for the Markov RF, obtained by MCMC.

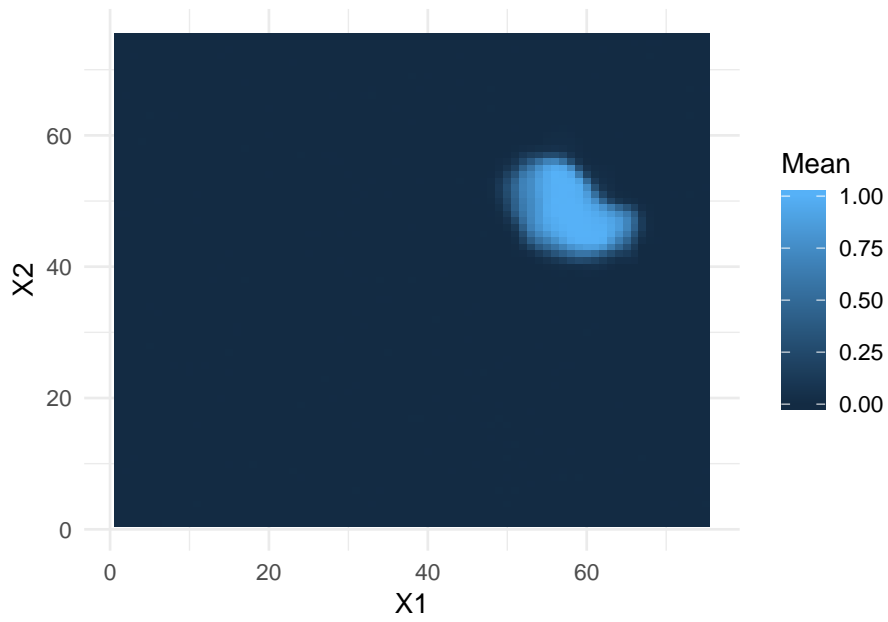


Figure 9: Posterior mean, obtained from a Markov RF prior model using MCMC.

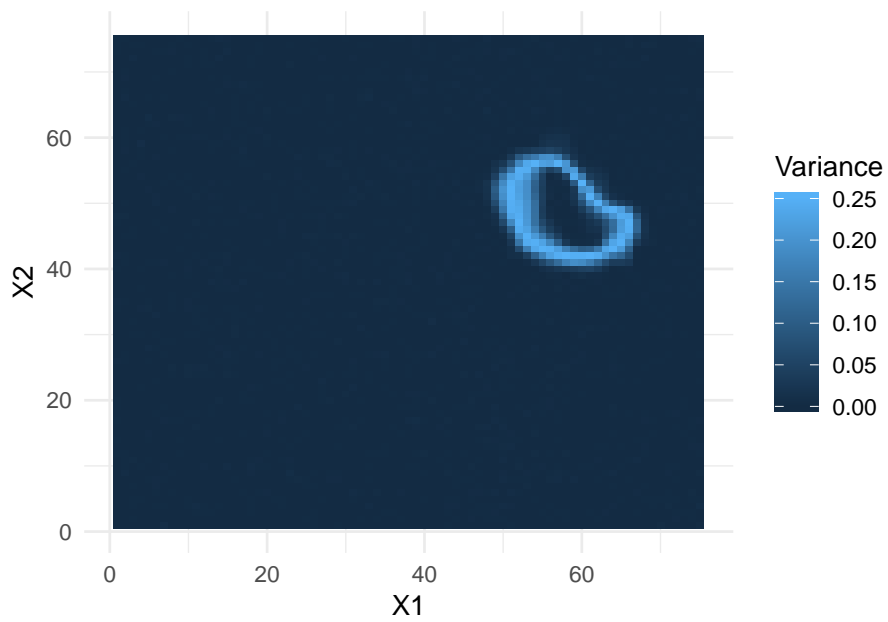


Figure 10: Posterior variance, obtained from a Markov RF prior model using MCMC

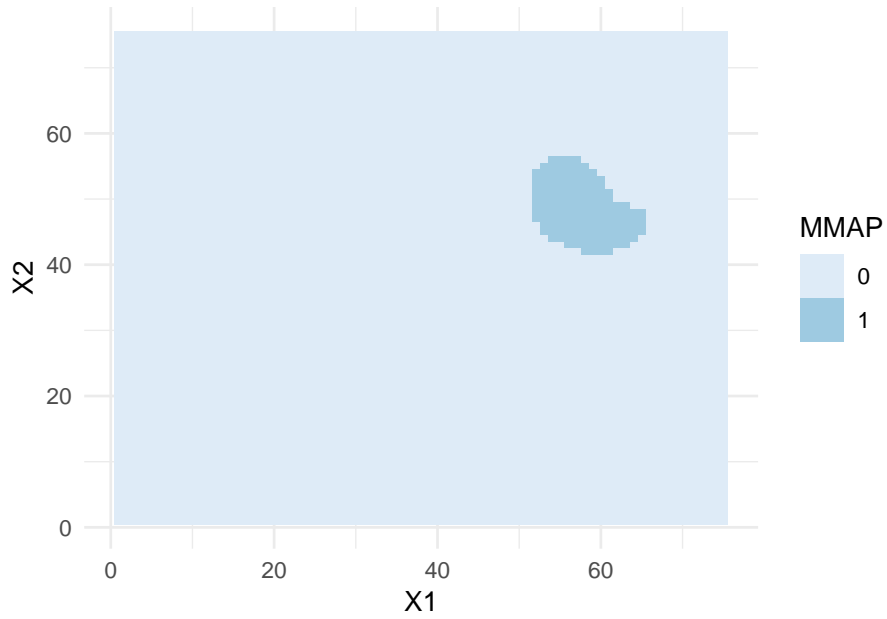


Figure 11: Maximum marginal posterior predictor, obtained from a Markov RF prior model using MCMC

right corner, just as we have observed the highest values in this corner, but the picture is a lot clearer in c). Here, the values drawn from the posterior are a lot more clustered. This shows that adding a dependency of only the four closest neighbors yields a significant difference.