

# Exercise 3

*Kwaku Peprah Adjei, Amir Ahmed*

## Problem 1: Markov RF

Assume that we have observed seismic data over a domain  $D \in \mathbb{R}^2$ . We want to identify the underlying lithology distribution over  $D$ , the underlying lithology of a point is either sand or shale,  $\{1, 0\}$  respectively.

The observations have been collected on a regular  $(75 \times 75)$  grid  $L_d$ , with seismic data being  $\{d(\mathbf{x}); \mathbf{x} \in L_d\}$ . Where  $d(\mathbf{x}) \in \mathbb{R}$ .

We have observed the lithology distribution in a geologically comparable domain  $D_c \in \mathbb{R}^2$ . Assume that this was collected on a regular  $(66 \times 66)$  grid  $L_{D_c}$ .

We assume that the underlying lithology distribution can be represented by a Mosaic RF  $\{l(\mathbf{x}); \mathbf{x} \in L_D\}, l(\mathbf{x}) \in \{0, 1\}$ .

### Problem 1a)

We start by looking at  $L_d$ . Let the seismic data collection procedure follow the following likelihood model:

$$[d_i | \mathbf{l}] = \begin{cases} 0.02 + U_i & \text{if sand, } l_i = 0 \\ 0.08 + U_i & \text{if shale, } l_i = 1 \end{cases}$$

$i = 1, 2, \dots, n$ . With  $U_i$  being identically independently distributed  $U_i \sim N(0, 0.06^2)$ . This would make each observation point  $d_i$  conditionally independent on  $\mathbf{l}$ . That will say:

$$p(d_i | \mathbf{l}) = p(d_i | l_i) = \phi(d_i; \mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \quad (1)$$

Where  $\phi$  is the pdf of the normal distribution. As all observations are independent we thus have:

$$p(\mathbf{d} | \mathbf{l}) = \prod_{i=1}^n p(d_i | l_i) = \prod_{i=1}^n \phi(d_i; \mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \quad (2)$$

We display the observations from  $L_D$  as a map in Figure 1, there seems to be one large gathering where  $d(\mathbf{x})$  takes on relatively large values, there also seems to be some smaller gatherings of large  $d(\mathbf{x})$  in areas centered around the large one.

### Problem 1b)

We now consider a uniform, independence prior model on  $\mathbf{l}$ . That will say:

$$p(\mathbf{l}) = \text{const} \quad (3)$$

We note that since the prior is constant we have:

$$p(\mathbf{l} | \mathbf{d}) \propto p(\mathbf{d} | \mathbf{l}) \quad (4)$$

We get the following posterior model using bayes law and the law of total probability:

$$p(\mathbf{l} | \mathbf{d}) = \frac{p(\mathbf{d} | \mathbf{l})}{\sum_{\mathbf{l} \in \mathbb{L}^n} p(\mathbf{d} | \mathbf{l})} \quad (5)$$

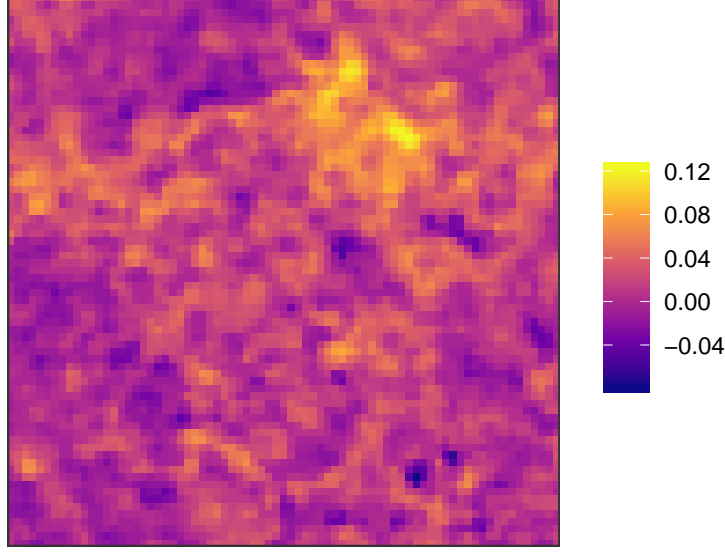


Figure 1: Display of seismic data  $L_D$ .

Inserting from (2) we get:

$$p(\mathbf{l}|\mathbf{d}) = \frac{\prod_{i=1}^n \phi(d_i; \mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2)}{\sum_{\mathbf{l}' \in \mathbb{L}^n} \prod_{i=1}^n \phi(d_i; \mu = 0.02 + 0.06l'_i, \sigma^2 = 0.06^2)} \quad (6)$$

Where  $\mathbb{L}^n$  is the n-dimensional space representing all possible values which  $l$  can take.

As the prior is independent each point would also be conditional independent, for each point we thus get the following. Let:

$$\begin{aligned} p_i &= p(l_i = 1|d_i) \\ &= \frac{p(d_i|l_i = 1)}{p(d_i|l_i = 0) + p(d_i|l_i = 1)} \\ &= \frac{\phi(d_i; \mu = 0.08, \sigma^2 = 0.06^2)}{\phi(d_i; \mu = 0.02, \sigma^2 = 0.06^2) + \phi(d_i; \mu = 0.08, \sigma^2 = 0.06^2)} \end{aligned} \quad (7)$$

As each point either is sand or shale we get:

$$1 - p_i = p(l_i = 0|d_i) \quad (8)$$

We recognize this conditioned model as something Bernoulli-distributed with probability  $p_i$ .

We thus have:

$$E(l_i|d_i) = p_i \quad (9)$$

and

$$Var(l_i|d_i) = p_i(1 - p_i) \quad (10)$$

We simulate 6 trials with the data and display the results in Figure ??.

The maximum marginal posterior predictor ( $MMAP\{\mathbf{l}|\mathbf{d}\}$ ) is defined as:

$$MMAP\{\mathbf{l}|\mathbf{d}\} = \hat{\mathbf{l}} = \underset{\mathbf{l} \in \mathbb{L}^n}{\operatorname{argmax}} \{p(\mathbf{l}|\mathbf{d})\} \quad (11)$$

Due to the conditional independence of the points we see:

$$\hat{l}_i = \begin{cases} 0, & \text{if } p_i < 0.5 \\ 1, & \text{if } p_i \geq 0.5 \end{cases} \quad (12)$$

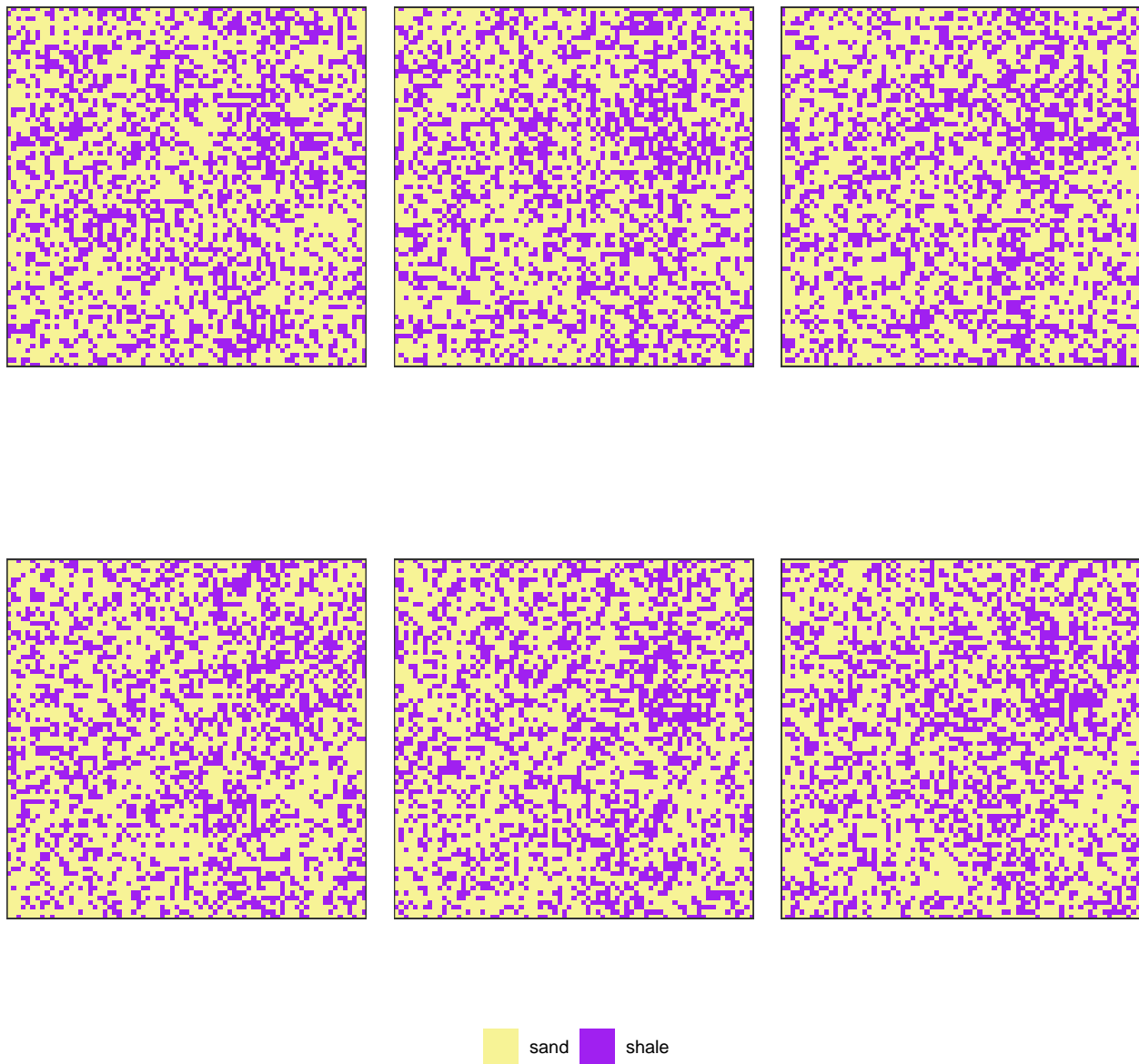


Figure 2: Display of six posterior realizations of  $L_D$ .

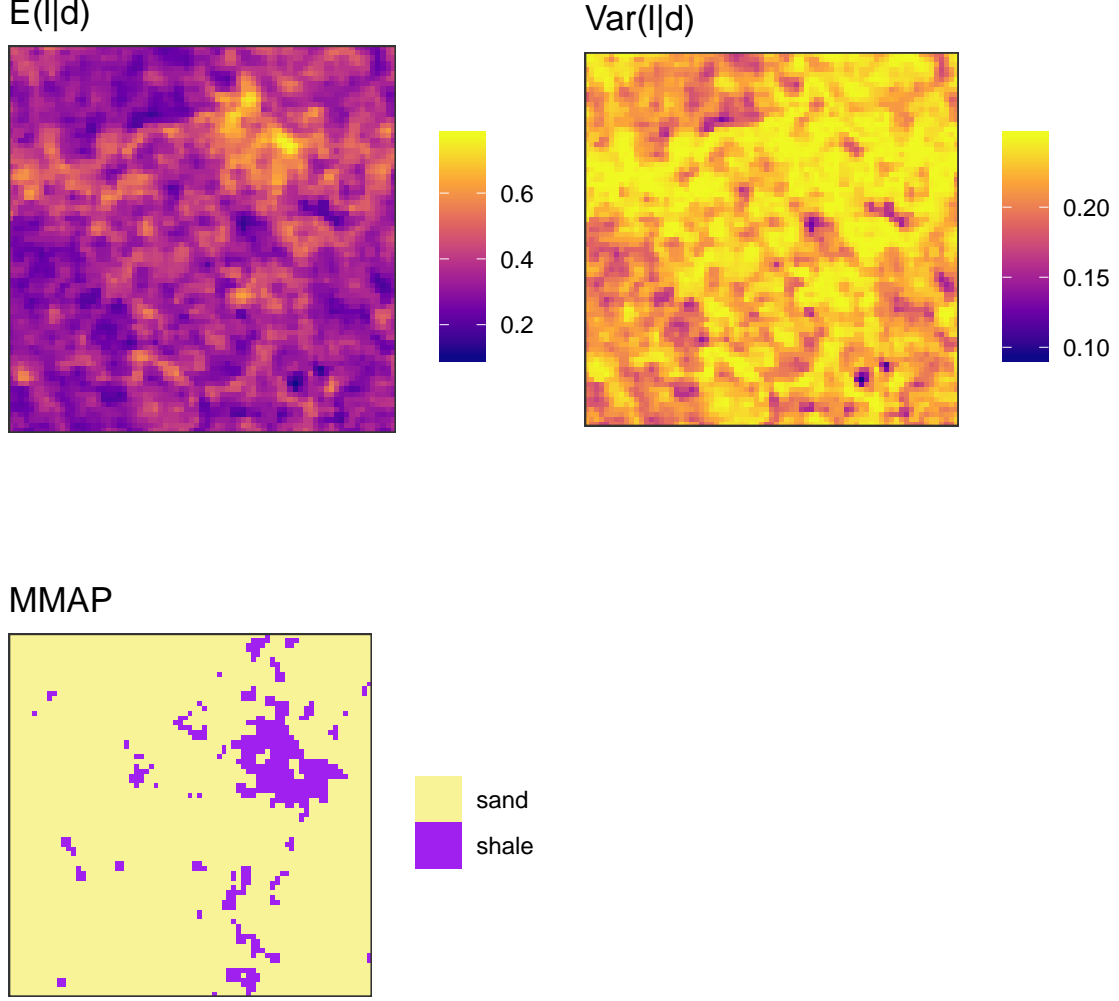


Figure 3: Plot of MMAP, expected values and variance with uniform prior

is a MMAP solution.

We plot MMAP solution, expectance and variance in Figure 3. We see that there is relatively high variance in most parts of the map, this is reflected in the large difference we see between the simulations. From both the expected value and the *MMAP* we see that there is one large spot (top-right) where we expect a large cluster of shale as with some bands of shale on mid-left. This is somewhat reflected in the simulations. A difference between MMAP and the simulations is that the simulations tend to expect more shale than the MMAP. The map of the expected values seems to be closer to the simulations than the MMAP. The MMAP result seems much less noisy.

### Problem 1c)

Now consider a Markov RF prior model for  $\{l(\mathbf{x}); \mathbf{x} \in L_D\}$ . Represented by the n-vector  $\mathbf{l}$  with the clique system  $\mathbf{c}_L$  consisting of two closest neighbors on the grid  $L_D$ .

The corresponding Gibbs formulation is:

$$\begin{aligned}
 p(\mathbf{l}) &= \text{const} \times \prod_{\mathbf{c} \in \mathbf{c}_L} v_{\mathbf{l}l}(l_i, i \in \mathbf{c}) = \text{const} \times \prod_{\langle i, j \rangle \in L_D} \beta^{I(l_i = l_j)} \\
 &= \text{const} \times \beta^{\sum_{\langle i, j \rangle \in L_D} I(l_i = l_j)}
 \end{aligned} \tag{13}$$

With  $\langle i, j \rangle \in L_d$  defining the set of two closest neighbors on the grid  $L_D$ .

Want to find expressions for the posterior models and want to specify the Markov formulation for the Markov RF.

Want to find the Markov formulation for the Markov RF. First see:

$$p(l_i | \mathbf{l}_{-i}) = \frac{p(\mathbf{l})}{\sum_{l'_i \in \mathbb{L}} p(l'_i, \mathbf{l}_{-i})} = \frac{p(\mathbf{l})}{p(l_i = 1, \mathbf{l}_{-i}) + p(l_i = 0, \mathbf{l}_{-i})} \quad (14)$$

This reduces to:

$$p(l_i | \mathbf{l}_{-i}) = p(l_i | l_j, j \in n_i) \quad (15)$$

We note that the joint distribution is given by:

$$\begin{aligned} p(\mathbf{d}, \mathbf{l}) &= p(\mathbf{d} | \mathbf{l}) p(\mathbf{l}) \\ &= \text{const} \times \prod_{i=1}^n p(d_i | l_i) \prod_{\mathbf{c} \in \mathbf{c}_L} v_{1l}(l_i, i \in \mathbf{c}) \\ &= \text{const} \times \prod_{i=1}^n \phi(d_i; \mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \prod_{\langle i, j \rangle \in L_D} \beta^{I(l_i=l_j)} \end{aligned} \quad (16)$$

We input the above into the following:

$$p(\mathbf{l} | \mathbf{d}) = \frac{p(\mathbf{l}, \mathbf{d})}{p(\mathbf{d})} = \text{const} \times \prod_{i=1}^n \phi(d_i; \mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \prod_{\langle i, j \rangle \in L_D} \beta^{I(l_i=l_j)} \quad (17)$$

Also:

$$\begin{aligned} p(l_i, \mathbf{d} | \mathbf{l}_{-i}) &= p(\mathbf{d} | \mathbf{l}) p(l_i | \mathbf{l}_{-i}) \\ &= p(\mathbf{d} | \mathbf{l}) p(l_i | \mathbf{l}_{-i}) \\ &= \frac{p(\mathbf{l})}{p(l_i = 1, \mathbf{l}_{-i}) + p(l_i = 0, \mathbf{l}_{-i})} \prod_{i=1}^n \phi(d_i; \mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \end{aligned} \quad (18)$$

Let  $n_i$  be the neighborhood around the  $i$ th node. Then have:

$$p(l_i, \mathbf{d} | \mathbf{l}_{-i}) = \frac{\prod_{l_j \in n_i} \beta^{I(l_i=l_j)}}{\prod_{l_j \in n_i} \beta^{I(l_i=0=l_j)} + \prod_{l_j \in n_i} \beta^{I(l_i=1=l_j)}} \prod_{i=1}^n \phi(d_i; \mu = 0.02 + 0.06l_i, \sigma^2 = 0.06^2) \quad (19)$$

Now want to develop expressions for the posterior model  $p(l_i | \mathbf{d}, \mathbf{l}_{-i})$  have:

$$\begin{aligned} p(l_i | \mathbf{d}, \mathbf{l}_{-i}) &= p(l_i | d_i, l_j; j \in n_i) \\ &= \frac{p(l_i, d_i, l_j; j \in n_i)}{p(d_i, l_j; j \in n_i)} \\ &= \frac{p(d_i | l_i) \prod_{l_j \in n_i} \beta^{I(l_i=l_j)}}{\phi(d_i | \mu = 0.02, \sigma^2 = 0.06^2) \prod_{l_j \in n_i} \beta^{I(l_j=0)} + \phi(d_i | \mu = 0.08, \sigma^2 = 0.06^2) \prod_{l_j \in n_i} \beta^{I(l_j=1)}} \end{aligned} \quad (20)$$

as the Markov formulation.

We now display the observations in  $D_c$  in Figure 4.

Now want to use these observations to estimate  $\beta$  by a maximum pseudo-likelihood procedure. In an optimal solution we would use the assumed  $p(\mathbf{l})$  distribution to do this, however we would need  $2^n$  calculations to

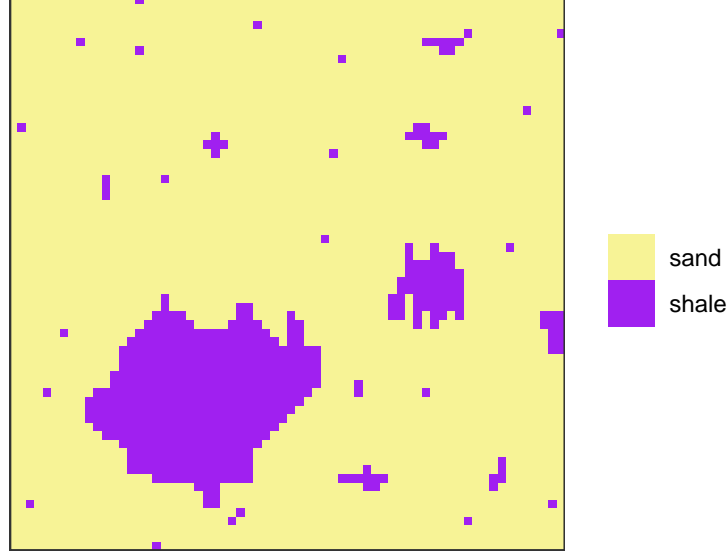


Figure 4: Display of observed data in  $D_c$

evaluate the normalizing constant, which in our case is infeasible. We rather use the Markov formulation to create an approximation. (here  $\mathbf{d} \in \mathbb{L}^n$ )

$$p(\mathbf{d}|\beta) \approx \hat{p}(\mathbf{d}|\beta) = \text{const} \times \prod_{i=1}^n \sum_{\{l'_i, l'_j | l'_j \in n_i\} \in L} \prod_{j=i, j \in n_i} p(d_j | l'_j) p(l'_i | l'_j) \quad (21)$$

As we assume the observations to  $\mathbf{l}$  from  $D_c$  be exact the model reduces as:

$$p(d_i | l'_i) \rightarrow \delta_{d_i}(l'_i) = \begin{cases} 1, & l'_i = l_i \\ 0, & \text{else} \end{cases} \quad (22)$$

This reduces (21) to:

$$\hat{p}(\mathbf{d}|\beta) \propto \prod_{i=1}^n p(d_i | l_j; j \in n_i; \beta) = \prod_{i=1}^n \frac{\beta^{\sum_{j \in n_i} I(l_i = l_j)}}{\beta^{\sum_{j \in n_i} I(0 = l_j)} + \beta^{\sum_{j \in n_i} I(1 = l_j)}} \quad (23)$$

A good estimation of  $\beta$  would be to maximize the above giving:

$$\hat{\beta} = \text{argmax}_{\beta} \sum_{i=1}^n \left\{ \left( \sum_{j \in n_i} I(l_i = l_j) \right) \log \beta - \log \left( \beta^{\sum_{j \in n_i} I(0 = l_j)} + \beta^{\sum_{j \in n_i} I(1 = l_j)} \right) \right\} \quad (24)$$

Using the Optim function in R this value can be “easily” found. Implementing the function and running we get that  $\hat{\beta} \approx 3.64$ .

We can use the information we learned from  $D_c$  and apply it to our model on  $L_d$  as their lithology is comparable.

Focus is on realizations from  $p(\mathbf{l}|\mathbf{d})$ , with related predictions  $E(\mathbf{l}|\mathbf{d})$ , variances in the diagonal terms of  $\text{Var}(\mathbf{l}|\mathbf{d})$ , and alternative predictions  $MMA\!P(\mathbf{l}|\mathbf{d})$ .

To estimate these, we use a MCMC/Gibbs algorithm with a single-site proposal based on the Markov formulation for the Markov RF. The pseudocode for the algorithm is given below.

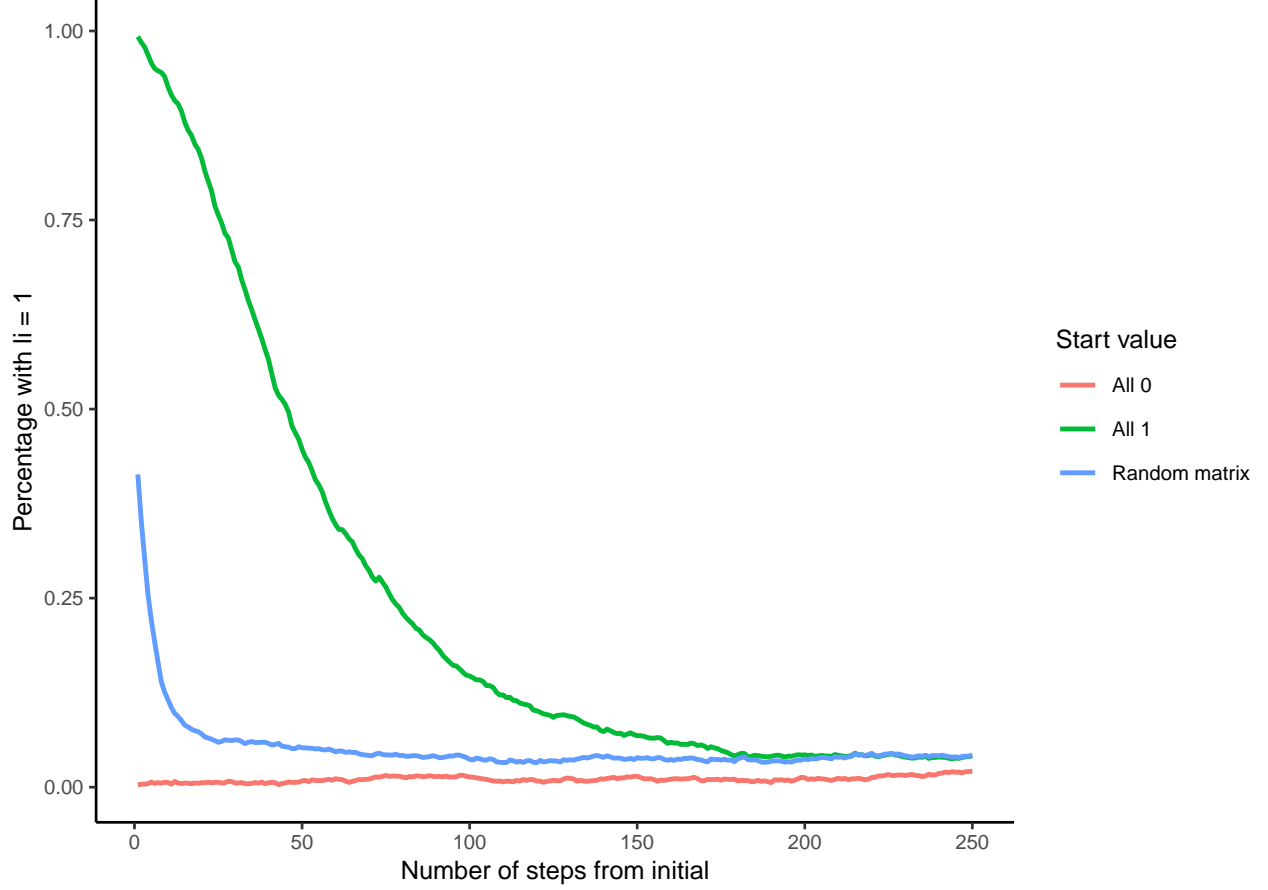


Figure 5: Display of the proportion of shale using different initial values after 250 iterations.

---

**Algorithm 1** MCMC/Gibbs algorithm for  $p(\mathbf{l}|\mathbf{d})$  realizations

---

```

 $\beta = \hat{\beta}$ 
 $\mathbf{l}^0 = p(\mathbf{l}^0|\mathbf{d}) > 0$ 
for all  $j = 1, 2, \dots$  do
   $\mathbf{l}^j \sim g(\mathbf{l}|\mathbf{l}^{j-1})$ 
end for

```

---



---

**Algorithm 2** Function  $g(\mathbf{l}'|\mathbf{l})$

---

```

 $i \sim Uniform(1, \dots, n)$ 
 $\mathbf{l}'_i \sim p(l_i|d_i, l_j, j \in n_i)$ 
 $\mathbf{l}' = (l_1, \dots, l_{i-1}, \mathbf{l}'_i, l_{i+1}, \dots, l_n)$ 
return  $\mathbf{l}'$ 

```

---

We implement and run the algorithm for three different initial values. One where we start with  $\mathbf{l}$  as all zeroes. One where we start with  $\mathbf{l}$  as all ones. And one where  $\mathbf{l}$  elements are randomly 1 or 0 with equal chance for each. To check convergence we plot percentage of  $l_i$ -s that are 1 compared to iteration numbers. The results are displayed in figure ???. We run the sampler for 250 iterations in each case. We see that the sampler starting at all 0 and all 1 meet after about 200 iterations we thus assume that they have converged. The random start reach this point the fastest, and seems to keep stationary after about 50 iterations. The one with all ones take 4 times as many iterations to reach the same point. The sampler starting at all 0 seems to

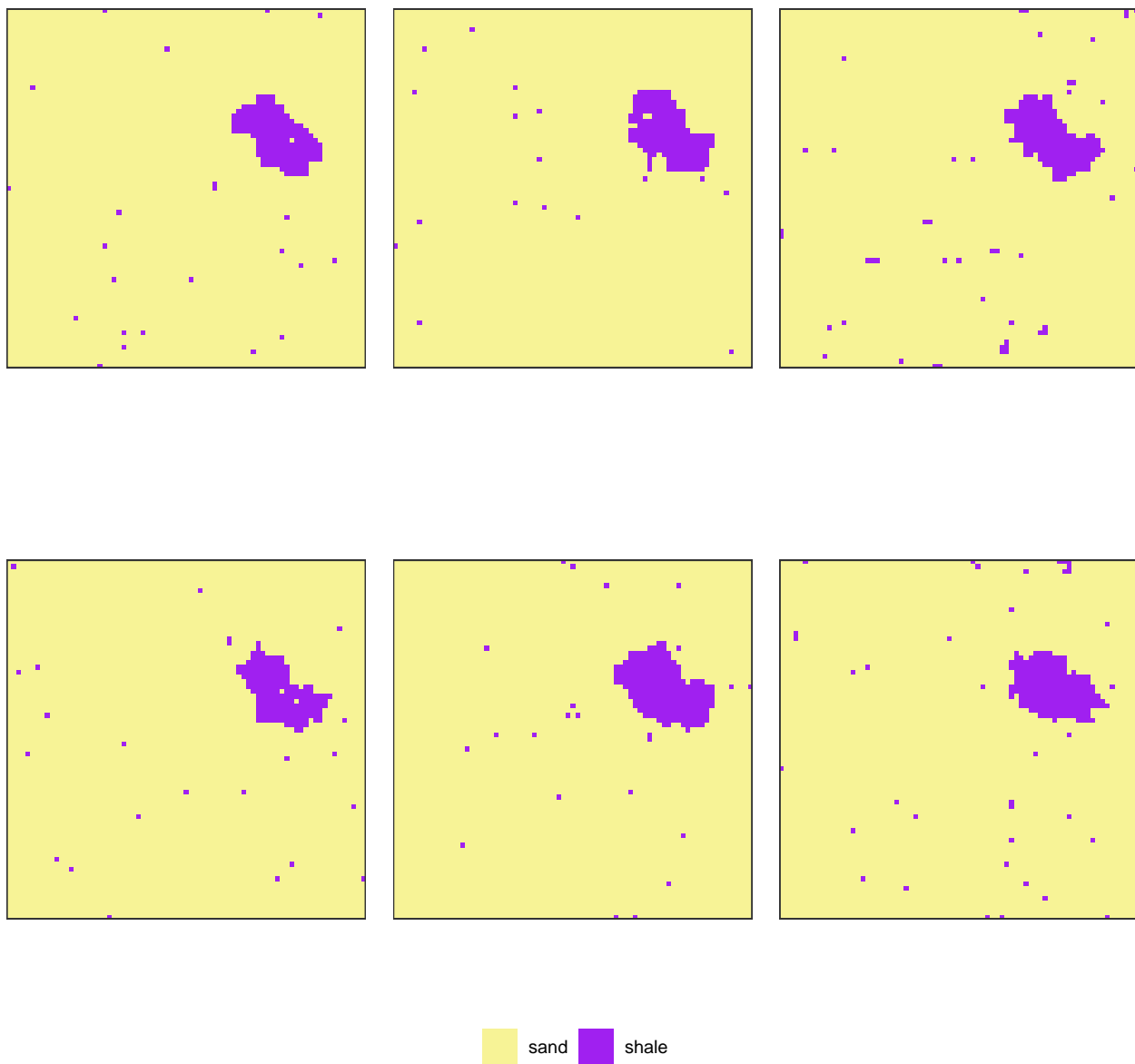


Figure 6: Six realisations from the MCMC/Gibbs algorithm.



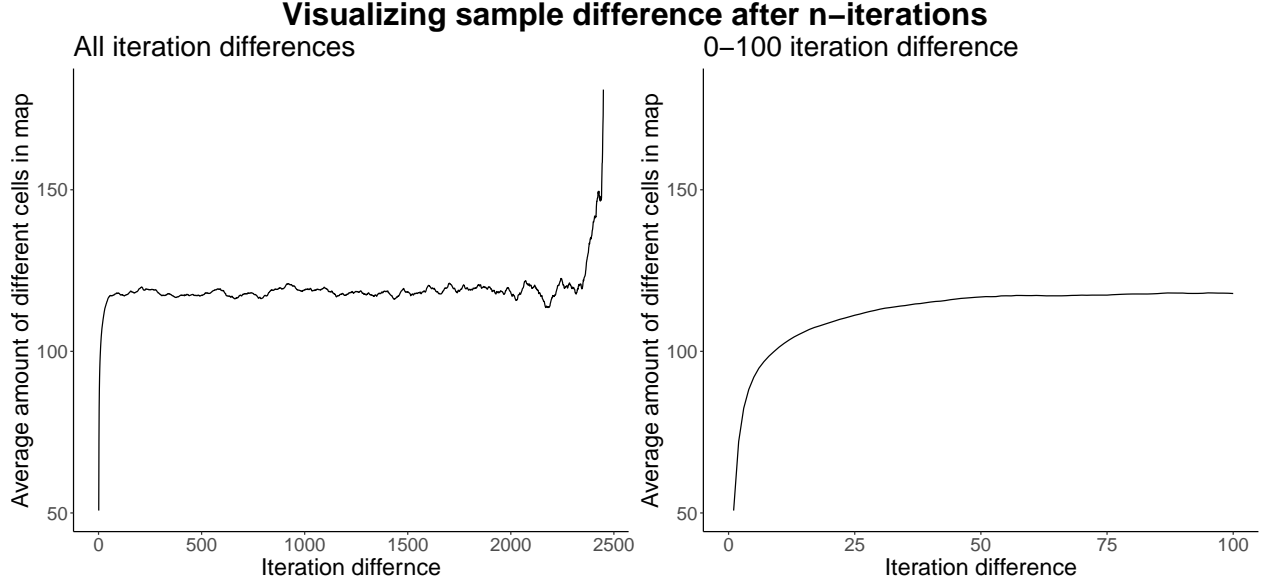


Figure 7: Average sample difference, 2450 samples, to check independence

slowly move towards the two others, but have not managed to do so after 250 iterations. We conclude that it is best to start with the random matrix.

To simulate realizations we run 2500 iterations with the sampler starting with the random matrix. We assume burnin after 50 iterations. After that we take each sweep as a sample. First and last iteration is displayed in Figure 6.

To control for sample independence we calculate how many different cells we have between each cell in the sample. We then plot the average difference controlled by iteration difference. We remove burnin data. The results are displayed in Figure 7. After about 50 iterations the difference seems to be stabilizing at approximately 110 different cells in each map.

We thus decide on taking each 50-th step as a independent sample.

An alternative to this method could i.e. be based on correlation.

Using our independent samples we have following estimator:

$$E(l_i|d_i) = p_i = \frac{1}{m} \sum_{j=1}^m l_i^j \quad (25)$$

Where  $m$  is the number of independent samples, and  $l_i^j$  is the  $i$ -th cell in the  $j$ -th sample. Estimators for variance and MMAP are (10) and (11) respectively. The results have shown in Figure 8. Compared with the model with the constant prior this model gives a result with a solid body, in the top-right part of the plot. Variance is close to zero, except at the edges of the solid shale area. The MMAP reflects the expected value, and only shows one connected area where one expect shale.

### Problem 1d)

Comparing the two methods, the first in 1b) with a uniform prior and the Markov RF in 1c), the Markov RF seems to produce less noise and samples from it seems much more stable and shows much less variance. Samples from the Markov RF seems match what is observed in  $D_c$  much better than 1a). Although it takes much more time to implement and run, the Markov RF produces much better results than its competitor.

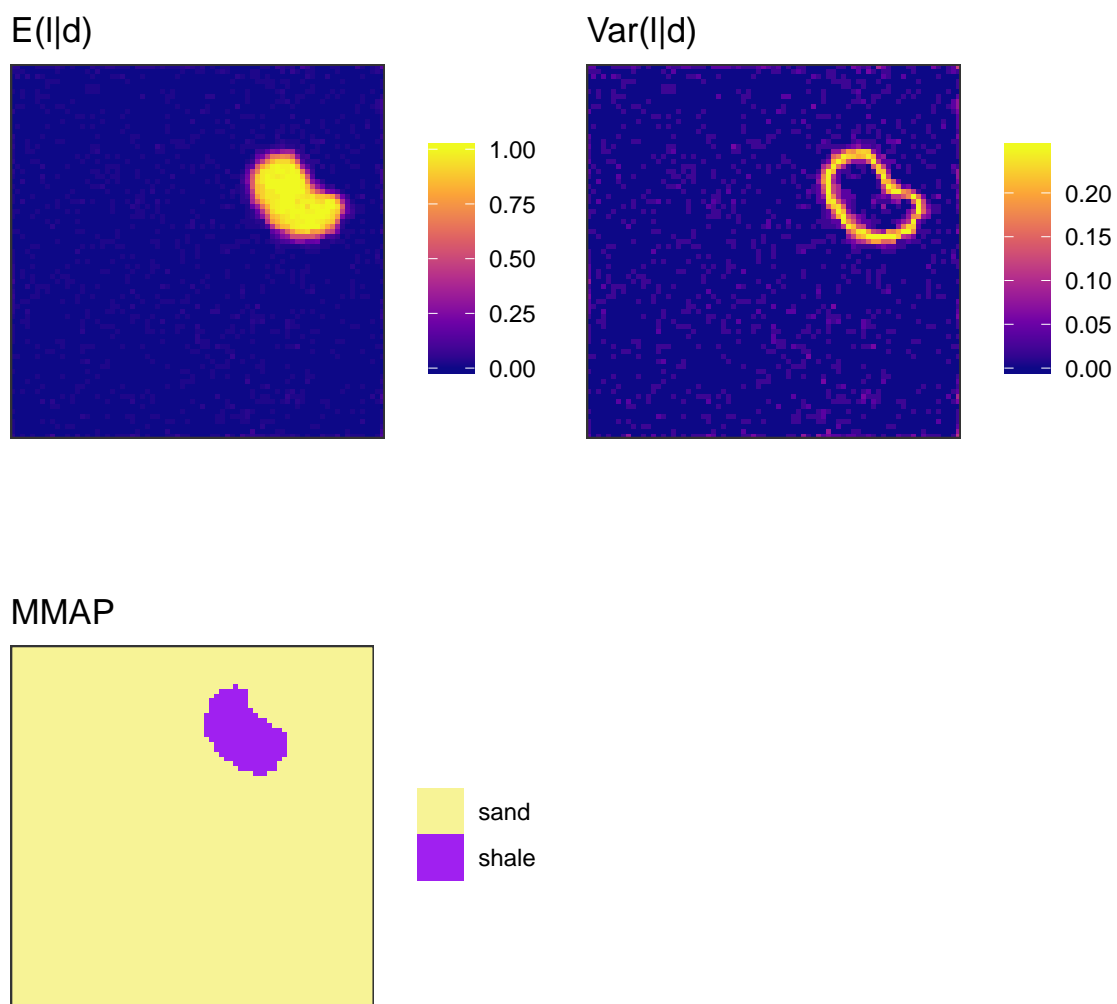


Figure 8: Estimators from independent gibbs sample

On the other hand, 1c) might be a bit too strict on what is classified as a shale-area, compared to  $D_c$  there seems to be few outlier shale fields, and small shale fields. A possibility to improve on this would be to manually alter the  $\beta$  estimate somewhat.

In both cases we also note that the MMAP gave satisfying results.