Problem set 3

TMA4250 - Spatial Statistics

Mads Adrian Simonsen and Karina Lilleborge

April 2021

Introduction

This assignment contains problems related to Markov random fields (RF).

Problem 1: Markov RF

We consider seismic data on a domain $\mathtt{D} \subset \mathbb{R}^2$. The aim is to determine the underlying {sand, shale} lithology distribution over $\mathtt{D}$, which is represented by $\{0, 1\}$, respectively.

We are given the data, `seismic.dat`. This data consists of collected observations $\{d(\mathbf{x}) \in \mathbb{R}; \mathbf{x} \in \mathtt{L}_\mathtt{D}\}$, represented by the n-vector $\mathbf{d}$, on a regular (75×75) grid $\mathtt{L}_\mathtt{D}$.

Furthermore, we are given a second data, `complit.dat`. This data consists of observations with code $\{0, 1\}$ for {sand, shale} on a regular (66×66) grid $\mathtt{L}_{\mathtt{D}_c}$ over a geologically comparable domain $\mathtt{D}_c$. That is, $\mathtt{L}_{\mathtt{D}_c}$ and $\mathtt{L}_\mathtt{D}$ have equal spacing.

We assume that the underlying lithology distribution can be represented by a Mosaic RF $\{l(\mathbf{x}); \mathbf{x} \in \mathtt{L}_\mathtt{D}\}; l(\mathbf{x}) \in \mathbb{L} : \{0, 1\}$, represented by the vector $\mathbf{l} \in \mathbb{L}^n$.

a)

The seismic observations data is shown in Figure 1. We assume the observations follows the likelihood model,

$$[d_i \mid \mathbf{l}] = \begin{cases} \mu_0 + U_i, & \text{if } l_i = 0 \\ \mu_1 + U_i, & \text{if } l_i = 1 \end{cases}, \quad i = 1, 2, \ldots, n,$$

where $\mu_0 = 0.02$, $\mu_1 = 0.08$ and $U_i \sim \mathrm{N}(0, \sigma^2)$ are all independent, with $\sigma^2 = 0.06^2$. Then, the likelihood $p(\mathbf{d} \mid \mathbf{l})$ is found to be

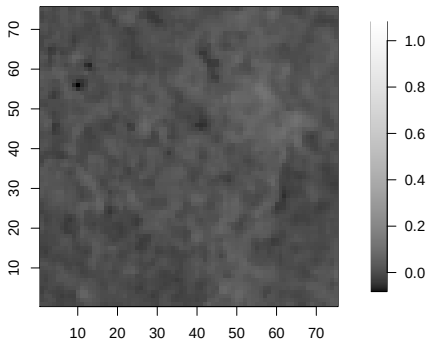


Figure 1: Observations of the seismic data.

$$
\begin{aligned}
[\mathbf{d} \mid \mathbf{l}] \sim p(\mathbf{d} \mid \mathbf{l}) &= \prod_{i=1}^{n} p(d_i \mid \mathbf{l}) = \prod_{i=1}^{n} p(d_i \mid l_i) \\
&= (2\pi\sigma^2)^{-n/2} \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2\sigma^2}(d_i - \mu_{l_i})^2 \right\}.
\end{aligned} \tag{1}
$$

b)

Consider a uniform, independence prior model for $\mathbf{l}$, that is

$$\mathbf{l} \sim p(\mathbf{l}) = \text{const.}$$

The posterior is then computed as follows:

$$
\begin{aligned}
[\mathbf{l} \mid \mathbf{d}] \sim p(\mathbf{l} \mid \mathbf{d}) &= \text{const}_1 \times p(\mathbf{l}, \mathbf{d}) = \text{const}_1 \times p(\mathbf{d} \mid \mathbf{l}) p(\mathbf{l}) \\
&= \text{const}_2 \times \prod_{i=1}^{n} p(d_i \mid l_i) \\
&= \text{const}_3 \times \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2\sigma^2}(d_i - \mu_{l_i})^2 \right\} \\
&= \text{const}_3 \times \prod_{i=1}^{n} \left(\mathrm{e}^{-\frac{1}{2\sigma^2}(d_i - \mu_1)^2} \right)^{l_i} \left(\mathrm{e}^{-\frac{1}{2\sigma^2}(d_i - \mu_0)^2} \right)^{1-l_i},
\end{aligned}
$$

where we notice that the marginal posteriors are independent Bernoulli variables. That is,

$$[l_i \mid \mathbf{d}] \sim p(l_i \mid \mathbf{d}) = \theta_i^{l_i}(1-\theta_i)^{1-l_i} \sim \text{Bernoulli}(\theta_i), \quad i = 1, \ldots, n,$$

with "success" ($l_i = 1$) probability

$$\theta_i = \frac{\exp\left\{ -\frac{1}{2\sigma^2}(d_i - \mu_1)^2 \right\}}{\exp\left\{ -\frac{1}{2\sigma^2}(d_i - \mu_0)^2 \right\} + \exp\left\{ -\frac{1}{2\sigma^2}(d_i - \mu_1)^2 \right\}}. \tag{2}$$

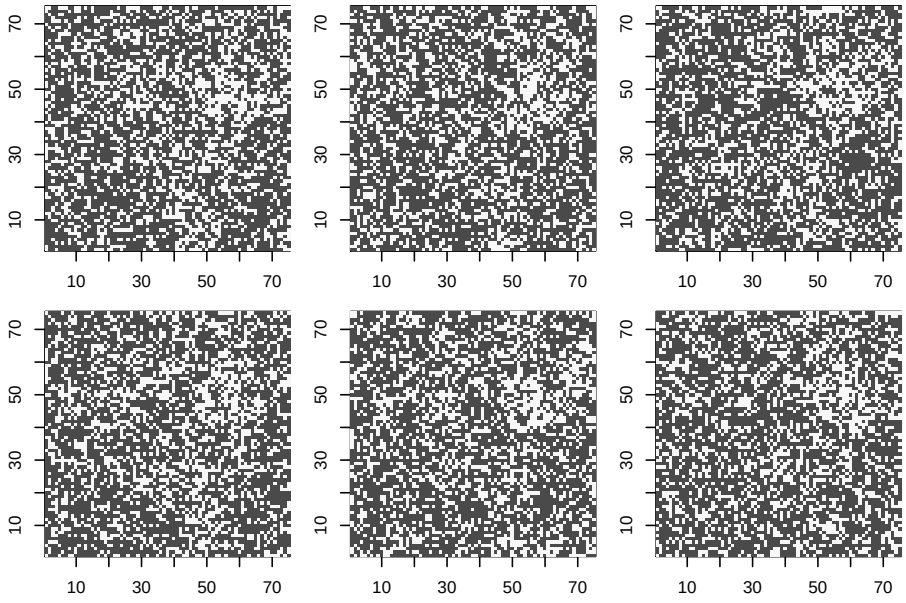


Figure 2: Six realizations of $[\mathbf{l} \mid \mathbf{d}]$, with uniform prior, where black and white colors corresponds to sand and shale, respectively.

Thus, the resulting expression of the posterior is

$$[\mathbf{l} \mid \mathbf{d}] \sim p(\mathbf{l} \mid \mathbf{d}) = \prod_{i=1}^{n} p(l_i \mid \mathbf{d}) = \prod_{i=1}^{n} \theta_i^{l_i} (1-\theta_i)^{1-l_i}, \tag{3}$$

where θ_i is given by Expression (2).

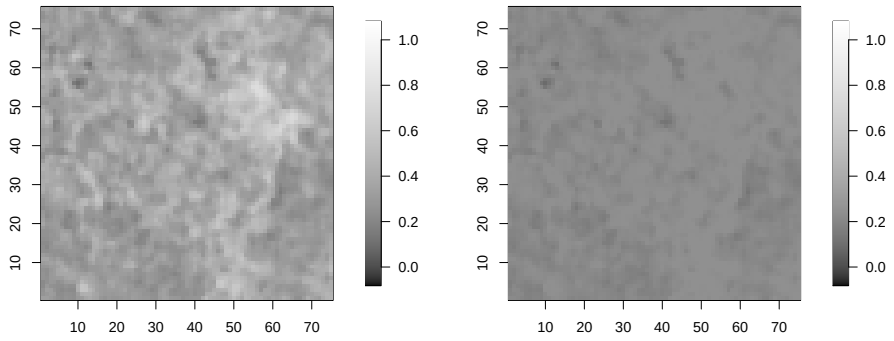
Six realizations of the posterior with a uniform prior is shown in Figure 2. The sand and shale distribution appears to be very mixed, with a higher proportion of shale in the upper right quadrant. This brighter area corresponds well with the observations in Figure 1.

As the marginal posteriors $[l_i \mid \mathbf{d}], i = 1, \ldots, n$ are independent, we can easily attain the posterior expectation $\mathrm{E}[\mathbf{l} \mid \mathbf{d}]$, and the diagonal terms of the posterior variance matrix $\mathrm{Var}[\mathbf{l} \mid \mathbf{d}]$:

$$\begin{aligned} \big[\mathrm{E}[\mathbf{l} \mid \mathbf{d}]\big]_i &= \mathrm{E}[l_i \mid \mathbf{d}] = \theta_i, \\ \big[\mathrm{diag}(\mathrm{Var}[\mathbf{l} \mid \mathbf{d}])\big]_i &= \mathrm{Var}[l_i \mid \mathbf{d}] = \theta_i(1-\theta_i), \end{aligned} \qquad i = 1, \ldots, n, \tag{4}$$

with θ_i as defined in Expression (2).

A map of the posterior expectation and the diagonal terms of the posterior variance is shown in Figure 3. We see that the posterior expectation matches the pattern seen in the observations in Figure 1. The posterior variance appears to be rather constant, and close to maximal variance, which is $\max_{\theta \in (0,1)} \{\theta(1-$



(a) Posterior expectation (b) Posterior variance

Figure 3: Posterior expectation $\mathrm{E}[\mathbf{l} \mid \mathbf{d}]$ and variance $\mathrm{diag}(\mathrm{Var}[\mathbf{l} \mid \mathbf{d}])$ with uniform prior.

$\theta)\} = 0.25$. That is, it is hard to capture any information with confidence based only on the observations and a uniform prior.

The i'th entry of the maximum marginal posterior (MMAP) predictor is found as follows:

$$\begin{aligned}
\big[\,\mathrm{MMAP}[\mathbf{l} \mid \mathbf{d}]\big]_i &= \arg\max_{l_i \in \{0,1\}} \{p(l_i \mid \mathbf{d})\} \\
&= \arg\max_{l_i \in \{0,1\}} \left\{ \exp\left\{ -\frac{1}{2\sigma^2}(d_i - \mu_{l_i})^2 \right\} \right\} \\
&= \arg\min_{l_i \in \{0,1\}} \left\{ (d_i - \mu_{l_i})^2 \right\} \\
&= I\left(d_i > \frac{1}{2}(\mu_0 + \mu_1) \right) \\
&= I(d_i > 0.05), \quad i = 1, \ldots, n,
\end{aligned} \tag{5}$$

where $I(A)$ equal 1 if A is true and equal 0 else. In fact, since the marginal posteriors are independent, the maximum posterior (MAP) is equal to the maximum marginal posterior (MMAP). The MMAP is shown in Figure 4. It shows that we have an area with shale in the upper quadrant, as we have seen in the realizations in Figure 2 and the observations in Figure 1. This is a reasonable guess based only on the observations, and no prior knowledge about the lithology distribution.

c)

Consider a Markov RF prior for $\{l(\mathbf{x}); \mathbf{x} \in \mathtt{L_D}\}$, with clique-system $\mathbf{c}_\mathtt{L}$ consisting of the two closest neighbors on the grid $\mathtt{L_D}$. The Gibbs formulation for the prior is

$$\mathbf{l} \sim p(\mathbf{l}) = \mathrm{const} \times \prod_{\mathbf{c} \in \mathbf{c}_\mathtt{L}} v_{1l}(l_i; i \in \mathbf{c}) = \mathrm{const} \times \prod_{<i,j> \in \mathtt{L_D}} \beta^{I(l_i = l_j)}, \tag{6}$$

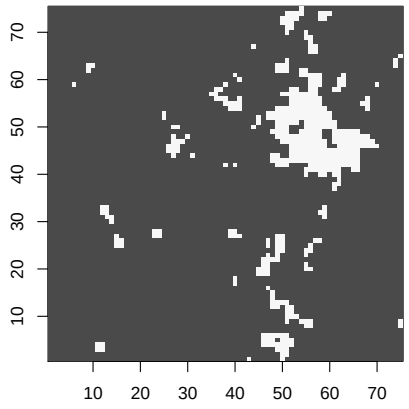


Figure 4: Maximum marginal posterior (MMAP) predictor, where the colors black and white corresponds to sand and shale, respectively.

with $\beta \in \mathbb{R}_{[1,\infty)}$, and where $< i, j > \in \mathtt{L_D}$ defines the set of two-closest neighbors on the grid $\mathtt{L_D}$.

The corresponding Markov formulation for $k = 1, 2, \ldots, n$, is,

$$\begin{aligned}[l_k \mid \mathbf{l}_{-k}] \sim p(l_k \mid \mathbf{l}_{-k}) &= p(l_k \mid l_j; j \in \mathbf{n}_k) \\ &= \text{const} \times \prod_{j \in \mathbf{n}_k} \beta^{I(l_k = l_j)} \end{aligned} \tag{7}$$

where $\mathbf{n}_k$ is the neighborhood for node k, given by

$$\mathbf{n}_k = \bigcup_{\mathbf{c} \in \mathbf{c}_{\mathtt{L}|k}} \{m \mid m \in \mathbf{c}\} \setminus k, \quad \text{with} \quad \mathbf{c}_{\mathtt{L}|k} = \{\mathbf{c} \in \mathbf{c}_{\mathtt{L}} \mid k \in \mathbf{c}\}.$$

The posterior model $p(\mathbf{l} \mid \mathbf{d})$ is obtained as follows,

$$\begin{aligned}[\mathbf{l} \mid \mathbf{d}] \sim p(\mathbf{l} \mid \mathbf{d}) &= \text{const} \times p(\mathbf{l}, \mathbf{d}) = \text{const} \times p(\mathbf{d} \mid \mathbf{l}) p(\mathbf{l}) \\ &= \text{const}_1 \times \prod_{i=1}^{n} p(d_i \mid l_i) \times \prod_{<i,j> \in \mathtt{L_D}} \beta^{I(l_i = l_j)} \\ &= \frac{1}{\text{const}_G} \times \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2\sigma^2} (d_i - \mu_{l_i})^2 \right\} \times \prod_{<i,j> \in \mathtt{L_D}} \beta^{I(l_i = l_j)}, \end{aligned} \tag{8}$$

where the normalizing constant

$$\text{const}_G = \sum_{\mathbf{l}' \in \mathbb{L}^n} \left(\prod_{i=1}^{n} \exp\left\{ -\frac{1}{2\sigma^2} (d_i - \mu_{l'_i})^2 \right\} \times \prod_{<i,j> \in \mathtt{L_D}} \beta^{I(l'_i = l'_j)} \right),$$

is computationally demanding to calculate. The posterior model with the Markov formulation, $p(l_k \mid \mathbf{d}, \mathbf{l}_{-k})$, $k = 1, \ldots, n$, is obtained similarly,

$$\begin{aligned}
\left[l_k \,\middle|\, \begin{matrix} \mathbf{d} \\ \mathbf{l}_{-k} \end{matrix} \right] &\sim p(l_k \mid \mathbf{d}, \mathbf{l}_{-k}) \\
&= \text{const} \times p(l_k, d_k \mid \mathbf{d}_{-k}, \mathbf{l}_{-k}) \\
&= \text{const} \times p(d_k \mid \mathbf{d}_{-k}, \mathbf{l}) p(l_k \mid \mathbf{d}_{-k}, \mathbf{l}_{-k}) \\
&= \text{const} \times p(d_k \mid l_k) p(l_k \mid \mathbf{l}_{-k}) \\
&= \text{const}_1 \times \exp\left\{ -\frac{1}{2\sigma^2} (d_k - \mu_{l_k})^2 \right\} \\
&\quad \times \prod_{j \in \mathbf{n}_k} \beta^{I(l_k = l_j)} \\
&= \text{const}_1 \times \left(\exp\left\{ -\frac{1}{2\sigma^2} (d_k - \mu_1)^2 + \log \beta \times \sum_{j \in \mathbf{n}_k} l_j \right\} \right)^{l_k} \\
&\quad \times \left(\exp\left\{ -\frac{1}{2\sigma^2} (d_k - \mu_0)^2 + \log \beta \times \sum_{j \in \mathbf{n}_k} (1 - l_j) \right\} \right)^{1 - l_k} \\
&\sim \text{Bernoulli}(\lambda_k),
\end{aligned} \tag{9}$$

with "success" probability

$$\lambda_k = \frac{\mathrm{e}^{-\frac{1}{2\sigma^2}(d_k - \mu_1)^2 + \log \beta \times \sum_{j \in \mathbf{n}_k} l_j}}{\mathrm{e}^{-\frac{1}{2\sigma^2}(d_k - \mu_0)^2 + \log \beta \times \sum_{j \in \mathbf{n}_k} (1 - l_j)} + \mathrm{e}^{-\frac{1}{2\sigma^2}(d_k - \mu_1)^2 + \log \beta \times \sum_{j \in \mathbf{n}_k} l_j}}.$$

In the last equality of Expression (9), the following identities were used,

$$\beta^{I(l_k = l_j)} = \left(\beta^{l_j}\right)^{l_k} \left(\beta^{1 - l_j}\right)^{1 - l_k}, \quad j \in \mathbf{n}_k, \quad k = 1, \ldots, n, \quad \text{and,}$$

$$\prod_{j \in \mathbf{n}_k} \beta^{l_j} = \beta^{\sum_{j \in \mathbf{n}_k} l_j} = \exp\left\{ \log \beta \times \sum_{j \in \mathbf{n}_k} l_j \right\}, \quad k = 1, \ldots, n.$$

The observations from the geologically comparable domain $\mathtt{D}_c$ given by the data, `complit.dat`, is shown in Figure 5. Assuming these observations are exact, we have a training image. We denote the observations from $\mathtt{D}_c$ by the m-vector $\mathbf{l}^o, l_k^o \in \mathbb{L} : \{0, 1\}, k = 1, \ldots, m$. We use this training image to estimate

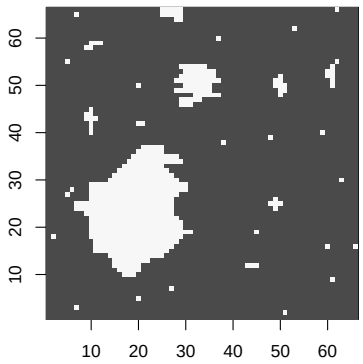


Figure 5: Observations from the geologically comparable domain $\mathtt{D}_c$, where the colors black and white corresponds to sand and shale, respectively.

the maximum marginal pseudo-likelihood (MMpL) estimator of β, as follows,

$$
\begin{aligned}
\hat{\beta} &= \arg\max_{\beta} \left\{ \prod_{k=1}^{m} p(l_i^o \mid l_j^o; j \in \mathbf{n}_k; \beta) \right\} \\
&= \arg\max_{\beta} \left\{ \prod_{k=1}^{m} \left(\left[\sum_{l'_k \in \{0,1\}} \prod_{j \in \mathbf{n}_k} \beta^{I(l'_k = l_j^o)} \right]^{-1} \prod_{j \in \mathbf{n}_k} \beta^{I(l_k^o = l_j^o)} \right) \right\} \\
&= \arg\max_{\beta} \left\{ \sum_{k=1}^{m} \log \left(\left[\sum_{l'_k \in \{0,1\}} \prod_{j \in \mathbf{n}_k} \beta^{I(l'_k = l_j^o)} \right]^{-1} \prod_{j \in \mathbf{n}_k} \beta^{I(l_k^o = l_j^o)} \right) \right\} \\
&= \arg\max_{\beta} \left\{ \sum_{k=1}^{m} \left(\log \beta \sum_{j \in \mathbf{n}_k} I(l_k^o = l_j^o) \right. \right. \\
&\qquad \left. \left. - \log \sum_{l'_k \in \{0,1\}} \exp \left\{ \log \beta \times \sum_{j \in \mathbf{n}_k} I(l'_k = l_j^o) \right\} \right) \right\},
\end{aligned}
$$

which we solve numerically with optimization. The MMpL estimate is found to be

$$\hat{\beta} = 3.557691.$$

We use this estimate for β to generate realizations from the posterior, $p(\mathbf{l} \mid \mathbf{d})$. This is done with a McMC procedure, more precisely, a Gibbs algorithm, which is presented in Algorithm 1, where one simulation sweep, s, corresponds to one visit per node in expectation. To avoid boundary issues when we visit the neighborhood $\mathbf{n}_k$, $k = 1, \ldots, n$, we wrap the domain $\mathtt{D}$ like a torus.

We run the algorithm three times, for three different initial values for $\mathbf{l}^0$; all sand (all zeros), all shale (all ones) and lastly a uniformly random distribution

Algorithm 1 Gibbs algorithm to sample from posterior

procedure GIBBSPOSTERIOR($\hat{\beta}, n, n_s$)
  Initiate :
  $\mathbf{l}^0$ such that $p(\mathbf{l}^0 \mid \mathbf{d}) > 0$
  for $s = 1, 2, ..., n_s$ **do**
    $\mathbf{l}^s \leftarrow \mathbf{l}^{s-1}$
    for $i = 1, 2, ..., n$ **do**
      $k \sim \text{Uniform}(1, 2, ..., n)$
      $l_k^s \sim p(l_k \mid d_k, l_j^s; j \in \mathbf{n}_k, \hat{\beta})$
  return $\{\mathbf{l}^0, \mathbf{l}^1, ..., \mathbf{l}^{n_s}\}$ ▷ The sample including the burn-in

of sand and shale. Additionally, we compute the proportion of sand for each sweep, s, and use this as a convergence criterion. The convergence plot is shown in Figure 6. We observe that the initial guess of uniformly distributed sand/shale converges after less than 50 sweeps. The all shale distribution takes about 150 sweeps before it converges to the right proportion of sand, while the initial guess of all sand uses about 250 sweeps before it has the desired sand proportion.

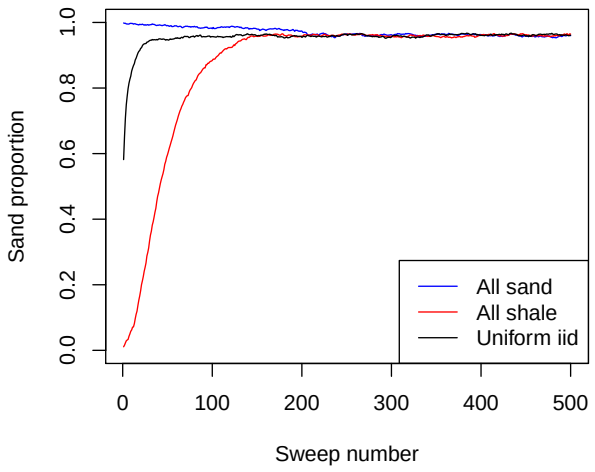


Figure 6: Convergence plot in terms of proportion of sand.

Six realizations from the posterior in Expression (9) are displayed in Figure 7. We choose to show two realizations from each of the three chains with initial guesses; all sand, all shale and uniformly iid. We choose two realizations with more than 150 sweeps in between them, to ensure independence. We observe

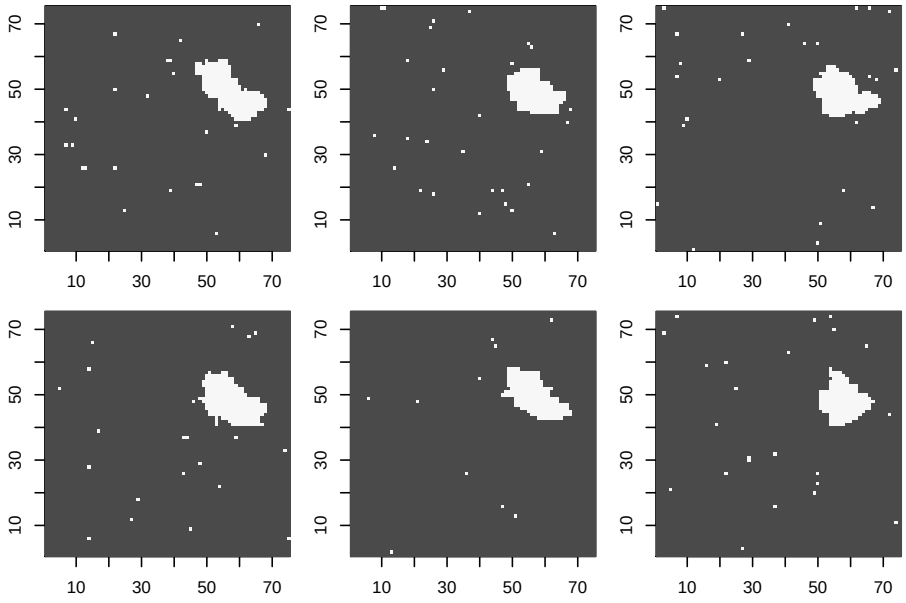


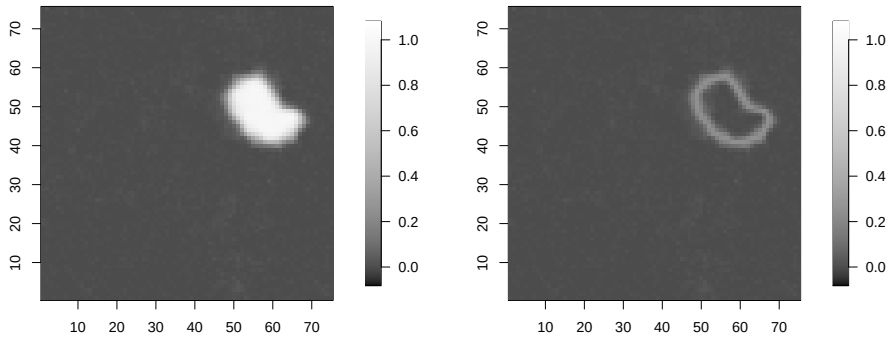
Figure 7: Six realizations of $[\mathbf{l} \mid \mathbf{d}]$, with a Markov RF prior, where black and white colors corresponds to sand and shale, respectively.

that all realizations have a white area in the upper right corner. Additionally there are some white areas scattered around the domain for all six realizations. This seems reasonable in comparison to the geologically comparable domain shown in Figure 5, which has the same characteristics.

Based on realizations from the posterior using Algorithm 1, we would like to find estimates of $\mathrm{E}[\mathbf{l} \mid \mathbf{d}]$ and the diagonal entries of $\mathrm{Var}[\mathbf{l} \mid \mathbf{d}]$. To do so, we use the three chains displayed in Figure 6 after convergence. That is after 250, 150 and 50 iteration sweeps for the all sand, all shale and uniform iid initial guesses for $\mathbf{l}^0$, respectively. As the converged chains should come from the same distribution, we compute mean and variance based on all realizations collectively. The estimated predictor $\mathrm{E}[\mathbf{l} \mid \mathbf{d}]$ and diagonal terms of $\mathrm{Var}[\mathbf{l} \mid \mathbf{d}]$ is shown in Figure 8.

We observe that the mean of the posterior, $\mathrm{E}[\mathbf{l} \mid \mathbf{d}]$, has shale in the upper right quadrant as observed in the realizations in Figure 7. Outside of this area, the expected value is close to zero, which suggests that this region is more likely to consist of sand. Further, the posterior variance, $\mathrm{diag}(\mathrm{Var}[\mathbf{l} \mid \mathbf{d}])$, displayed in Figure 8b tells us that the uncertainty is located on the border between the shale area in the upper right corner, and sand surrounding it. Elsewhere, the variance seems to be constant, and close to zero.

Figure 9 displays the MMAP predictor for the posterior when we use a Markov RF prior. The MMAP in this case can be approximated by the realiza-



(a) Posterior expectation (b) Posterior variance

Figure 8: Posterior expectation $\mathrm{E}[\mathbf{l} \mid \mathbf{d}]$ and variance $\mathrm{diag}(\mathrm{Var}[\mathbf{l} \mid \mathbf{d}])$ with a Markov RF prior.

tions as follows.

$$\begin{aligned}\left[\mathrm{MMAP}[\mathbf{l} \mid \mathbf{d}]\right]_k &= \underset{l_k \in \{0,1\}}{\arg\max} \{p(l_k \mid \mathbf{d})\} \\ &\approx \underset{l \in \{0,1\}}{\arg\max} \left\{ \sum_{s=1}^{n_s} I\left([l_k \mid \mathbf{d}]^s = l\right) \right\} \\ &= I\left(\frac{1}{n_s} \sum_{s=1}^{n_s} [l_k \mid \mathbf{d}]^s > 0.5 \right), \quad k = 1, \ldots, n,\end{aligned} \tag{10}$$

where $[l_k \mid \mathbf{d}]^s$ denotes the s'th realization, and n_s the total number of realizations. Again, we observe the shale in the upper right quadrant of the domain, very similar to the predictor $\mathrm{E}[\mathbf{l} \mid \mathbf{d}]$, in Figure 8a, with the only visible difference being at the border between sand and shale where the diagonal terms of the variance $\mathrm{Var}[\mathbf{l} \mid \mathbf{d}]$ is nonzero, as seen in Figure 8b.

d)

The greatest noticeable difference between the methods in subsection **b)** and **c)** is seen in the realizations of the two posteriors, in Figure 2 and 7, respectively. With the training image with a geologically comparable domain shown in Figure 5, we see that the realizations with a uniform matches poorly, as we see no connection between the patterns. The realizations with the Markov prior shown in Figure 7 matches quite well with the training image, although the training image suggests there should be several "medium"-sized regions of shale, as the realizations only show one large region of shale and several single-nodes of shale.

When comparing Figure 3 and 8 with the observations in Figure 1, we note that the expected posterior with a uniform prior is more similar to the observations. On the other hand, the expected posterior with a Markov RF prior has a more distinct prediction on the distribution of sand and shale. This is

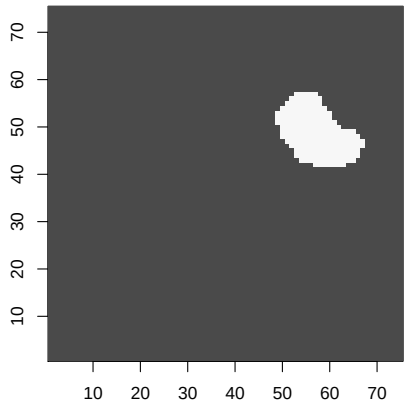


Figure 9: Maximum marginal posterior (MMAP) predictor with a Markov RF prior, where the colors black and white corresponds to sand and shale, respectively.

also stressed by the posterior variance shown in Figure 3b and 8b. The variance is larger and almost constant on the whole domain with a uniform prior, while when applying the Markov RF prior, we achieve a posterior with lower variance.

Lastly we would like to compare the MMAP predictions in Figure 4 with a uniform prior to the MMAP prediction in Figure 9 with a Markov RF prior. We observe that the MMAP prediction with a uniform prior predicts more shale sporadic on the domain, while the same prediction with a more informative prior predict only one area with shale in the upper right corner.

Through inspection of the posterior distribution $[\mathbf{l} \mid \mathbf{d}]$ using two different priors; constant prior and a Markov RF prior, we have learned the effect of an informative prior. This assumption must be done a priori, and that requires some knowledge about the system to be studied. In this problem, we approximated the prior distribution assuming interactions between two nearest nodes, using a maximum marginal pseudo-likelihood estimate β on the training image to estimate the interaction. If no such information is known a priori, a constant prior will result in a posterior distribution not too far from the observations, and thus, not providing any new interesting discoveries.