

Sequential Markov random fields and Markov mesh random fields for modelling of geological structures

Marita Stien

Master of Science in Physics and Mathematics
Submission date: June 2006
Supervisor: Håkon Tjelmeland, MATH

Problem Description

The candidate should study the use of sequential Markov random fields and Markov mesh random fields for modelling of geological structures. In particular she should consider how to parametrize the models and how to do parameter estimation.

Assignment given: 20. January 2006
Supervisor: Håkon Tjelmeland, MATH

Preface

This thesis is part of the Master of Science degree in Industrial Mathematics at the Norwegian University of Science and Technology. The work is confined to a period of 21 weeks (20 weeks plus 1 week as compensation for the Easter holidays), from January 2006 to June 2006.

Working with the thesis has been a very instructive process for me. I have learned a lot about writing reports as well as improved my knowledge in statistics and developed by computer programming skills.

I would like to thank Håkon Tjelmeland for his excellent supervision. He has spent a lot of time guiding me through the theory related to this work, and given me very helpful feedback.

Trondheim, June 16th 2006

Marita Stien

Abstract

We have been given a two-dimensional image of a geological structure. This structure is used to construct a three-dimensional statistical model, to be used as prior knowledge in the analysis of seismic data. We consider two classes of discrete lattice models for which efficient simulation is possible; sequential Markov random field (sMRF) and Markov mesh random field (MMRF). We first explore models from these two classes in two dimensions, using the maximum likelihood estimator (MLE). The results indicate that a larger neighbourhood should be considered for all the models. We also develop a second estimator, which is designed to match the model with the observation with respect to a set of specified functions. This estimator is only considered for the sMRF model, since that model proved to be flexible enough to give satisfying results. Due to time limitation of this thesis, we could not wait for the optimization of the estimator to converge. Thus, we can not evaluate this estimator. Finally, we extract useful information from the two-dimensional models and specify a sMRF model in three dimensions. Parameter estimation for this model needs approximative techniques, since we only have given observations in two dimensions. Such techniques have not been investigated in this report, however, we have adjusted the parameters manually and observed that the model is very flexible and might give very satisfying results.

Contents

1	Introduction	1
2	Discrete lattice models	2
2.1	Random fields and definitions	2
2.2	Markov random fields	3
2.3	Sequential Markov random fields	5
2.3.1	Efficient simulation of Markov random fields in one dimension	8
2.4	Markov mesh random fields	10
3	Parameter estimation	14
3.1	Maximum likelihood estimation	14
3.1.1	Sequential Markov random field	14
3.1.2	Markov mesh random field	15
3.2	Matching specified functions	16
4	Models in two dimensions	18
4.1	Evaluation of the models	19
4.2	SMRF model	21
4.2.1	Maximum likelihood estimation	22
4.2.2	Matching specified functions	25
4.3	MMRF model1	28
4.3.1	Maximum likelihood estimation	29
4.4	MMRF model2	32
4.4.1	Maximum likelihood estimation	33
4.5	Discussion of the results in two dimensions	35
5	Models in three dimensions	36
5.1	Model specification	36
5.2	Maximum likelihood estimation with observations in three dimensions	44
5.3	Discussion of the results in three dimensions	46
6	Closing remarks	46
	References	48

1 Introduction

Analysis of seismic data provides knowledge about the geological subsea structures. However, such data is affected by “noise” making them hard to interpret. To strengthen the accuracy of the interpretation, it might be of help to use prior knowledge of the geological structures. Geologists have found areas, above the sea level, with geological structures they think are similar to the subsea structures. In Figure 1 we have given such a structure in the form of a two-dimensional image. The white fields are shale and the black fields are sandstone. Our aim is to construct a three-dimensional statistical model reproducing similar structures, that in a larger context is to be used as a prior distribution in a Bayesian framework, where we condition on given seismic data.



Figure 1: A two dimensional slice of a geological structure where the white fields are shale and the black fields are sandstone. The resolution of the image is 145×271

Since we are dealing with two-colored images, the class of binary discrete lattice models is an appropriate class to work with. The Markov random field (MRF) models are a subclass of the discrete lattice models and are widely used in this area. However, to simulate and estimate the parameters of a MRF model generally require approximative techniques, and for large images such techniques are very time-consuming, and convergence is not guaranteed. We want to find a model that besides being able to reproduce geological structures with the same features as the one in Figure 1, is efficient both when it comes to simulation and parameter estimation.

In this text we explore two alternative classes of models to the MRF, where approximative techniques are not required for simulation and parameter estimation. The

first one is the sequential Markov random field (sMRF), also known as multidimensional Markov chain model in Qian and Titterton (1991). The second class is called Markov Mesh random field (MMRF) (Abend et al., 1965; Gray et al., 1994). We will first focus on models in two dimensions, and use what we learn from these models when we specify a model in three dimensions. Both sMRF and MMRF have been studied in two dimensions for the same purpose in Eriksen (2005). Also in Stien (2005) a two-dimensional sMRF model has been specified and explored, where the maximum likelihood estimator (MLE) has been used for the parameter estimation. We will further explore the same parametrized model by applying a second method of parameter estimation based on matching a set of specified functions. In addition to investigation of these two classes in two dimensions, we also study a three-dimensional sMRF model.

The models will be evaluated on how well they fit the geological structure in Figure 1, both visually and by comparing the correlation function for the realizations with the empirical correlation.

Regarding the organization of the text, we first establish the notation and theory needed to fully understand the models and how we simulate from them, followed by a review of the methods of parameter estimation. The result part is divided into two parts, one for the models in two dimensions and one for the three-dimensional model. Discussions of the results will be given immediately after each of these parts.

2 Discrete lattice models

The sMRF and the MMRF are subclasses of the MRF, which again is a subclass of random fields. In this section we work our way through the different classes one by one starting with random fields, where we also introduce definitions and notation. For simplicity reasons we choose to work with a general notation, and make it more specific once we specify the models.

2.1 Random fields and definitions

A lattice is a countable collection of sites $i \in \mathcal{I}$ for $i = 1, \dots, n$, where each site i has a variable that takes different values, x_i . We let χ be the set of possible values taken by the x_i 's. These values can either be discrete or continuous, and if they are assumed to be random they will be associated with a probability, $\pi(x_i)$. We use x_i as both a random variable and the realization of the variable, the interpretation will be clear from the context. A lattice of random variables is called a random field, and this field is denoted

by x , where $x = (x_1, \dots, x_n)$. The probability of a given realization of the random field is $\pi(x)$, and is the joint probability of the x_i 's. The conditional probability at site i given the rest is written as $\pi(x_i|x_{-i})$.

In this report we view our data as a finite collection of binary random variables, i.e. $x_i \in \chi = \{0, 1\}$ where 0 and 1 represent black and white, respectively. The numbering i of the sites determines the ordering of the variables. This ordering is essential for certain classes of models, where it also determines the order of simulation.

To efficiently simulate a random field, a sequential sampling scheme must be obtained. We can always factorize a random field x according to the multiplication rule which states that any proper probability distribution can be factorized into the product of the conditionals, i.e.

$$\begin{aligned} \pi(x) &= \pi(x_n, x_{n-1}, x_{n-2}, \dots, x_2, x_1) = \pi(x_n|x_{n-1}, x_{n-2}, \dots, x_2, x_1) \\ &\quad \times \pi(x_{n-1}|x_{n-2}, \dots, x_2, x_1) \times \dots \times \pi(x_2|x_1) \times \pi(x_1). \end{aligned} \quad (1)$$

If we have expressions for each of the conditionals in (1), the field x can be sampled sequentially by first sampling x_1 , then $x_2|x_1$, and continue sampling the conditional distributions all the way to $x_n|x_{n-1}, x_{n-2}, \dots, x_2, x_1$, resulting in a sample from $\pi(x)$.

2.2 Markov random fields

For a thorough introduction to the MRFs we refer to Hurn et al. (2003). A MRF is a random field in which the spatial dependency between the sites is local, i.e. the field has a Markov property. The scope of this dependency is determined by specifying a neighbourhood. By neighbourhood we usually mean a set of sites close to each other. If site i is a neighbour of site j , we write $i \sim j$. The neighbour relation for a MRF is symmetric, i.e. if i is a neighbour of j then j has to be a neighbour of i . Figure 2 shows two examples of neighbourhoods for a two-dimensional MRF on a regular lattice, where the gray square is the site for which the white squares are neighbours. In (a) only the closest neighbours are included, and we have a first order neighbourhood. In (b) we have a second order neighbourhood which also includes the sites that are second closest. The Markov property allows us to write the conditional probability at site i given the rest as

$$\pi(x_i|x_{-i}) = \pi(x_i|x_{\delta_i}) \quad (2)$$

where δ_i is the set of neighbours to site i , i.e. $\delta_i = \{j : j \sim i\}$.

A clique is a set of variables which either consists of a single site, or where every pair

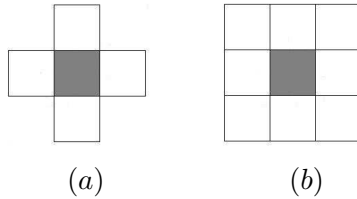


Figure 2: Examples of neighbourhoods for a two-dimensional MRF on a regular lattice. The gray square is the site for which the white squares are neighbours. (a) first order neighbourhood. (b) second order neighbourhood.

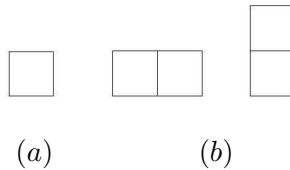


Figure 3: The set of possible cliques for the first order neighbourhood in Figure 2(a). (a) c_1 (b) c_2 .

of sites in the set are neighbours. We let C denote the set of all cliques, and c_k is a clique in C containing k sites. For the first order neighbourhood in Figure 2(a) the clique can be of size 1 or 2, where c_2 contains the largest possible number of sites, see Figure 3 for a visualization of the cliques. Each clique configuration in a field have a potential given by a potential function. A potential function, $V(\cdot)$, is a function that takes arguments associated with the sites in a clique.

Given the Markov assumption in (2) and a positivity condition, $\pi(x) > 0$ for all possible x , it follows from the Hammersley-Clifford theorem (Hurn et al., 2003) that we have a unique joint distribution $\pi(x)$ if it is on the form

$$\pi(x) = K \times \exp \left\{ - \sum_{c \in C} V(x_c) \right\}, \quad (3)$$

where x_c are the elements in x that belong to clique c and K is a normalizing constant,

$$\frac{1}{K} = \sum_x \exp \left\{ - \sum_{c \in C} V(x_c) \right\} < \infty. \quad (4)$$

As an example we consider a one-dimensional MRF. The neighbourhood is chosen such that the largest clique is of size l , c_l , and without loss of generality we set the potential

for the rest of the possible cliques equal to zero. Then the joint probability is

$$\pi(x) = K \times \exp \left\{ - \sum_{i=1}^{n-l+1} V(x_i, x_{i+1}, \dots, x_{i+l-1}) \right\}. \quad (5)$$

Markov chain Monte Carlo (MCMC) methods are generally required for simulation of MRF models in more than one dimension. As the name indicates, MCMC methods involve a construction of a Markov chain that has the desired probability distribution as its limiting probability distribution. However, there are many approaches for specifying a MRF. As long as the joint probability fulfills the requirements of the Hammersley-Clifford, there are lots of possibilities. For instance, we can specify the field in terms of the conditionals, $\pi(x_i|x_{i-1}, \dots, x_1)$, and directly obtain a sequential sampling scheme. The MMRFs have this property.

2.3 Sequential Markov random fields

A similar deduction of this class is given in Qian and Titterton (1991). The sMRF consists of MRFs in one dimension, which are linked together as a first or higher order Markov chain (MC). With the property of a MC, we can specify the sMRF in terms of the conditional probabilities for the one-dimensional MRFs given a subset of their preceding MRFs. The number of fields in the subset determines the order of the chain. Note that we do not specify the conditional probabilities at each site, however, we are still able to sample the sites sequentially by using a recursive algorithm which will be introduced later.

To determine the spatial dependency between the MRFs, we must first determine their ordering. We let the r one-dimensional MRFs in the sMRF be denoted by ϕ_i for $i = 1, 2, \dots, r$, where each ϕ_i contains n_i sites, giving a total number of $\sum_{i=1}^r n_i = n$ sites in the field. An arbitrary MRF, ϕ_i , in the MC depends only on a subset of the MRFs with a lower ordering number, i.e. a subset of $\{\phi_j | j < i\}$. For the sMRF we have the following ordering

$$\begin{aligned} \phi_1 &= (x_1, x_2, \dots, x_{n_1}) \\ \phi_2 &= (x_{n_1+1}, x_{n_1+2}, \dots, x_{n_1+n_2}) \\ &\vdots \\ \phi_r &= (x_{n_1+\dots+n_{r-1}+1}, \dots, x_{n_1+\dots+n_r}). \end{aligned} \quad (6)$$

See Figure 4 for an illustration of the sMRF on a two-dimensional regular lattice, where

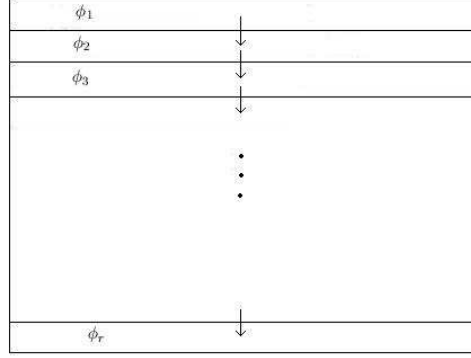


Figure 4: Illustration of a sMRF on a two-dimensional regular lattice, where the MRFs are linked together by a first order MC. The arrows visualize the order of simulation.

the MRFs are linked together as a first order MC. The arrows visualize the order of simulation, the MRFs must be simulated according to their ascending ordering. If the arrows in the illustration were pointing in both directions the whole field would be a MRF.

We factorize the field x with respect to the MRFs

$$\begin{aligned} \pi(x) &= \pi(\phi_r, \phi_{r-1}, \phi_{r-2}, \dots, \phi_1) = \pi(\phi_r | \phi_{r-1}, \phi_{r-2}, \dots, \phi_1) \\ &\quad \times \pi(\phi_{r-1} | \phi_{r-2}, \phi_{r-3}, \dots, \phi_1) \times \dots \times \pi(\phi_2 | \phi_1) \times \pi(\phi_1). \end{aligned} \quad (7)$$

Each ϕ_i is only dependent of a few of the preceding fields, and we denote the set of these fields by ϕ_{D_i} for all i 's. Thus, (7) simplifies to

$$\pi(x) = \pi(\phi_r | \phi_{D_r}) \times \pi(\phi_{r-1} | \phi_{D_{r-1}}) \times \dots \times \pi(\phi_E), \quad (8)$$

where ϕ_E is the set of ϕ_i 's that do not have the same set of dependent preceding fields, ϕ_{D_i} , as the rest. The conditional probability $\pi(\phi_i | \phi_{D_i})$ will have the distribution of a one-dimensional MRF, and by using the same assumptions as for the expression in (5) we get

$$\begin{aligned} \pi(\phi_i | \phi_{D_i}) &= K(\phi_{D_i}) \\ &\quad \times \exp \left\{ - \sum_{j=1}^{n_i-l+1} V(x_{N_i+j}, x_{N_i+j+1}, \dots, x_{N_i+j+l-1} | \phi_{D_i}) \right\}, \end{aligned} \quad (9)$$

where $K(\phi_{D_i})$ is the normalizing constant for field ϕ_i given ϕ_{D_i} , and $N_i = \sum_{k=1}^{i-1} n_k$ ensures a correct numbering of the sites. By combining (8) and (9) the probability distribution for the whole field x becomes

$$\begin{aligned} \pi(x) &= \pi(\phi_E) \prod_{i \notin E} \pi(\phi_i | \phi_{D_i}) = \pi(\phi_E) \\ &\times \prod_{i \notin E} K(\phi_{D_i}) \exp \left\{ - \sum_{j=1}^{n_i-l+1} V(x_{N_i+j}, x_{N_i+j+1}, \dots, x_{N_i+j+l-1} | \phi_{D_i}) \right\}, \end{aligned} \quad (10)$$

where $\pi(\phi_E)$ is the probability for the MRFs in ϕ_E .

Only a limited number of sites from the preceding fields ϕ_{D_i} influence the value at site x_{N_i+j} . We denote the set of these sites together with the sites in the clique by $x_{S_{N_i+j}}$, where the number of sites in the set is n_S . We can then write (10) as

$$\pi(x) = \pi(\phi_E) \prod_{i \notin E} K(\phi_{D_i}) \exp \left\{ - \sum_{j=1}^{n_i-l+1} V(x_{S_{N_i+j}}) \right\}. \quad (11)$$

We will specify the potential function such that we consider all possible interactions between its arguments, rather than restricting ourselves to lower order interactions. Since we are considering binary data the number of configurations of the sites in $x_{S_{N_i+j}}$ is 2^{n_S} . We define a function $\gamma(\cdot)$, which takes arguments associated with the sites in $x_{S_{N_i+j}}$. This function returns a positive integer t for $t = 1, 2, \dots, s$, where $s \leq 2^{n_S}$. In relation to the $\gamma(\cdot)$ -function we introduce a set of s parameters,

$$\theta_{\gamma(x_{S_{N_i+j}})} = V(x_{S_{N_i+j}}). \quad (12)$$

We let $I(\gamma(x_{S_{N_i+j}}) = t)$ be an indicator function,

$$I(\gamma(x_{S_{N_i+j}}) = t) = \begin{cases} 1 & \text{if } \gamma(x_{S_{N_i+j}}) = t, \\ 0 & \text{otherwise,} \end{cases} \quad (13)$$

and replace the potential function with the set of parameters in (12). This yields

$$\pi(x) = \pi(\phi_E) \prod_{i \notin E} K(\phi_{D_i}) \exp \left\{ - \sum_{j=1}^{n_i-l+1} \sum_{t=1}^s I(\gamma(x_{S_{N_i+j}}) = t) \theta_t \right\}. \quad (14)$$

This can be further simplified by defining n_t^i , which denotes the number of times the

value of the $\gamma(\cdot)$ -function is t in ϕ_i , i.e.

$$n_t^i = \sum_{j=1}^{n_i-l+1} I(\gamma(x_{S_{N_i+j}}) = t). \quad (15)$$

We can now remove one of the summations in the exponential factor and get

$$\pi(x) = \pi(\phi_E) \prod_{i \notin E} K(\phi_{D_i}) \exp \left\{ - \sum_{t=1}^s n_t^i \theta_t \right\}. \quad (16)$$

Recall that the sites included in the potential function are sites forming the clique and sites from the conditioned MRFs. When the configuration of the conditioned variables is known, we assign a potential to each clique configuration. The potential is identical to the corresponding parameter, θ_t . The relative relation between the different potentials can be linked to the probabilities of occurrence for the cliques. For instance, if a specific clique configuration has a high potential compared to the other clique configurations, this implies a low probability of occurrence, and vice versa. However, in some situations this interpretation must be modified. An extreme case can be where a configuration with a relatively low potential occurs only as a consequence of the occurrence of a configuration with a relatively high potential. Then the low potential configuration will have a small probability of occurrence, since it is shaded by a configuration that has a small probability of occurrence.

Finally, we will discuss how to simulate the sMRFs. First of all, we let the sites in E be independent and identically distributed with equal probability for both black and white. Then the expression in (16) becomes

$$\pi(x) = \left(\frac{1}{2} \right)^{n_E} \prod_{i \notin E} K(\phi_{D_i}) \exp \left\{ - \sum_{t=1}^s n_t^i \theta_t \right\}, \quad (17)$$

where n_E is the number of sites in E . We first sample the sites in E , and then sequentially sample the rest of the ϕ_i 's according to their ascending ordering number. For the sites within each of the ϕ_i 's we must, as mentioned earlier, use a recursive algorithm to obtain a sequential sampling scheme, and this is presented in the following section.

2.3.1 Efficient simulation of Markov random fields in one dimension

For MRFs in more than one dimension it requires a high amount of CPU time to calculate each of the conditionals in the multiplication rule in (1). However, for one-dimensional

MRFs there exists a so called forward-backward algorithm which takes advantage of the Markov property, and recursively finds exact expressions for the conditional probabilities. We choose to demonstrate the forward-backward algorithm with a general notation to avoid using the extensive labelling introduced for the sMRF. Thus, each ϕ_i is denoted by z where $z = (z_1, z_2, \dots, z_p)$, and from (9) we get

$$\pi(\phi_i|\phi_{D_i}) = \pi(z|\phi_{D_i}) = K(\phi_{D_i}) \exp \left\{ - \sum_{j=1}^{p-l+1} V(z_j, z_{j+1}, \dots, z_{j+l-1}|\phi_{D_i}) \right\}. \quad (18)$$

The forward-backward algorithm works as follows. We use the summation formula

$$\sum_{z_p, z_{p-1}, \dots, z_{q+1}} \pi(z|\phi_{D_i}) = \pi(z_q, z_{q-1}, \dots, z_1|\phi_{D_i}), \quad (19)$$

which states that if you sum over all the possible outcomes of the variables $z_p, z_{p-1}, \dots, z_{q+1}$ you will be left with the (conditional) joint probability of z_q, z_{q-1}, \dots, z_1 . This summation is valid for $0 < q < p - 1$, which makes it possible to find the (conditional) marginal distribution for the first variable z_1 . During the summation process in (19) we gather numerical information in recursive $g(\cdot)$ -functions, defined by

$$g_{q-l+1, \dots, q-1}(z_{q-l+1}, \dots, z_{q-1}) = \sum_{z_q} g_{q-l+2, \dots, q}(z_{q-l+2}, \dots, z_q) \exp \{-V(z_{q-l+1}, \dots, z_q|\phi_{D_i})\}. \quad (20)$$

The connection between the $g(\cdot)$ -functions and the probabilities is

$$\begin{aligned} \pi(z_1, \dots, z_q|\phi_{D_i}) &= g_{q-l+2, \dots, q}(z_{q-l+2}, \dots, z_q) \\ &\times K(\phi_{D_i}) \exp \left\{ - \sum_{j=1}^{q-l+1} V(z_j, z_{j+1}, \dots, z_{j+l-1}|\phi_{D_i}) \right\}, \end{aligned} \quad (21)$$

where the exponential term vanishes when q is less than the clique size l . For the first site, $q = 1$, we get

$$\pi(z_1|\phi_{D_i}) = K(\phi_{D_i}) \sum_{z_2} g_{1,2}(z_1, z_2|\phi_{D_i}) = K(\phi_{D_i}) g_1(z_1). \quad (22)$$

To be a valid distribution the marginal probability for z_1 has to sum to 1 when summing

over the possible outcomes of z_1 , i.e.

$$\sum_{z_1} \pi(z_1 | \phi_{D_i}) = K(\phi_{D_i}) \sum_{z_1} g_1(z_1) = 1, \quad (23)$$

and with this relation it is possible to compute the normalizing constant $K(\phi_{D_i})$,

$$K(\phi_{D_i}) = \left[\sum_{z_1} g_1(z_1) \right]^{-1}. \quad (24)$$

The conditional probabilities $\pi(z_q | z_{q-1}, \dots, z_1, \phi_{D_i})$ are given by the relation

$$\pi(z_q | z_{q-1}, \dots, z_1, \phi_{D_i}) = \frac{\pi(z_q, \dots, z_1 | \phi_{D_i})}{\pi(z_{q-1}, \dots, z_1 | \phi_{D_i})}, \quad (25)$$

where we after the summation process have expressions for both the nominator and denominator in the fraction. Thus, we have obtained complete expressions for all the conditionals in the factorization in (1), and the field can be sampled in one scan.

The forward-backward algorithm is efficient since, due to the Markov property, the $g(\cdot)$ -functions only take a small number of arguments. The number of these arguments corresponds to the order of the neighbourhood in the MRF. For MRFs in higher dimensions than one, the $g(\cdot)$ -functions can no longer take advantage of the Markov property and the number of arguments is considerably larger. For a more detailed discussion on the forward-backward algorithm, we refer to the project work in Stien (2005) where slightly different and more specific notations have been used.

2.4 Markov mesh random fields

In the literature we refer to Abend et al. (1965) and Gray et al. (1994). The MMRF is a generalization of a MC to two or higher dimensions. We specify the field in terms of the conditional probabilities, $\pi(x_i | x_{i-1}, \dots, x_1)$, for each site, and directly obtain a sequential sampling scheme.

The ordering of the sites in a MMRF is crucial, and the sites must be sampled according to the ascending ordering. For instance, for a two-dimensional MMRF on a regular grid where we wish to sample from left to right and from top to bottom, the

field must have the underlying ordering

$$\begin{bmatrix} 1 & 2 & \dots & k \\ k+1 & k+2 & \dots & 2k \\ \vdots & \vdots & \ddots & \vdots \\ (l-1)k+1 & (l-1)k+2 & \dots & n \end{bmatrix}, \quad (26)$$

where k and l give the horizontal and vertical extension of the field, respectively. Another possible ordering of the sites can be if we wish to simulate from left to right and from right to left for every other row, and from top to bottom. For the whole field this yields

$$\begin{bmatrix} 1 & 2 & \dots & k \\ 2k & 2k-1 & \dots & k+1 \\ \vdots & \vdots & \ddots & \vdots \\ (l-2)k+1 & (l-2)k+2 & \dots & (l-1)k \\ n & n-1 & \dots & (l-1)k+1 \end{bmatrix}. \quad (27)$$

The predecessors of a site are the set of sites in x with an ordering number less than the ordering number of the site they proceed. We let η_i be the set of predecessors for site i , then this set is given by $\eta_i = \{1, 2, \dots, i-1\}$. The predecessors for an arbitrary site, with the ordering in (26), are illustrated in Figure 5. The black sites are the predecessors of the gray site. This illustration is also valid for the ordering in (27), by ensuring an odd row number. For a MMRF, the factorization of the field x can be expressed by

$$\pi(x) = \pi(x_n | x_{\eta_n}) \times \pi(x_{n-1} | x_{\eta_{n-1}}) \times \dots \times \pi(x_2 | x_{\eta_2}) \times \pi(x_1). \quad (28)$$

A site is only affected by a limited number of sites in its set of predecessors. These sites constitute the sequential neighbourhood. We denote the set of sequential neighbours to site x_i by $x_{p\delta_i}$, where the number of sites in $x_{p\delta_i}$ is $n_{p\delta}$, with $2^{n_{p\delta}}$ possible configurations. Those sites that do not have the same sequential neighbourhood relation as the rest are denoted by x_E , and the number of these sites depends on the choice of the sequential neighbourhood. Figure 6 shows three examples of sequential neighbourhoods on a regular two-dimensional lattice, where the white squares are the sequential neighbours of the gray square. In (a) and (b) there are sequential neighbourhoods valid for both orderings (26) and (27), and in (c) the sequential neighbourhood is only valid for the ordering in (27) for even row numbers.



Figure 5: Illustration of the predecessor of an arbitrary site for the ordering in (26). The black sites are the predecessors of the gray site. Notice that for odd row numbers, this illustration is also valid for the ordering in (27).

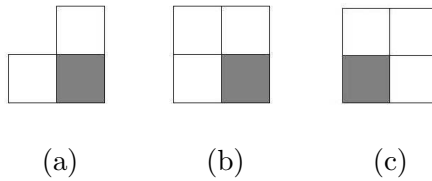


Figure 6: Examples of sequential neighbourhoods for the Markov mesh random fields. (a) $n_{p\delta} = 2$ (b) $n_{p\delta} = 3$ (c) $n_{p\delta} = 3$. The neighbourhood in (c) is not valid for the ordering in (26).

The conditional probabilities $\pi(x_i|x_{\eta_i})$ can now be written

$$\pi(x_i|x_{\eta_i}) = \pi(x_i|x_{p\delta_i}). \quad (29)$$

We combine (28) and (29) and obtain the joint probability

$$\pi(x) = \pi(x_E) \prod_{i \notin E} \pi(x_i|x_{p\delta_i}), \quad (30)$$

where $\pi(x_E)$ is the probability for the sites in E . To get a complete expression for the joint probability, it remains to specify the conditional probabilities $\pi(x_i|x_{p\delta_i})$ and the probability $\pi(x_E)$.

We will specify the models such that we do not consider lower order interactions between the sites in the sequential neighbourhood. Similarly as we did for the sMRF, we define a function $\gamma(\cdot)$, which takes arguments associated with the sites in the sequential neighbourhood, $\gamma(x_{p\delta_i})$. This function returns a positive integer t for $t = 1, \dots, s$, where $s \leq 2^{n_{p\delta}}$. In addition, we introduce a set of parameters such that the conditional probability $\pi(x_i = 0|x_{p\delta_i}) = \theta_{\gamma(x_{p\delta_i})}$ is the probability of site i being black given the

values of the sequential neighbourhood. Since there are only two possible outcomes, the conditional probability of the site being white is simply $\pi(x_i = 1|x_{p\delta_i}) = 1 - \theta_{\gamma(x_{p\delta_i})}$. This yields the joint expression

$$\pi(x_i|x_{p\delta_i}) = \theta_{\gamma(x_{p\delta_i})}^{1-x_i} (1 - \theta_{\gamma(x_{p\delta_i})})^{x_i}. \quad (31)$$

So for our MMRF model the parameters represent probabilities, making their interpretation very simple.

We use the indicator function from (13) and replace the form of the conditional probabilities in (31) with the ones in (30), and get

$$\pi(x) = \pi(x_E) \prod_{i \notin E} \sum_{t=1}^l I(\gamma(x_{p\delta_i}) = t) \theta_t^{1-x_i} (1 - \theta_t)^{x_i}. \quad (32)$$

We let n_t^0 denote the number of times $\gamma(x_{p\delta_i})$ equals t and the value at i is zero when going through the field, and the same for n_t^1 where the value at i is one, i.e

$$n_t^0 = \sum_{i \notin E} I(\gamma(x_{p\delta_i}) = t) (1 - x_i) \quad (33)$$

$$n_t^1 = \sum_{i \notin E} I(\gamma(x_{p\delta_i}) = t) x_i. \quad (34)$$

Thus, rather than taking the product of the conditionals for each site, we collect the conditional factors with the same parameter, and take the product over the different parameters

$$\pi(x) = \pi(x_E) \prod_{t=1}^s \theta_t^{n_t^0} (1 - \theta_t)^{n_t^1}. \quad (35)$$

Finally, when the $\gamma(\cdot)$ -function is specified we can start sampling. This is done by first sampling the sites in E . Then we sample sequentially site by site according to the ascending ordering number. As for the sMRF, we let the n_E sites in E be independent and identically distributed with probability $1/2$, and get

$$\pi(x) = \left(\frac{1}{2}\right)^{n_E} \prod_{t=1}^s \theta_t^{n_t^0} (1 - \theta_t)^{n_t^1}. \quad (36)$$

3 Parameter estimation

In this section we introduce two methods of parameter estimation. The first one is the MLE and the second is a method designed to minimize the square distance between the model and the observation with respect to a set of specified functions.

3.1 Maximum likelihood estimation

The likelihood function has the same form as the probability function except that the x values are known and the parameters are unknown, i.e.

$$L(\theta|x) = \pi(x|\theta). \quad (37)$$

Maximum likelihood parameter estimation is a method that determines the parameters which maximize the likelihood function,

$$\hat{\theta} = \arg \max_{\theta} L(\theta|x). \quad (38)$$

3.1.1 Sequential Markov random field

From the probability distribution in (14) we get a likelihood on the form

$$L(\theta|x) = \pi(x|\theta) = \pi(\phi_E|\theta) \prod_{i \notin E} K(\phi_{D_i}) \exp \left\{ - \sum_{t=1}^s n_t^i \theta_t \right\}. \quad (39)$$

If $\pi(\phi_E|\theta)$ is unspecified, we can not compute the exact likelihood. This results in a slightly less informative likelihood function, the conditional likelihood

$$\begin{aligned} L_c(\theta|x) &= \frac{\pi(x|\theta)}{\pi(\phi_E|\theta)} = \pi(\phi_{-E}|\phi_E, \theta) \\ &= \prod_{i \notin E} K(\phi_{D_i}) \exp \left\{ - \sum_{t=1}^s n_t^i \theta_t \right\}, \end{aligned} \quad (40)$$

where ϕ_{-E} are those sites that are not in ϕ_E .

To compute the likelihood we need the normalizing constant, and this can easily be computed by using the forward-backward algorithm. However, numerical problems may arise in the computations. We often get some very high and very low values which result in unstable computations. This problem is solved by working with the logarithm of the likelihood instead of the likelihood. The parameter values that optimize the likelihood

function will also optimize the logarithm of the likelihood function.

It is not possible to optimize the parameters analytically, so we have to use iterative procedures. For that purpose we use a Non-Linear Minimization algorithm in R. This function carries out a minimization of the negative logarithm of the likelihood using a Newton-type algorithm (R Development Core Team, 2005).

3.1.2 Markov mesh random field

For the MMRF we have the probability distribution from (36), giving a likelihood function on the form

$$L(\theta|x) = \pi(x|\theta) = \pi(x_E|\theta) \prod_{t=1}^s \theta_t^{n_t^0} (1 - \theta_t)^{n_t^1}. \quad (41)$$

With the same reasoning as for the sMRF we must work with the conditional likelihood function

$$L_c(\theta|x) = \prod_{t=1}^s \theta_t^{n_t^0} (1 - \theta_t)^{n_t^1}. \quad (42)$$

By partial differentiation we can obtain the estimates in (38). Differentiating the logarithm of the likelihood function is much easier than differentiating the likelihood function itself, and the estimates are the same. The logarithm of the likelihood is

$$\log L_c(\theta|x) = \sum_{t=1}^s (n_t^0 \log \theta_t + n_t^1 \log(1 - \theta_t)). \quad (43)$$

All the partial derivatives of $\log L_c(\theta|x)$ with respect to θ_t is on the same form, namely

$$\frac{\partial}{\partial \theta_t} \log L_c(\theta|x) = \frac{n_t^0}{\theta_t} - \frac{n_t^1}{1 - \theta_t}, \quad \text{for } t = 1, 2, \dots, s. \quad (44)$$

When the partial derivatives are all equal to zero, the likelihood function is either at its maximum or minimum. Taking the double derivative proves that we have a maximum point. So when we solve the following set of equations

$$\frac{n_t^0}{\theta_t} - \frac{n_t^1}{1 - \theta_t} = 0, \quad \text{for } t = 1, 2, \dots, s, \quad (45)$$

we obtain the MLE for θ , and these are

$$\hat{\theta}_t = \frac{n_t^0}{n_t^0 + n_t^1}, \quad \text{for } t = 1, 2, \dots, s. \quad (46)$$

3.2 Matching specified functions

This estimation procedure is based on matching the model with the observation using a set of optional specified functions. That is, we minimize the square distance, $g(\theta)$, between the expected value for the model $\pi(x|\theta)$ and the empirical value, with respect to these functions. We define a set with p functions, $h_1(\cdot), h_2(\cdot), \dots, h_p(\cdot)$, and get a distance function given by

$$g(\theta) = \sum_{k=1}^p (E_{\theta}[h_k(x)] - \hat{\mu}_{h_k})^2, \quad (47)$$

where $\hat{\mu}_{h_k}$ are the empirical function values, and x is simulated from the model parametrized by θ . The parameter estimate is the argument minimizing $g(\theta)$

$$\hat{\theta} = \arg \min_{\theta} g(\theta) = \arg \min_{\theta} \sum_{k=1}^p (E_{\theta}[h_k(x)] - \hat{\mu}_{h_k})^2. \quad (48)$$

The expectation values in (48) can not be computed analytically for either of our models and must be estimated. For this purpose we use the Monte Carlo approach, which estimates the expectation by taking averages over a large number of samples from the model. Our fields are efficient to sample from, making the Monte Carlo method an easy task. We get the following estimate for the expectation

$$E_{\theta}[\widehat{h_k(x)}] = \frac{1}{N} \sum_{i=1}^N h_k(x_i), \quad (49)$$

where the realizations, x_i for $i = 1, 2, \dots, N$, are sampled from the model $\pi(x|\theta)$. The number of realization is N , and the higher this number is the more accurate the estimate will be. The estimate for the distance function, $\hat{g}(\theta)$, becomes

$$\hat{g}(\theta) = \arg \min_{\theta} \sum_{k=1}^p \left(E[\widehat{h_k(x)}|\theta] - \hat{\mu}_{h_k} \right)^2 \quad (50)$$

In order to minimize the estimate $\hat{g}(\theta)$, we must use a minimization algorithm. Our minimization algorithm will be based on comparing pairs of $\hat{g}(\theta)$ -functions, each with a separate parameter value. For instance, we wish to test if the parameter value θ gives a better fit than θ_0 , and measure the difference $\hat{d} = \hat{g}(\theta_0) - \hat{g}(\theta)$. If \hat{d} is positive, θ is better. However, as for every stochastic variable, there is variance present in the calculations,

and the variance of \hat{d} is

$$\text{Var}[\hat{d}] = \text{Var}[\hat{g}(\theta_0)] + \text{Var}[\hat{g}(\theta)] - 2\text{Cov}[\hat{g}(\theta_0), \hat{g}(\theta)], \quad (51)$$

where the covariance between the two $\hat{g}(\cdot)$ -functions is zero since they are computed from independent realizations of the model. If the variance of the variable \hat{d} is large, the optimization may prove to be difficult. We wish to reduce the variance, and use a variance reduction technique called importance sampling. In importance sampling we construct a case where the covariance between the two estimates is positive. We want to calculate

$$E_{\theta}[\widehat{h_k(x)}] = \frac{1}{N} \sum_{i=1}^N h_k(x_i), \quad (52)$$

which can be re-written

$$E_{\theta}[\widehat{h_k(x)}] = \frac{1}{N} \sum_{i=1}^N h_k(x_i) \frac{\pi(x_i|\theta)}{\pi(x_i|\theta_0)} = E_{\theta_0}[\widehat{\psi_k(x_i)}], \quad (53)$$

where $\psi_k(x_i) = h_k(x_i) \frac{\pi(x_i|\theta)}{\pi(x_i|\theta_0)}$. Hence, we sample from the same distribution, $\pi(x|\theta_0)$ for both the estimates $\hat{g}(\theta_0)$ and $\hat{g}(\theta)$. They are no longer independent, resulting in a positive covariance. With a positive covariance between the two $\hat{g}(\cdot)$ -functions, the total variance of \hat{d} is reduced. In Tjelmeland (1996) an equivalent procedure was carried out to estimate the normalizing constant in a MRF, in order to make MLE possible.

With this new formulation, we also need new estimates. We let $\tilde{g}(\theta)$ denote the estimate for the distance function, and similarly we let $\tilde{\theta}$ be the parameter estimate minimizing $\tilde{g}(\theta)$, i.e.

$$\tilde{\theta} = \arg \min_{\theta} \tilde{g}(\theta) = \arg \min_{\theta} \sum_{k=1}^p \left(\frac{1}{N} \sum_{i=1}^N h_k(x_i) \frac{\pi(x_i|\theta)}{\pi(x_i|\theta_0)} - \hat{\mu}_{h_k} \right)^2. \quad (54)$$

The fraction $\frac{\pi(x_i|\theta)}{\pi(x_i|\theta_0)}$ are weights, w_i , with values in the range $(0, \infty)$. These weights determine the influence each sample have on the estimate. If the two models $\pi(x|\theta_0)$ and $\pi(x|\theta)$ are approximately equal all the weights will have a value around 1, and we get a good estimate. However, when the models have most of their mass located in different areas the majority of the weights will equal zero and the estimate will be less accurate. Thus, in order for the models to be sufficiently close, the distance between the parameters, θ and θ_0 , must be chosen carefully.

Our routine for minimizing $\tilde{g}(\theta)$ goes as follows. We start with a set of initial pa-

parameter values, $\theta_0 = (\theta_{01}, \dots, \theta_{0s})$, and both increase and decrease each of the elements in θ_0 one at a time by a small value ϵ . For each of the s steps we get three sets of parameters. For instance, step t yields the parameters θ_0 , $\theta_1 = (\theta_{01}, \dots, \theta_{0t} + \epsilon, \dots, \theta_{0s})$ and $\theta_3 = (\theta_{01}, \dots, \theta_{0t} - \epsilon, \dots, \theta_{0s})$, with corresponding function estimates $\tilde{g}(\theta_0)$, $\tilde{g}(\theta_1)$ and $\tilde{g}(\theta_3)$. The θ value giving the smallest value of the $\tilde{g}(\cdot)$ -function is chosen as the new θ_0 . This procedure is repeated until convergence. The algorithm is more formally written down in Algorithm 1.

Algorithm 1 Pseudo code for the optimization routine minimizing $\tilde{g}(\theta)$.

```

Initialize  $\theta_0$ 
while not at minima do
  for  $j=1, \dots, s$  do
    Set  $\theta_1 = \theta_2 = \theta_0$ 
    Set  $\theta_1(j) = \theta_1(j) + \epsilon$  and  $\theta_2(j) = \theta_2(j) - \epsilon$ 
    Calculate the three estimates  $\tilde{g}(\theta_0)$ ,  $\tilde{g}(\theta_1)$  and  $\tilde{g}(\theta_2)$ 
    Find the smallest estimate and set  $\theta_0$  equal to the corresponding parameter
  end for
end while

```

We primarily focus on finding an appropriate estimator for our models. Therefore, we have constructed a simple minimization routine and disregarded the fact that it is very inefficient. To increase efficiency it would be a good idea to approximate the gradient of $\tilde{g}(\theta)$, which always finds the direction in which the function decreases the most.

It now remains to find appropriate $h(\cdot)$ -functions that give good parameter estimates. The MLE is a special case of $g(\theta)$, where the $h(\cdot)$ -functions are equivalent to the parametrization of the model. In Eriksen (2005) the covariance function was used as an $h(\cdot)$ -function for estimation of the parameters in a slightly differently specified sMRF model. However, there was a problem with the convergence. We will specify a set of our own functions, and these will be introduced later on when this estimator is applied to one of our models.

4 Models in two dimensions

In this section we will study one model from the sMRF class and two models from the MMRF class. This involves model specification, estimation of the parameters, and evaluation of the simulation results. The work of specifying a model from the sMRF class has previously been done in Stien (2005). Here we give a short review of the modelling process and the results. For the MMRF we look at two slightly differently specified

models, denoted by MMRF model1 and MMRF model2. In Eriksen (2005) a similar study of the MMRF is performed, and we take these results into consideration in our study. The MLE is used for estimation of the parameters for all three models, and in addition we also specify a set of functions for the second estimator, to match the sMRF model with the observation.

First, we change to a more specific and convenient notation. Since an image is being analyzed, the sites in the lattice correspond to pixels in the image. The image is two-dimensional with m pixels in the vertical direction and n pixels in the horizontal direction, so it becomes natural to express the lattice by an $m \times n$ -matrix. The sites in the lattice are the row and column indices in the matrix, $\mathcal{I} = \{(i, j) | i = 1, 2, \dots, m \text{ and } j = 1, 2, \dots, n\}$. We let x_{ij} be the variable in row i and column j of the matrix, $x_{i\star} = (x_{i1}, \dots, x_{in})$ are the variables in row i and $x_{\star j} = (x_{1j}, \dots, x_{mj})$ are the variables in column j .

4.1 Evaluation of the models

For evaluation of our models we compare the correlation function for the simulated images with the empirical correlation function for the observation in Figure 1. The correlation function is given by

$$\begin{aligned} \rho(x_{ij}, x_{i+k, j+l}) &= \frac{\text{Cov}(x_{ij}, x_{i+k, j+l})}{\sqrt{\text{Var}(x_{ij})\text{Var}(x_{i+k, j+l})}} \\ &= \frac{E[(x_{ij} - E[x_{ij}])(x_{i+k, j+l} - E[x_{i+k, j+l}])]}{\sqrt{(E[x_{ij}^2] - E[x_{ij}]^2)(E[x_{i+k, j+l}^2] - E[x_{i+k, j+l}]^2)}}, \end{aligned} \quad (55)$$

where the integers k and l are lags in the vertical and horizontal direction, respectively. The correlation function can not be calculated analytically, and must be estimated.

We assume that our fields are weakly stationary. Firstly, this means that the expectation values are constant and do not depend on i and j , i.e.

$$E[x_{ij}] = E[x_{i+k, j+l}] = \mu. \quad (56)$$

Secondly, $\rho(x_{ij}, x_{i+k, j+l})$ only depends on the difference $((i+k, j+l) - (i, j)) = (k, l)$ between the sites, so we can write

$$\rho(x_{ij}, x_{i+k, j+l}) = \rho(k, l). \quad (57)$$

It is also worth noticing that for binary weakly stationary fields all the moments are the

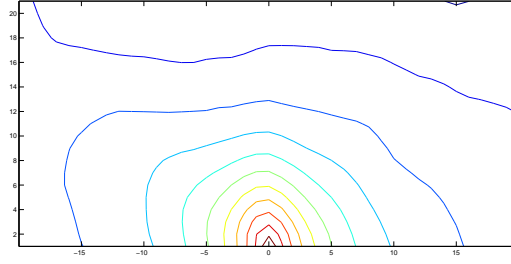


Figure 7: Contour plot of the empirical correlation function for the observation in Figure 1. The center of the elliptic contour lines has the highest correlation value equal to 1, and it decreases with distance from the center.

same, i.e. $\mu = E[x_{ij}] = E[x_{ij}^2] = E[x_{ij}^3] = \dots$, and the expression for the variance can be written on the simple form

$$\text{Var}[x_{ij}] = \mu - \mu^2 = \mu(1 - \mu). \quad (58)$$

The assumption of stationarity also results in the equality $\rho(k, l) = \rho(-k, -l)$, which can easily be verified by swapping the two factors in the nominator in (55). The fields are not assumed to be isotropic, and we therefore have to calculate the correlation function for positive vertical direction and both positive and negative horizontal direction. A natural unbiased estimator for μ is the mean $\hat{\mu} = \frac{1}{nm} \sum_{i,j} x_{ij}$, and together with the above assumptions the estimate for the correlation function is

$$\hat{\rho}(k, l) = \frac{\frac{1}{(m-|k|)(n-l)} \sum_{i=\max\{1, -k\}}^{\min\{m, m-k\}} \sum_{j=1}^{n-l} (x_{ij} - \hat{\mu})(x_{i+k, j+l} - \hat{\mu})}{\hat{\mu}(1 - \hat{\mu})}. \quad (59)$$

Figure 7 shows a contour plot of the estimated empirical correlation function. The closer the distance between the contour lines, the faster the correlation function decreases. If the contour lines are shaped like half circles we have equal properties in all directions, and asymmetry will therefore indicate an anisotropic field. For the observation the half circles have a more elliptic shape, where the longest axis is in the vertical direction. This does not necessarily indicate that the white objects are longer in the vertical direction than in the horizontal. Within each object the vertical extensions are approximately constant, and the horizontal extensions are very fluctuating. Thus, even though the white objects are horizontally longer, their irregular shape result in many low correlated areas in the horizontal direction, which explains the behaviour of the

correlation function.

4.2 SMRF model

To specify our sMRF model, we start with determining the neighbourhood, cliques and the set of conditioned variables we wish to include in the potential function. This will give us the arguments for the $\gamma(\cdot)$ -function, and we can determine its return values.

In the model we let the rows in the image matrix be the one-dimensional MRFs, where the first row corresponds to the first field, ϕ_1 . For each row we have a second order neighbourhood, i.e. the two nearest pixels both to the left and right of a site is a neighbour. The coordinates of the neighbours to pixel x_{ij} are shown in Figure 8.

$(i, j - 2)$	$(i, j - 1)$	(i, j)	$(i, j + 1)$	$(i, j + 2)$
--------------	--------------	----------	--------------	--------------

Figure 8: The coordinates of the neighbours to the pixel in (i, j) .

As a consequence of the choice of neighbourhood, there are three types of cliques c_1 , c_2 and c_3 . Without loss of generality, we set the potential for the two first cliques, c_1 and c_2 , equal to zero.

We let the set of dependent fields include the two previous fields, i.e. $D_i = \{i - 1, i - 2\}$. With the neighbourhood and the cliques defined above, the potential function is a function of three clique variables conditioned on variables from the two preceding rows. We include six conditioned variables, and get the set $x_{S_{N_i+j}} = \{x_{ij}, x_{i,j+1}, x_{i,j+2}, x_{i-1,j}, x_{i-1,j+1}, x_{i-1,j+2}, x_{i-2,j}, x_{i-2,j+1}, x_{i-2,j+2}\}$ of variables in the potential function. Figure 9 shows the coordinates of the nine variables included in the potential function.

$(i - 2, j)$	$(i - 2, j + 1)$	$(i - 2, j + 2)$
$(i - 1, j)$	$(i - 1, j + 1)$	$(i - 1, j + 2)$
(i, j)	$(i, j + 1)$	$(i, j + 2)$

Figure 9: The coordinates of the pixels included in the potential function.

The clique variables and the conditioned variables will together have $2^9 = 512$ different configurations. The $\gamma(\cdot)$ -function could be given a separate value for each of these configurations, but that would give 512 parameters to estimate. To reduce the number of parameters, the configurations are placed in 13 different groups, i.e. $\gamma(x_{S_{N_i+j}}) = t$

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
-2.57	-2.70	-3.51	-3.13	-1.05	-0.89	-1.23
θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	
-1.69	-1.64	-1.19	-1.34	-1.17	0.00	

Table 1: MLE parameters for the sMRF model. The last parameter, θ_{13} , is fixed to zero.

for $t = 1, 2, \dots, 13$. The model is then given by

$$\pi(x) = \pi(\phi_1, \phi_2) \prod_{i=3}^m K(\phi_{i-1}, \phi_{i-2}) \times \exp \left\{ - \sum_{t=1}^{13} n_t^i \theta_t \right\}, \quad (60)$$

where $\pi(\phi_1, \phi_2)$ is the probability for the two first rows.

We will not discuss the details of the grouping scheme, but instead give a short explanation of the idea behind it. We consider the conditioned variables together with the clique variables as pieces of a puzzle, and select those pieces that seem to appear most frequently in the observation and place them into 9 different groups. This way we can control the potential of the clique for the most important pieces. The rest of the pieces are placed in four separate groups, called rest groups, according to the number of horizontal changes between black and white within the clique. Figure 10 shows the pieces included in the 13 groups. We have groups of homogeneous pieces, corners (black and white), edges (vertical, horizontal and inclined) and rest groups. The gray color indicates that the pixels can take any value, except for values resulting in any of the configurations from the other groups.

4.2.1 Maximum likelihood estimation

The nonlinear minimization function $nlm()$ in R has no difficulty in finding the values of the parameters that minimize the negative logarithm of the likelihood function. Two realizations from the model are visualized in Figure 11 using the MLE parameter values in Table 1. To avoid the randomness effect for the first rows, we have simulated larger images and visualize only a subsample of them. The resolution of the images is the same as for the observation, 145×271 .

The MLE results are not satisfying. The realizations in Figure 11 contain a vertical pattern, and the estimator does not capture the fact that there should be white objects on a black background. The correlation function for the realizations in Figure 12 also

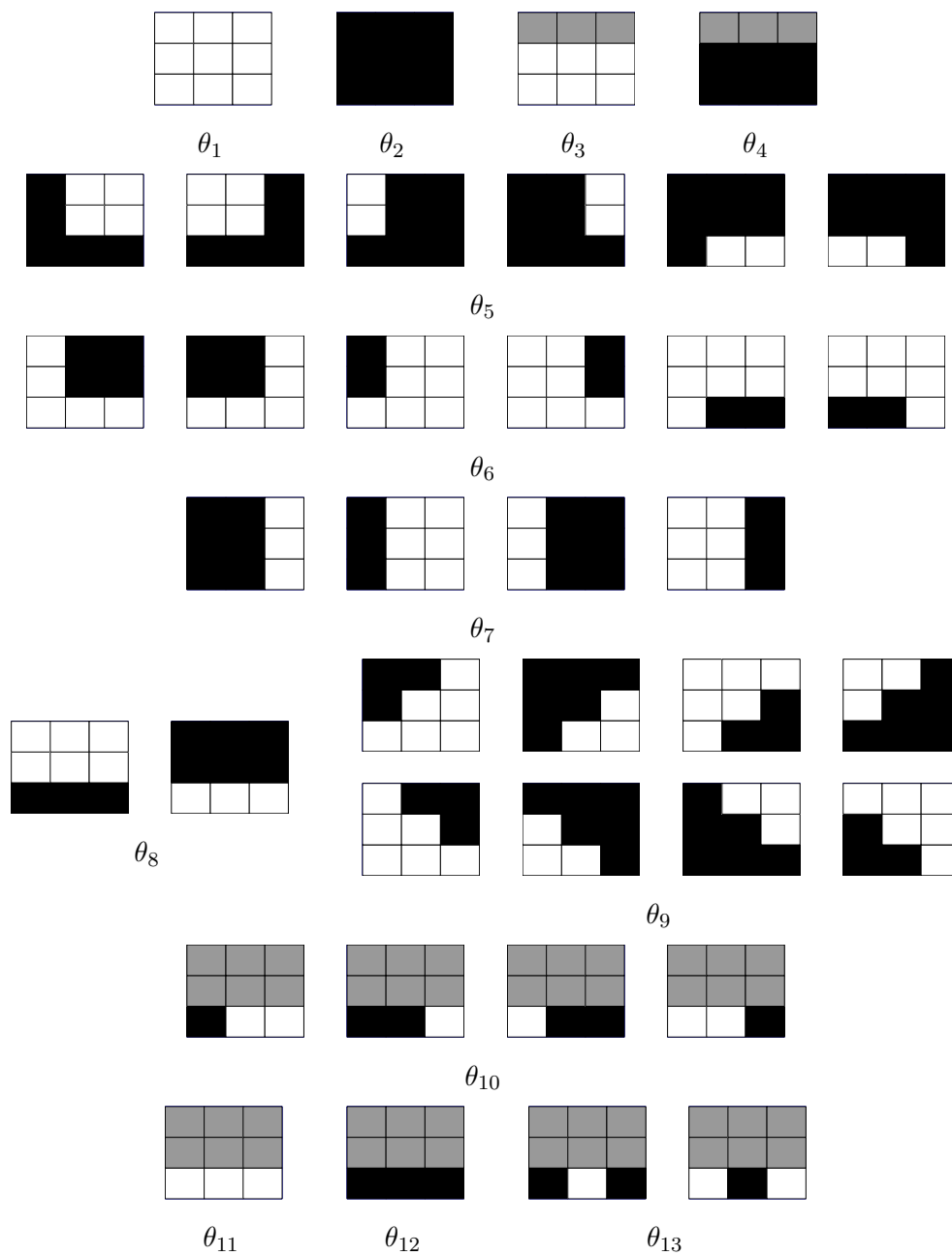


Figure 10: A visualization of the parametrization of the sMRF model. The gray color emphasize that the pixels can be either black or white as long as it does not result in any of the configurations from the other groups.



Figure 11: Realizations from the sMRF model with the MLE parameter values given in Table 1.

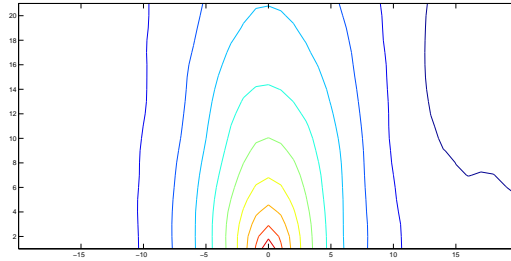


Figure 12: The correlation function for realizations from the sMRF model, using the MLE parameters in Table 1.

implies that the white objects are more correlated vertically compared to the observation. Horizontally, there is more resemblance with the empirical correlation.

The parameters, θ_i , can here be interpreted as probabilities. The parameters θ_3 and θ_4 can with low potentials create homogeneous areas, by giving two vertical changes between black and white within the configurations a low probability to occur. The vertical continuation of these homogeneous areas are decided by the potential for the first to parameters, θ_1 and θ_2 . Since the observation consists mostly of homogeneous areas, these four first parameters have, as expected, the lowest potentials. The fixed parameter, θ_{13} , has the highest potential. This is reasonable since the parameter represents the set of configurations where the clique is non-homogeneous, i.e. there are two changes between black and white within the clique.

It is also worth noticing that the potentials for the three first rest groups, θ_{10} , θ_{11} and θ_{12} , are approximately equal to the potentials for the groups of edges and corners, θ_6 , θ_7 , θ_8 and θ_9 , when we in fact would like them to be higher. The reason for this can be that

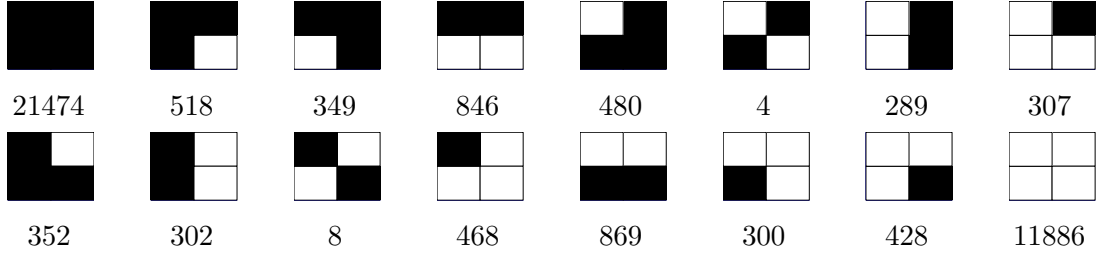


Figure 13: The 16 possible configurations used to match the sMRF model with the observation. Their corresponding frequencies in the observation are listed under each configuration.

the rest groups contain configurations that also can be considered as corners or edges. If the potentials for the corner and edges were lower, the homogeneous configurations would not be quite that dominant and the white objects would not extend so much vertically.

4.2.2 Matching specified functions

For this procedure we consider squares of four pixels. These squares will have $2^4 = 16$ different configurations, where each configuration is denoted by ζ_k for $k = 1, 2, \dots, 16$. For instance, if all the pixels in the square are black we get $\zeta_k = \{0, 0, 0, 0\}$. First, we define an indicator function

$$I(\{x_{ij}, x_{i,j-1}, x_{i-1,j}, x_{i-1,j-1}\} = \zeta_k) = \begin{cases} 1 & \text{if } \{x_{ij}, x_{i,j-1}, x_{i-1,j}, x_{i-1,j-1}\} = \zeta_k, \\ 0 & \text{otherwise,} \end{cases} \quad (61)$$

for $k = 1, \dots, 16$. Next we define 16 $h(\cdot)$ -functions, denoted by $h_k(\cdot)$ for $k = 1, 2, \dots, 16$. Each $h_k(\cdot)$ function measures the average number of times configuration ζ_k is repeated in realization x , i.e.

$$h_k(x) = \frac{1}{(m-1)(n-1)} \sum_{i=2}^m \sum_{j=2}^n I(\{x_{ij}, x_{i,j-1}, x_{i-1,j}, x_{i-1,j-1}\} = \zeta_k), \quad (62)$$

where $(m-1)(n-1)$ is the number of times a square configuration of four pixels can possibly be repeated in the image. These are the functions used in the estimate $\tilde{g}(\theta)$. In Figure 13, the 16 different configurations are visualized, together with the number of times they are repeated in the observation.

The minimization algorithm for $\tilde{g}(\theta)$, Algorithm 1, has iterated 945 times and run for 215 hours. Time limitation of this thesis made us interrupt the process before con-

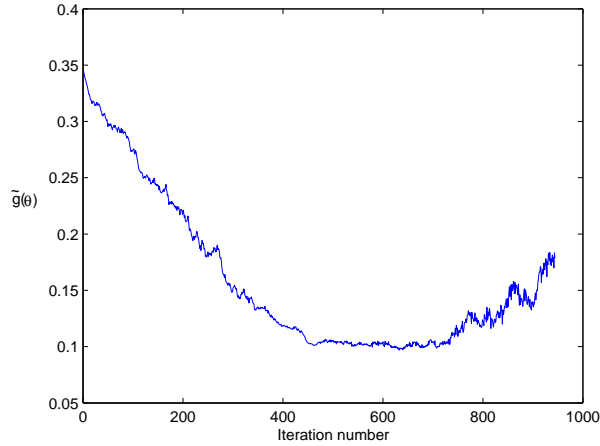


Figure 14: A plot showing how the estimate $\tilde{g}(\theta)$ moves as a function of the iteration number. The initial values for θ are the zero vector.

vergence was obtained. Our results are therefore incomplete and we are not in a place where we can draw conclusions about the estimator. However, we will show our temporary results and discuss them. Figure 14 shows how the function estimate, $\tilde{g}(\theta)$, moves with respect to the iteration number. The initial values of θ are the zero vector, the step length is set equal to $\epsilon = 0.01$, and the number of realizations in the estimate is $N = 500$. We have experimented with different step lengths, focusing on the weights, w_i , being sufficiently close to one. Optimization of the exact function $g(\theta)$ would be monotone, because we in the algorithm always move in the direction of a smaller value. However, optimization of the estimate $\tilde{g}(\theta)$, which is stochastic, might not be perfectly monotone, but should follow a downward trend. In Figure 14 we can see that $\tilde{g}(\theta)$ exhibits an upward trend for the last iterations. The reason for this trend is unclear. It might be because the estimate is biased, but this will not be investigated further in this thesis.

The temporary results from the estimation are tabulated in Table 2, and two realizations can be seen in Figure 15. From these realizations, it is not possible to extract information about the estimator's ability to reproduce the geological structure. Recall that the parameter θ_3 represents a configuration which creates white homogenous areas, whereas θ_1 represents the vertical continuation of these areas. If θ_3 continues to decrease, we might see the effect of the low potential for θ_1 and the white noise in the realizations will become white objects.

The frequencies of the different square configurations in the realizations, are listed in Figure 16. These frequencies are very different from the frequencies in the observation.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
-1.30	-1.53	-0.81	-2.10	-1.09	-1.20	-0.86
θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	
-0.21	-1.18	0.14	-0.79	-1.41	0.00	

Table 2: Temporary parameter values for the sMRF model, estimated by the optimization routine matching the $h(\cdot)$ -functions in (62).



Figure 15: Temporary simulation results from the sMRF model with the parameters in Table 2 estimated by matching the $h(\cdot)$ -functions in (62).

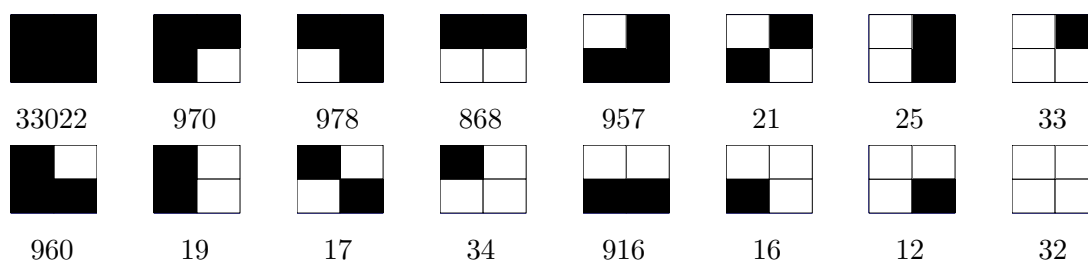


Figure 16: The 16 configurations used to match the sMRF model with the observation. The frequencies are measured from realizations of the model with the temporary parameter values in Table 2.



Figure 17: Realizations as a result of manually adjustment of the parameters, to show how flexible the sMRF model is.

The frequency of the black configuration is more than 10 000 higher than it should be, and opposite for the white configuration. Hopefully, this will even out if we let the optimization routine continue to run. However, it might be that the algorithm only finds a local minimum, and never obtain resemblance with the observation.

Even if the estimator finds a global minimum we might still end up with the same problem as for the MLE. That is, the frequencies of the 16 configurations are preserved, however, their distribution in the realizations do not resemble the observation. There are an infinite number of ways to specify the $h(\cdot)$ -functions, and we have only explored one option. The model is very flexible, meaning that by adjusting the parameter values we can get a wide range of different quantitative and qualitative features in the realizations. Figure 17 shows examples of two realizations which better resemble the observation. Thus, if our specifications are not satisfying, it is still possible that other $h(\cdot)$ -functions will improve the estimator.

4.3 MMRF model1

To specify a MMRF model we must first determine the ordering of the pixels and which pixels that should be included in the sequential neighbourhood. Once this is done, we will specify the $\gamma(\cdot)$ -function.

We choose the ordering where we sequentially sample the rows from top to bottom. Within each row we sample the sites sequentially from left to right. This ordering is also given in (26). In Eriksen (2005) the same ordering was considered, and it was concluded that the sequential neighbourhood was not large enough. The MLE results gave a highly directional pattern. We will therefore try to expand this neighbourhood to improve the results.

Eriksen (2005) considered two different sequential neighbourhoods, $x_{p\delta_i} = \{x_{i,j-1}, x_{i-1,j}\}$ and $x_{p\delta_i} = \{x_{i,j-1}, x_{i-1,j+1}, x_{i-1,j}, x_{i-1,j-1}\}$, and assigned a separate parameter for each configuration. For instance, for the largest sequential neighbourhood, the $\gamma(\cdot)$ -function was defined as $\gamma(x_{p\delta_i}) = k$ for $k = 1, \dots, 2^4$. For larger sequential neighbourhoods, some of the configurations may not occur in the data set, and we can not estimate the corresponding parameters. By assigning several configurations to the same parameter, we can use a larger sequential neighbourhood without getting too many parameters.

We consider a sequential neighbourhood of 6 sites, $n_{p\delta} = 6$, where $x_{p\delta_{ij}} = \{x_{i,j-1}, x_{i,j-2}, x_{i-1,j+1}, x_{i-1,j}, x_{i-1,j-1}, x_{i-2,j}\}$. With 6 sites in a sequential neighbourhood, we get a total number of $2^6 = 64$ different configurations of the neighbourhood. The frequencies of the 64 different configurations in the observation have been counted, ranging from 0 to 10198. In Figure 18, the 12 most frequently configurations are visualized. We assign a separate parameter for each of these configurations, and the rest of the configurations are placed in 8 different groups according to the configuration of the three nearest pixels, see Figure 19. The dark gray pixel will be black with a probability given by the corresponding parameter value, and the light gray color emphasize that the pixels can take any value except values resulting in any of the 12 configurations in Figure 18. The expression for the parameterized model is

$$\begin{aligned} \pi(x) &= \pi(x_{1\star}, x_{2\star}, x_{\star 1}, x_{\star 2}) \prod_{i=3}^m \prod_{j=3}^n \sum_{t=1}^{20} I(\gamma(x_{p\delta_{ij}}) = t) \theta_t^{x_{ij}} (1 - \theta_t)^{1-x_{ij}} \\ &\propto \prod_{t=1}^{20} \theta_t^{n_t^0} (1 - \theta_t)^{n_t^1}, \end{aligned} \quad (63)$$

where $\pi(x_{1\star}, x_{2\star}, x_{\star 1}, x_{\star 2})$ is the probability for the two first columns and rows.

4.3.1 Maximum likelihood estimation

The MLEs of the 20 parameters, $\theta_1, \theta_2, \dots, \theta_{20}$, are listed in Table 3. Two realizations with these parameters can be seen in Figure 20. The estimates generate realizations with different characteristics than what we see in the observation. Not only is it unclear whether there are white objects on a black background or vice versa, but their shape are also oriented vertically and slightly diagonally. The correlation function for the realizations can be seen in Figure 21. The function is not symmetric, which confirms the slightly diagonal direction of the pattern. Furthermore, the function decreases too rapidly in the horizontal direction and too slowly in the vertical direction, meaning that

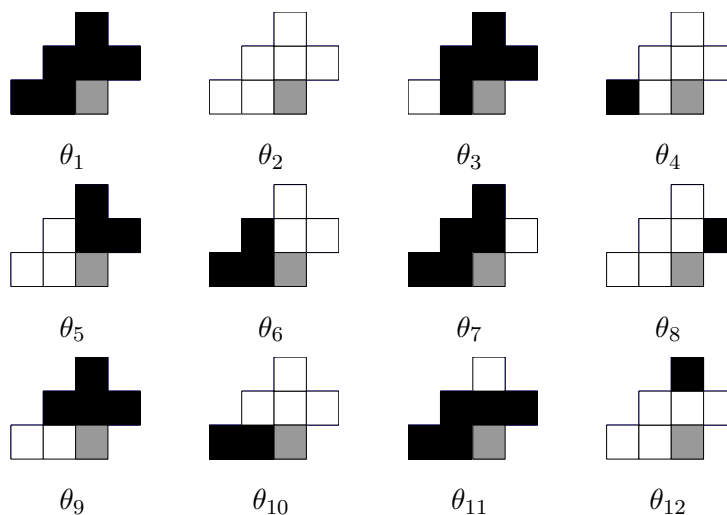


Figure 18: The 12 sequential neighbourhood configurations that appear most frequently in the observation, each assigned a separate parameter value. The dark gray color represent the pixels that are black with a probability given by the corresponding parameter value.

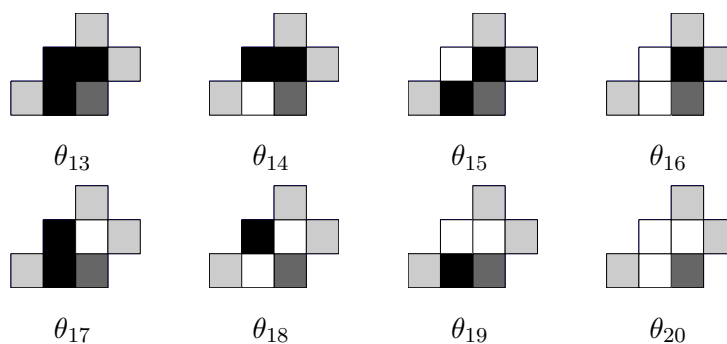


Figure 19: The 8 groups of the remaining configurations, grouped according to the configurations of the three nearest pixels. The light gray color emphasize that the pixels can take any value, except for values resulting in any of the configurations in Figure 18.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
0.99	0.01	0.98	0.02	0.75	0.32	0.74	0.25	0.67	0.34
θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}	θ_{18}	θ_{19}	θ_{20}
1.00	0.01	0.84	0.34	0.99	0.72	0.27	0.01	0.71	0.13

Table 3: MLE parameters for the MMRF model1.

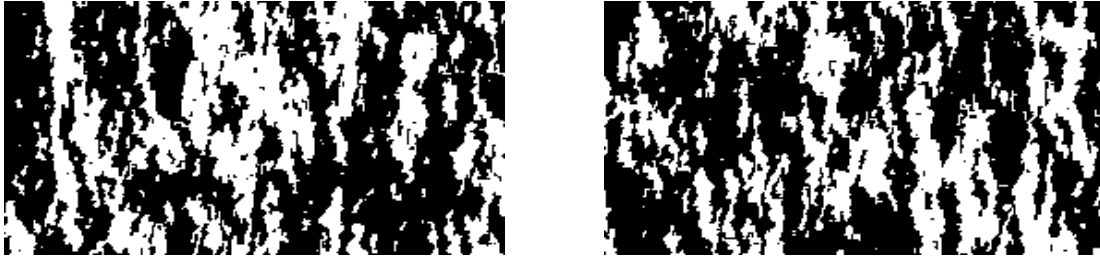


Figure 20: Realizations from the MMRF model1 with the MLE parameter values in Table 3.

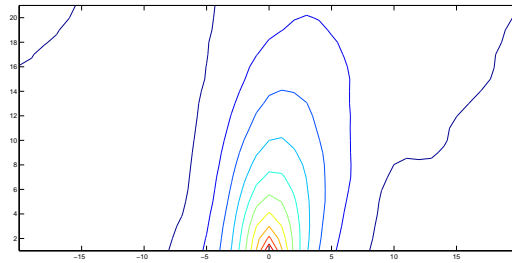


Figure 21: The correlation function for the realizations from the MMRF model 1, using the MLE parameter values in Table 3.

the white areas extend too much vertically and too little horizontally.

Recall that the parameter θ_t is the probability that a pixel is black, given the corresponding configuration of its sequential neighbourhood. For instance, $\theta_1 = 0.99$ means that the probability that a pixel is black when all the pixels in the sequential neighbourhood are black, is 0.99. We notice that the probability for a black pixel is high where it represents a black homogeneous area, i.e. the values of θ_1 and θ_{13} are high. The parameters θ_3 , θ_{11} and θ_{15} are also reasonable high. A low value of these parameters would result in many non-homogeneous areas. The opposite applies for the white areas. The rest of the parameters represent the three kinds of edges, vertical, horizontal and inclined. These have values more in the mid-range between zero and one.

We have further explored the model by adjusting the parameters manually and visually observed the changes. See Figure 22 for two examples of realizations that resemble the observation slightly better, as a result of manual adjustments. We could not find a set of parameters that gave us vertically shorter and horizontally longer white objects as

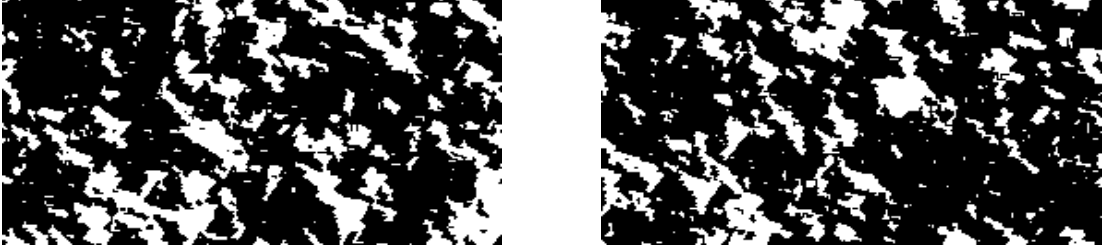


Figure 22: Realizations as a result of manually adjusting the parameters of the MMRF model2.

well as a non-diagonal pattern. Thus, we do not believe that it is the parameter values that causes the pattern to be oriented, but rather the ordering of the sites.

4.4 MMRF model2

We wish to improve the results obtained for the MMRF model1 by changing the order of simulation. The new model is denoted as MMRF model2. Rather than simulating every row from left to right, we change the direction for every other row, i.e. each row is simulated in the opposite direction as the one above. The ordering is given in (27).

We use the same sequential neighbourhood and parametrization as for the previous model, however, some modifications must be made. We are now dealing with two separate sequential neighbourhoods, one for the left to right simulation and one for the right to left simulation. Thus we get two different sets of pixels for the $\gamma(\cdot)$ -function, denoted by $x_{p\delta_{ij}}^{odd} = \{x_{i,j-1}, x_{i,j-2}, x_{i-1,j+1}, x_{i-1,j}, x_{i-1,j-1}, x_{i-2,j}\}$ and $x_{p\delta_{ij}}^{even} = \{x_{i,j+1}, x_{i,j+2}, x_{i-1,j+1}, x_{i-1,j}, x_{i-1,j-1}, x_{i-2,j}\}$. So far only the left to right sequential neighbourhood is specified, i.e. the $\gamma(\cdot)$ -function is only specified for the set $x_{p\delta_{ij}}^{odd}$. The specification of the $\gamma(\cdot)$ -function for $x_{p\delta_{ij}}^{even}$ is equivalent. If a configuration of $x_{p\delta_{ij}}^{odd}$ assigned a parameter θ_i is reflected with respect to column i , we get a configuration of $x_{p\delta_{ij}}^{even}$. This will be assigned the same parameter. So we have again restricted ourselves to a model of 20 parameters. See Figure 23 for an example of two separate sequential neighbourhoods with the same parameter value. In (a) we have a sequential neighbourhood for simulating from left to right and in (b) a sequential neighbourhood for simulating from right to left. The expression for the model is

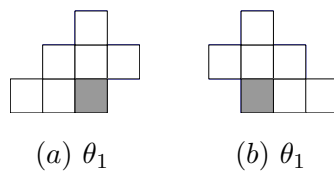


Figure 23: An illustration of how the sequential neighbourhood is reflected and assigned the same parameter value. (a) neighbourhood when simulating from left to right. (b) the same neighbourhood simulating from right to left.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
0.99	0.01	0.98	0.02	0.70	0.34	0.79	0.22	0.70	0.33
θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}	θ_{18}	θ_{19}	θ_{20}
1.00	0.01	0.91	0.38	0.99	0.63	0.30	0.01	0.66	0.10

Table 4: MLE parameter values for the MMRF model2.

$$\begin{aligned}
\pi(x) &= \pi(x_{1\star}, x_{2\star}, x_{\star 1}, x_{\star 2}, x_{\star, n-1}, x_{\star n}) \prod_{i \text{ odd}} \prod_{j=3}^n \sum_{t=1}^{20} I(\gamma(x_{p\delta_{ij}}^{\text{odd}}) = t) \theta_t^{x_{ij}} (1 - \theta_t)^{1-x_{ij}} \\
&\times \prod_{i \text{ even}} \prod_{j=n-2}^1 \sum_{t=1}^{20} I(\gamma(x_{p\delta_{ij}}^{\text{even}}) = t) \theta_t^{x_{ij}} (1 - \theta_t)^{1-x_{ij}}, \tag{64}
\end{aligned}$$

where $\pi(x_{1\star}, x_{2\star}, x_{\star 1}, x_{\star 2}, x_{\star, n-1}, x_{\star n})$ is the probability for the two first rows, and the two first and last columns.

4.4.1 Maximum likelihood estimation

The MLEs of the 20 parameters, $\theta_1, \theta_2, \dots, \theta_{20}$, are listed in Table 4 with two corresponding realizations visualized in Figure 24. The realizations, again, do not give satisfying results, although the diagonal structure has been removed. They still exhibit a vertical structure. The plot of the correlation function in Figure 25, confirms our visual conclusions. The contour lines are now symmetric, however, the correlation function decreases too slowly vertically and too rapidly horizontally. The parameter values are reasonable and we refer to the interpretation for the MMRF model1.

Manual adjustments of the parameters have given better resemblance with the observation, see Figure 26. However, a difficulty arises with the shape of the white objects, they have a larger vertical extension than horizontal and we have not been able to find

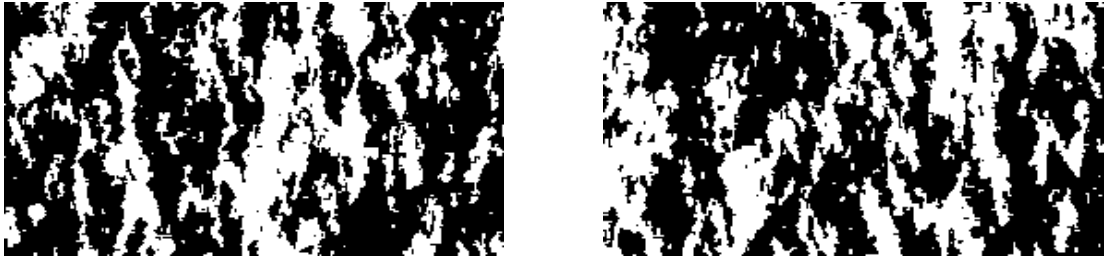


Figure 24: Realizations from the MMRF model2 with the MLE parameter values in Table 4.

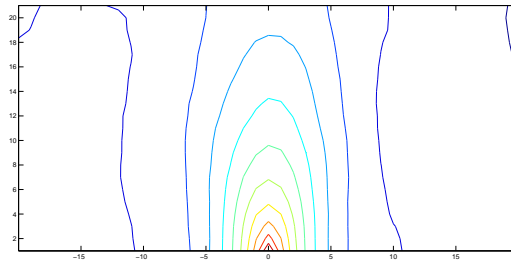


Figure 25: The correlation function for realizations from the MMRF model2, using the MLE parameter values in Table 4.

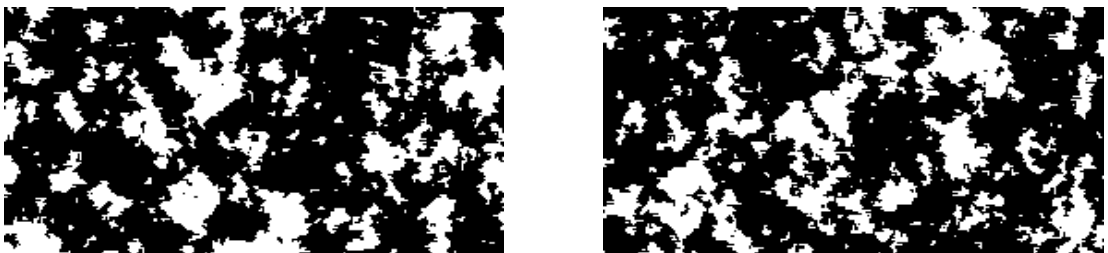


Figure 26: Realizations as a result of manually adjusting the parameters for the MMRF model2

a set of parameters where this is not the case. We suspect that the sequential neighbourhood is not large enough.

4.5 Discussion of the results in two dimensions

We are now able to draw some conclusions based on the study of our models in two dimensions. First of all, by looking at the realizations from the various models it is easy to tell the order in which they have been simulated. This lack of symmetry in the realizations is a result of how the model classes are defined. However, there are ways to specify models where the non-symmetric effect will be much less evident or even not evident at all.

First we start with the MLE results for the sMRF model. The homogenous configurations dominate, and combined with the small neighbourhood we consider it is natural that the pattern follows the downward direction of simulation. Vertically, we consider only three pixels simultaneously in the simulation, which might not be enough to control the vertical extension of the white objects. Horizontally we have symmetry in the simulation, and achieve more resemblance with the observation since we consider five pixels simultaneously.

The second estimator based on matching the $h(\cdot)$ -functions did not give us any final results, since the optimization routine did not converge in time. There exists an infinite number of possibilities of specifying this estimator and we have only explored a very simple one. One suggestion could be to specify functions that measure vertical and horizontal extensions as well as how the edges are distributed around the white objects. It is, however, important that the estimator not only coincide with the data but also with the model. The model has proved to be very flexible, but it might not be flexible enough. Another possibility is of course to find a model with a larger neighbourhood that will coincide with the MLE. This is the simplest estimator of the two and also more efficient, but we should remember that larger neighbourhoods result in more complications with the parametrization.

None of the two MMRF models we have studied have been able to reproduce the characteristics of the white objects. The MMRF model1 gave slightly diagonal realizations, which we managed to remove by changing the order of simulation. Also here, our results indicate that larger neighbourhoods might improve the results. When we sample the MMRF models we consider only three pixels both vertically and horizontally, and this might not be enough to control the extension of the white objects.

Since the MMRF models proved not to be rich enough to give satisfying realizations

for any set of parameters, we did not try the second estimator. However, this class of models is the simplest and most efficient of the two classes considered in this report, and a larger sequential neighbourhood with an appropriate parametrization should be further investigated.

5 Models in three dimensions

In this section we specify a model from the sMRF where we consider a three-dimensional regular lattice. The observed data is only available in two dimensions, which complicates the parameter estimation. Due to lack of time, we do not investigate the parameter estimation any further. Instead we adjust the parameters manually by visual measures in order to find a satisfying set of parameters. Then, finally, we test the MLE on a data set originating from the specified model, given the set of parameters.

The notation we will use in three dimensions is similar to the one in two dimensions, except that we express the lattice by a three-dimensional matrix, an $m \times n \times p$ -matrix. The pixels are now volume elements or voxels. We refer to the illustration in Figure 27 where the volume is visualized together with the following notation. Each voxel is given a location (i, j, k) in the matrix, for $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$ and $k = 1, 2, \dots, p$. We let x_{ijk} be the voxel in row i , column j and layer k of the matrix. The one-dimensional MRFs are still the rows in the image, however, they also have a layer location. Thus, we denote their location by ϕ_{ik} , and the ordering of the MRFs is therefore not immediately clear. To determine the ordering we define an order function, $\Omega(i, k)$, taking i and k as arguments. The set $\phi_{i\star} = \{\phi_{i1}, \dots, \phi_{ip}\}$ contains the MRFs with row number i and $\phi_{\star k} = \{\phi_{1k}, \dots, \phi_{mk}\}$ are the MRFs in layer k . The set of preceding dependent fields is denoted by D_{ik} .

To be able to extract visual information from the realizations, we have to view them through cross-sections. The three most natural are the cross-sections (i, j) , (i, k) and (j, k) . Also when we visualize parts in the parametrization, we use cross-sections.

5.1 Model specification

The sMRF in three dimensions can be regarded as a MMRF in two dimensions, where each site in the MMRF is a one-dimensional MRF. This is illustrated in Figure 28. For the MMRFs we must determine the ordering of the sites, which in three dimensions

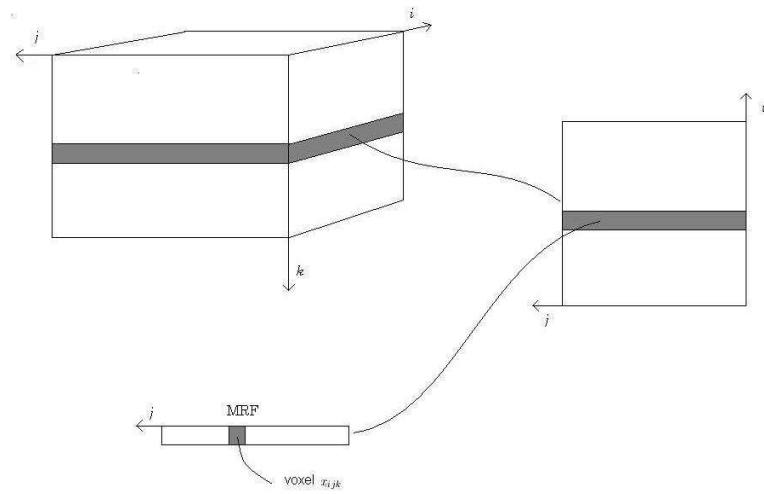


Figure 27: Illustration of a sMRF on a regular grid in three dimensions. Each row in the cross-sections (i, j) and (j, k) are one-dimensional MRFs.

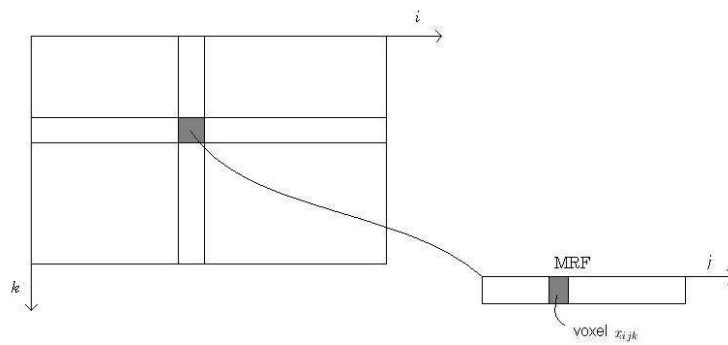


Figure 28: Illustration of the three-dimensional sMRF, viewed as a two-dimensional MMRF where each site is a one-dimensional MRF.

$$\boxed{(i, j - 2, k) \mid (i, j - 1, k) \mid (i, j, k) \mid (i, j + 1, k) \mid (i, j + 2, k)}$$

Figure 29: The coordinates of the neighbours to the voxel with coordinates (i, j, k) in ϕ_{ik} .

means that we must determine the ordering of the one-dimensional MRFs. For the MMRF models we studied in two dimensions, we obtained best results when changing the order of simulation for every other row. Thus, we will change the order of simulation for every other layer, resulting in the ordering

$$\Omega(i, k) = \begin{cases} m(k - 1) + i & \text{if } k \text{ is odd} \\ mk - i + 1 & \text{if } k \text{ is even} \end{cases} \quad (65)$$

for each MRF, and for the whole field the ordering is

$$\begin{bmatrix} 1 & 2 & \dots & m \\ 2m & 2m - 1 & \dots & m + 1 \\ \vdots & \vdots & \ddots & \vdots \\ (p - 2)m + 1 & (p - 2)m + 2 & \dots & (p - 1)m \\ mo & mo - 1 & \dots & (p - 1)m + 1 \end{bmatrix}. \quad (66)$$

With this ordering we sample each layer k as a sMRF with a direction opposite to the layer above, and the layers are sampled sequentially from top to bottom.

Even though the field is explained through MMRF we are still considering a three-dimensional sMRF. The MMRF is only used for determination of the ordering of the one-dimensional MRFs. Thus, as for the two-dimensional sMRF, we have to determine the set of voxels to be included in the potential function.

We let each MRF in the field have a second order neighbourhood, and use cliques of size three, c_3 . Figure 29 shows the coordinates of the neighbours to voxel x_{ijk} in ϕ_{ik} . With the ordering of the MRFs given in (66), we have to specify two sets of preceding dependent fields, $\phi_{D_{ik}}$, one for odd layers and one for even layers. For odd k 's we let $D_{ik}^{odd} = \{(i - 1, k), (i - 2, k), (i, k - 1), (i, k - 2)\}$ and for even k 's we let $D_{ik}^{even} = \{(i + 1, k), (i + 2, k), (i, k - 1), (i, k - 2)\}$. In the potential function we include one voxel from each of these fields, the voxels with column index j . Thus, the sets of voxels in the potential function are $x_{S_{ijk}^{odd}} = \{x_{i,j-1,k}, x_{ijk}, x_{i,j+1,k}, x_{i-1,j,k}, x_{i-2,j,k}, x_{i,j,k-1}, x_{i,j,k-1}\}$ and $x_{S_{ijk}^{even}} = \{x_{i,j-1,k}, x_{ijk}, x_{i,j+1,k}, x_{i+1,j,k}, x_{i+2,j,k}, x_{i,j,k-1}, x_{i,j,k-1}\}$. Figure 30 shows the set preceding dependent MRFs for both odd and even layer number, k . This is the cross-section (i, k) . For odd layer numbers we also visualize the two other cross-sections

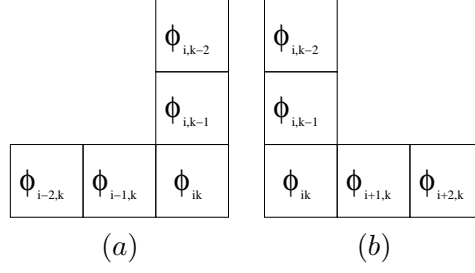


Figure 30: Cross-sections visualizing the set of dependent MRFs included in the potential function. (a) valid for odd layer numbers (b) valid for even layer numbers.

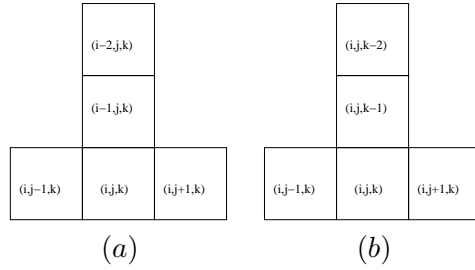


Figure 31: Cross-sections visualizing the coordinates of the voxels included in the potential function for odd layer numbers. (a) cross-section (i, j) (b) cross-section (j, k) .

to show the coordinates of the set of voxels from D_{ik}^{odd} that are included in the potential function, see Figure 31. We can now write the probability distribution as

$$\begin{aligned} \pi(x) &= \pi(\phi_E) \times \prod_{k \text{ odd}} \prod_{i=3}^m K(\phi_{D_{ik}^{odd}}) \exp \left\{ - \sum_{j=1}^{n-2} V(x_{S_{ijk}^{odd}}) \right\} \\ &\times \prod_{k \text{ even}} \prod_{i=m-2}^1 K(\phi_{D_{ik}^{even}}) \exp \left\{ - \sum_{j=1}^{n-2} V(x_{S_{ijk}^{even}}) \right\}, \end{aligned} \quad (67)$$

where $\phi_E = \{\phi_{\star 1}, \phi_{\star 2}, \phi_{1\star}, \phi_{2\star}, \phi_{m-1\star}, \phi_{m\star}\}$ are the MRFs at the edges, and their probability is $\pi(\phi_E)$.

The number of voxels included in the potential function is at all times equal to 7, and that gives a total number of $2^7 = 128$ possible configuration of the sites in the function. With two separate sets of voxels, x_{ijk}^{odd} and x_{ijk}^{even} , we get $128 \times 2 = 256$ parameters to estimate. To reduce the number of parameters, we divide the set of voxels into three separate groups and define a separate potential function for each of these, i.e. we let the

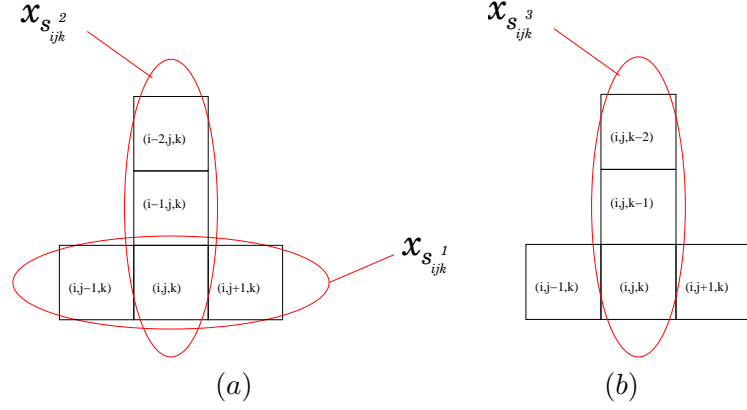


Figure 32: A visualization of how the 7 voxels in the potential function are divided into three sets of voxels, where each set are assigned a separate parameter value for every of its 8 possible configurations.

potential function, $V(\cdot)$, be a sum of three potential functions, $V_1(\cdot)$, $V_2(\cdot)$ and $V_3(\cdot)$, each taking three arguments. The sets of voxels are denoted by $x_{S_{ijk}^{odd}}^1$, $x_{S_{ijk}^{odd}}^2$ and $x_{S_{ijk}^{odd}}^3$ for odd k 's, and $x_{S_{ijk}^{even}}^1$, $x_{S_{ijk}^{even}}^2$ and $x_{S_{ijk}^{even}}^3$ for even k 's. Thus, we get the following potential functions

$$V(x_{S_{ijk}^{odd}}) = V_1(x_{S_{ijk}^{odd}}^1) + V_2(x_{S_{ijk}^{odd}}^2) + V_3(x_{S_{ijk}^{odd}}^3) \quad (68)$$

$$V(x_{S_{ijk}^{even}}) = V_1(x_{S_{ijk}^{even}}^1) + V_2(x_{S_{ijk}^{even}}^2) + V_3(x_{S_{ijk}^{even}}^3). \quad (69)$$

If we consider odd k 's, the three sets are $x_{S_{ijk}^{odd}}^1 = \{x_{ijk}, x_{i,j+1,k}, x_{i,j+2,k}\}$, $x_{S_{ijk}^{odd}}^2 = \{x_{i,j+1,k}, x_{i-1,j+1,k}, x_{i-2,j+1,k}\}$ and $x_{S_{ijk}^{odd}}^3 = \{x_{i,j+1,k}, x_{i,j+1,k-1}, x_{i,j+1,k-2}\}$. These three sets are also visualized in Figure 32. We notice that voxel x_{ijk} is included in all three groups. The three potential functions each have 8 possible return values, giving a total of $(8 + 8 + 8) \times 2 = 48$ different parameters. We assign the same parameter to odd and even sets with the same configuration. That is, when a configuration for even k 's is reflected with respect to row i we get a configuration for odd k 's and these two configurations have the same potential function value. This scheme reduces the number of parameters to 24.

We define three $\gamma(\cdot)$ -function, one for every potential function

$$\gamma_1(x_{S_{ijk}^1}^{odd}) = \gamma_1(x_{S_{ijk}^1}^{even}) = t \text{ for } t = 1, \dots, 8 \quad (70)$$

$$\gamma_2(x_{S_{ijk}^2}^{odd}) = \gamma_2(x_{S_{ijk}^2}^{even}) = t \text{ for } t = 1, \dots, 8 \quad (71)$$

$$\gamma_3(x_{S_{ijk}^3}^{odd}) = \gamma_3(x_{S_{ijk}^3}^{even}) = t \text{ for } t = 1, \dots, 8 \quad (72)$$

In addition, we introduce three sets of parameters corresponding to the $\gamma(\cdot)$ -functions, $\theta_{\gamma_1(\cdot)}^1 = V_1(\cdot)$, $\theta_{\gamma_2(\cdot)}^2 = V_2(\cdot)$ and $\theta_{\gamma_3(\cdot)}^3 = V_3(\cdot)$. Using this we get

$$\begin{aligned} \pi(x) = & \pi(\phi_E) \prod_{k \text{ odd}} \prod_{i=3}^m K(\phi_{D_{ik}}^{odd}) \exp \left\{ - \sum_{t=1}^8 n_t^{1ik} \theta_t^1 + n_t^{2ik} \theta_t^2 + n_t^{3ik} \theta_t^3 \right\} \\ & \times \prod_{k \text{ even}} \prod_{i=m-2}^1 K(\phi_{D_{ik}}^{even}) \exp \left\{ - \sum_{t=1}^8 n_t^{1ik} \theta_t^1 + n_t^{2ik} \theta_t^2 + n_t^{3ik} \theta_t^3 \right\}, \end{aligned} \quad (73)$$

where n_t^{1ik} , n_t^{2ik} and n_t^{3ik} represent the number of times the functions $\gamma_1(\cdot)$, $\gamma_2(\cdot)$ and $\gamma_3(\cdot)$ respectively return the value t in field ϕ_{ik} .

In Table 5 the different parameters are listed together with their corresponding configurations. The ordering in which the elements are listed in the three different sets corresponds to the ordering in the configurations. For instance for set $x_{S_{ijk}^1} = \{x_{i,j-1,k}, x_{ijk}, x_{i,j+1,k}\}$ the element $x_{i,j-1,k}$ will be represented by the first voxel in the configuration.

We have adjusted the parameters using visual measures and we have found a tolerable good set of values. These values are tabulated in Table 6, and two images for each cross-section are visualized in Figure 33. The resolution of the field is $200 \times 200 \times 200$. In two dimensions we sampled large fields and used only a sub-sample of the field for visualization and other measurements. In this way we avoided the effect of the sites in ϕ_E , i.e. the effect of the randomness at the edges. In three dimensions it is very time-consuming and it requires a high amount of memory to sample larger fields. The edge effect will therefore be present in the realizations from our model.

The observation we have been given represents a geological structure as we move downwards towards the center of the earth. Therefore, we do not know the appearance of the structure when viewed from above, in this context that would be the cross-section (i, j) . However, by focusing on making the two other cross-sections resemble the observation, we might obtain sufficiently accurate results for the (i, j) cross-section.

The realizations from cross-section (i, j) and (i, k) resemble the observation since









Configuration	$x_{S_{ijk}}^1$	$x_{S_{ijk}}^2$	$x_{S_{ijk}}^3$
	θ_1^1	θ_1^2	θ_1^3
	θ_2^1	θ_2^2	θ_2^3
	θ_3^1	θ_3^2	θ_3^3
	θ_4^1	θ_4^2	θ_4^3
	θ_5^1	θ_5^2	θ_5^3
	θ_6^1	θ_6^2	θ_6^3
	θ_7^1	θ_7^2	θ_7^3
	θ_8^1	θ_8^2	θ_8^3

Table 5: The list of parameters with their corresponding configuration. The ordering in which each voxel in the sets are listed corresponds to the ordering in the configuration.

θ_1^1	θ_2^1	θ_3^1	θ_4^1	θ_5^1	θ_6^1	θ_7^1	θ_8^1
-1.80	-0.80	6.20	-0.80	-0.80	6.20	-0.80	-1.80
θ_1^2	θ_2^2	θ_3^2	θ_4^2	θ_5^2	θ_6^2	θ_7^2	θ_8^2
0.47	-0.33	6.17	1.17	1.20	6.20	-0.30	0.50
θ_1^3	θ_2^3	θ_3^3	θ_4^3	θ_5^3	θ_6^3	θ_7^3	θ_8^3
0.00	-1.30	6.20	1.20	1.20	6.20	-1.30	0.00

Table 6: Parameter values based on manually adjustments of the parameters using visual measures.

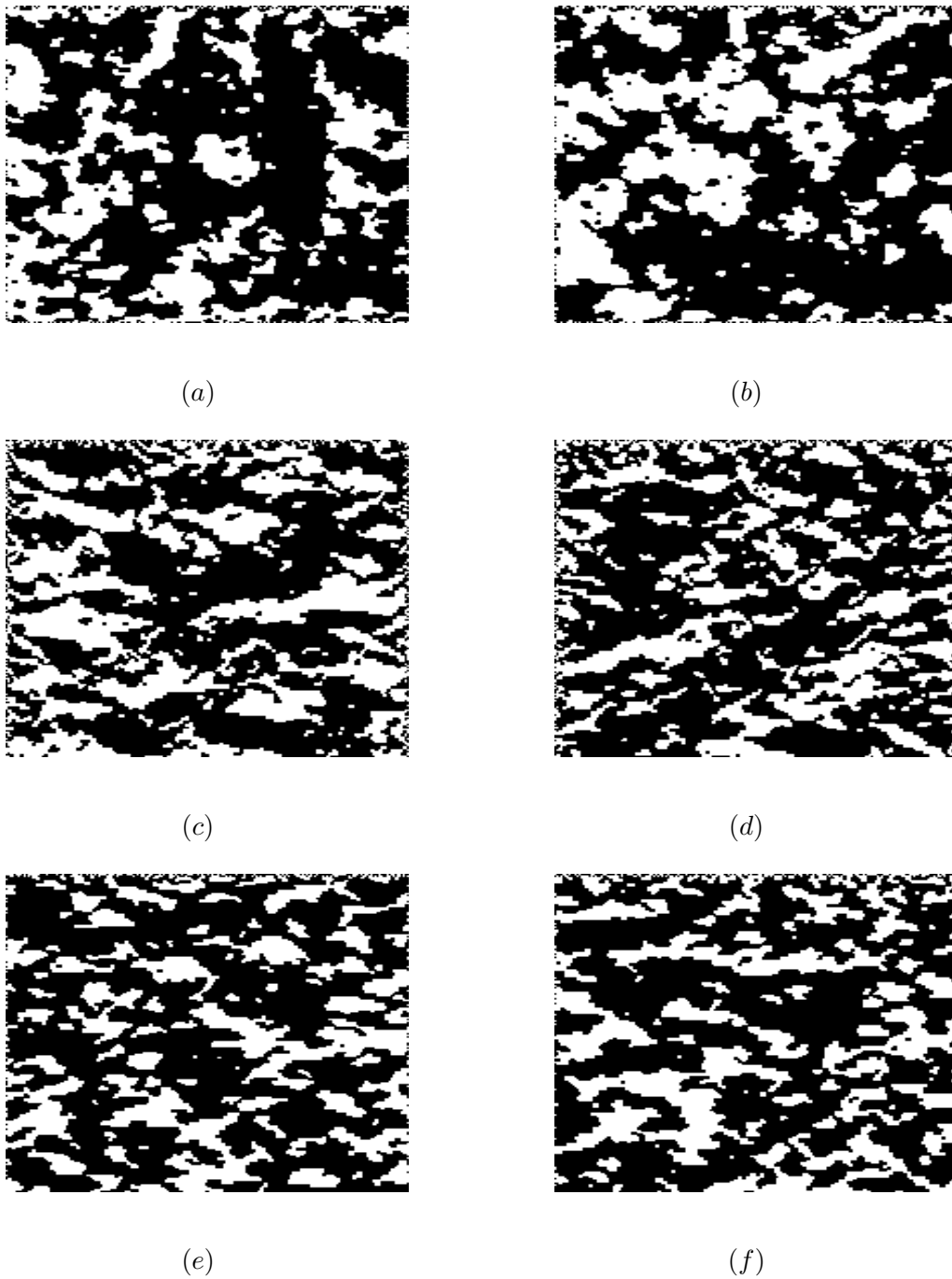


Figure 33: Realizations for the three cross-sections using the manually adjusted parameter values in Table 6. (a) and (b) cross-section (i, j) . (c) and (d) cross-section (i, k) . (e) and (f) cross-section (j, k) .

θ_1^1	θ_2^1	θ_3^1	θ_4^1	θ_5^1	θ_6^1	θ_7^1	θ_8^1
-1.90	-0.89	5.80	-0.80	-0.92	6.31	-0.78	-1.77
θ_1^2	θ_2^2	θ_3^2	θ_4^2	θ_5^2	θ_6^2	θ_7^2	θ_8^2
0.30	-2.58	3.88	1.00	0.96	3.84	-2.62	0.26
θ_1^3	θ_2^3	θ_3^3	θ_4^3	θ_5^3	θ_6^3	θ_7^3	θ_8^3
0.06	-3.07	4.34	1.36	1.27	4.33	-3.08	0.00

Table 7: Parameter values based on MLE with a realization from the model given the parameters in Table 6 as the observation.

we clearly have white objects on a black background, which extend more horizontally than vertically. For cross section (i, k) the edges of the white objects have a diagonal structure, which is not present in the observation. However, the model is very flexible and it might be possible to find even better parameter values, and maybe values where the shapes of the white objects are more satisfying.

The first column of parameters, $\theta_1^1, \dots, \theta_8^1$, represents the potential for the cliques, see Table 5. In Table 6, we see that the homogeneous cliques have been given the lowest potential, and the cliques with one horizontal change between black and white have a slightly higher potential. We do not want any of the configurations with two changes between black and white in the clique, so the parameters for these cliques are given a very high value. The second column of parameters in Table 5, $\theta_1^2, \dots, \theta_8^2$, represents the horizontal dependency in the field, and by adjusting these parameters we can determine the horizontal extension of the white and black areas, and how many changes we want between black and white. The same interpretation goes for the third column of parameters, but with the obvious modification that we adjust the vertical extension.

5.2 Maximum likelihood estimation with observations in three dimensions

We have used the above simulation results as an observation, and applied the MLE. The results are tabulated in Table 7, and Figure 34 shows two realizations for each cross-section. The realizations are very similar to the observation. We can therefore conclude that MLE works well for three-dimensional data originating from the statistical model.

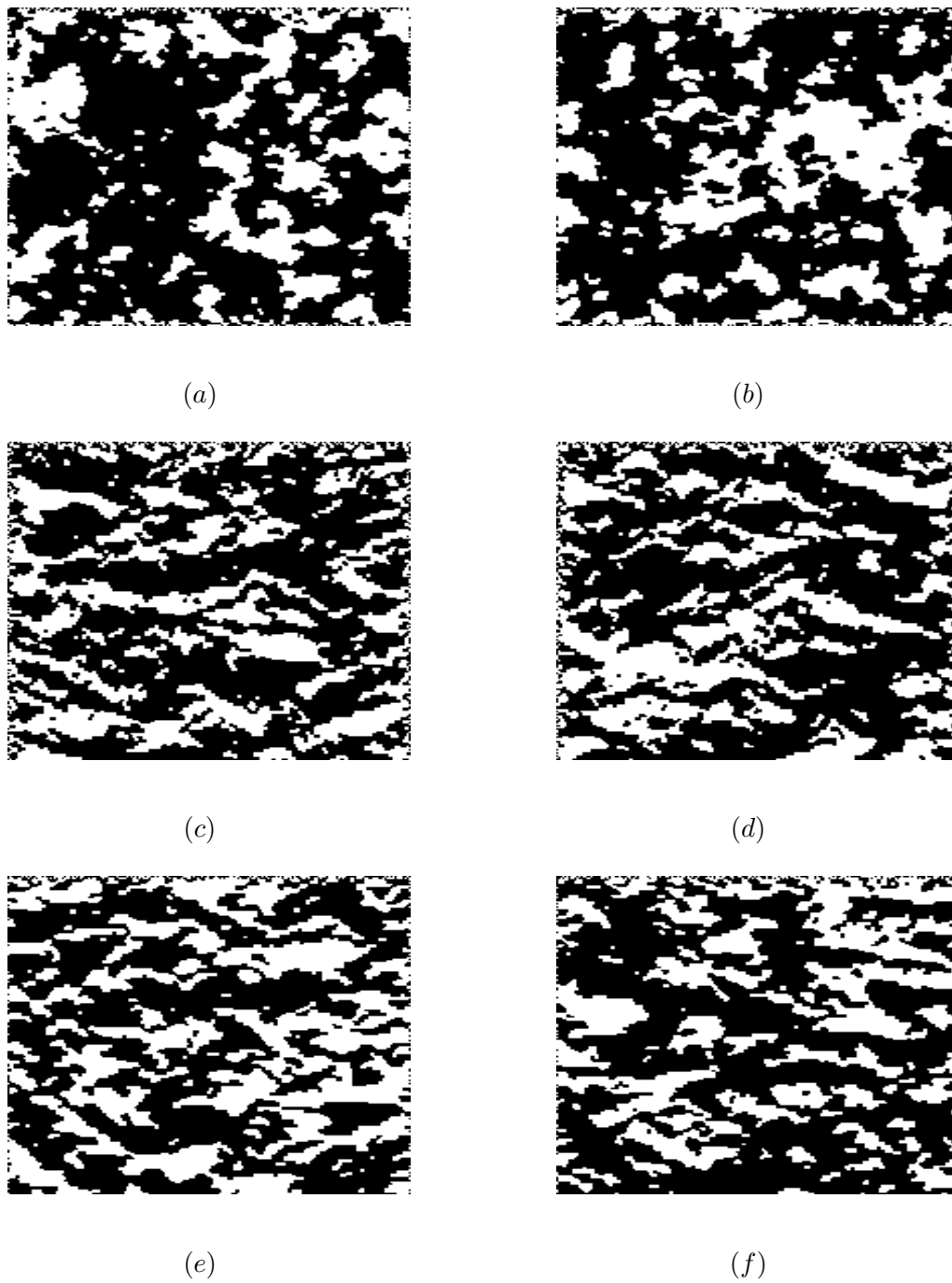


Figure 34: Realizations for the three cross-sections using the MLE parameter values in Table 7. (a) and (b) cross-section (i, j) (c) and (d) cross-section (i, k) (e) and (f) cross-section (j, k) . The observed data used for the estimation is a realization from the model using the parameter values in Table 6.

5.3 Discussion of the results in three dimensions

The model we have specified has given us some very promising results. However, we are not fully satisfied with the appearance of edges for the white objects in cross-section (i, k) . Maybe the model is not flexible enough to remove this diagonal effect, and other parametrizations should be considered. It might also be a good idea to include more voxels in the potential function, not only from the horizontal and vertical direction but also diagonally. This should be further investigated.

Regarding the estimation of the parameters, we run into some problems. We do not have available any three-dimensional observation, only a two-dimensional image, and this complicates the estimation procedure. We let y be a two-dimensional field from a cross-section of the whole field x , and x_{-y} is the whole field except for the cross-section y . Then, the likelihood can be written

$$L(\theta|y) = \sum_{x_{-y}} \pi(x|\theta) = \pi(y|\theta), \quad (74)$$

where we sum over the possible outcomes of x_{-y} to be left with a likelihood for two-dimensional observations. Thus, we can replace y with our observation and maximize the likelihood function. In practice this is of course an unattainable process, since our fields are too large for the summation in (74) to be performed. Approximative techniques are required, however, we do not have time to investigate such techniques any further in this thesis.

6 Closing remarks

The aim of this thesis has been to find an appropriate statistical prior model in three dimensions, that can reproduce geological structures. We have required the models to be efficient when it comes to simulation and parameter estimation, and two classes of models fulfilling these requirements are the sMRFs and MMRFs.

The observation is only available in two dimensions, which complicates the parameter estimation. For that reason, we have first focused on reproducing the geological structure in two dimensions to become familiar with the classes and possible estimators. The results for the two-dimensional models have not been fully satisfying. The MLE did not give good results for any of our models. A second estimator based on matching a set of square configurations was considered only for the sMRF model, and did not converge in time for us to evaluate it.

Further research of the models and estimators in two dimensions remains. However, we have used what we have learned from the study of these models to specify a three-dimensional sMRF model. This model proved to be very flexible and promising, even though some modifications of the parametrization might improve the results. The estimation of the parameters is challenging and remains to be further investigated.

References

- Abend, K., Harley, T., and Kanal, L. “Classification of binary patterns.” *IEEE Transactions on information theory*, 11:538 (1965).
- Eriksen, J. “On discrete lattice models for which efficient simulation is possible.” Master’s thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway (2005).
- Gray, A., Kay, J., and Titterington, D. “An empirical study of the simulation of various models used for images.” *IEEE Transactions on pattern analysis and machine intelligence*, 16:507 (1994).
- Hurn, M., Husby, O., and Rue, H. “A Tutorial on Image Analysis.” In Møller, J. (ed.), *Spatial Statistics and Computational Methods*, number 173 in Lecture Notes in Statistics, 87–139. Springer, Berlin (2003).
- Qian, W. and Titterington, D. “Multidimensional Markov chain models for image textures.” *Journal of the Royal Statistical Society, Series B*, 53(3):661 (1991).
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2005). ISBN 3-900051-07-0. URL <http://www.R-project.org>
- Stien, M. “Sequential Markov random fields for modelling of geological structures.” (2005). Project report, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
- Tjelmeland, H. “Stochastic models in reservoir characterization and Markov random fields for compact objects.” Ph.D. thesis, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway (1996).