NTNU
Norwegian University of
Science and Technology

Markov Random Field Modelling of Diagenetic Facies in Carbonate Reservoirs

Elisabeth Finserås Larsen

Master of Science in Physics and Mathematics
Submission date: July 2010
Supervisor: Karl Henning Omre, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

To define models for non-stationary profile Markov random fields with associated simulation algorithm. Implement the methodology and apply it to diagenetic facies inversion in a real carbonate reservoir.

This MSc-thesis is to be carried out at the Department of Mathematical Sciences under guidance by Professor Henning Omre. Statoil shall provide the reservoir data and Siv.Ing. Daniel Berge Sollien, Statoil, will support the study.

Assignment given: 09. February 2010
Supervisor: Karl Henning Omre, MATH

Preface

This work is a result of the course TMA4905 Statistics, Master Thesis. The course amounts to 30 credits or one semester at NTNU. The problem to be addressed was given by Professor Henning Omre in cooperation with Daniel Berge Sollien, a geologist at Statoil.

The background of the problem is the development of a model for a carbonate reservoir in the Middle East, and the aim of the thesis is to improve a specific part of this model. Starting the thesis I had close to no knowledge about geology, so needless to say I encountered many words and expressions new to me. I quickly learned that geologists and mathematicians converse on different levels, so finding a middle ground was essential for communication.

One of the main challenges in developing the model presented in this thesis was not overfitting the model. Emphasis has been put on making sure that all reasoning presented is mathematically justifiable.

I wish to thank my supervisor Henning Omre for his help and input and Daniel Berge Sollien for contributing with reservoir observations and teaching me what I needed to know about geology.

Elisabeth Larsen

Trondheim, 06.07.2010

Abstract

Bayesian inversion is performed on real observations to predict the diagenetic classes of a carbonate reservoir where the proportions of carbonate rock and depositional properties are known. The complete solution is the posterior model. The model is first developed in a 1D setting where the likelihood model is generalized Dirichlet distributed and the prior model is a Markov chain. The 1D model is used to justify the general assumptions on which the model is based. Thereafter the model is expanded to a 3D setting where the likelihood model remains the same and the prior model is a profile Markov random field where each profile is a Markov chain. Lateral continuity is incorporated into the model by adapting the transition matrices to fit a given associated limiting distribution, two algorithms for the adjustment are presented. The result is a good statistical formulation of the problem in 3D. Results from a study on real observations from a 2D reservoir show that simulations reproduce characteristics of the real data and it is also possible to incorporate conditioning on well observations into the model.

Contents

1 Introduction

Prediction of reservoir characteristics is increasingly important as petroleum resources become scarcer. With the help of good mathematical models, a geologist can use his prior knowledge to make a reliable prediction of reservoir characteristics. In Norway the majority of petroleum reserves are located in sandstone reservoirs, hence the competence on other types of reservoirs is limited.

The Norwegian oil and gas company Statoil are now venturing outside the North Sea in search of petroleum reserves, which entails developing fields other than the familiar sandstone reservoirs. The background of this thesis is a carbonate field located in the Middle East for which Statoil wishes to develop a model. Statoil has already predicted depositional properties and the proportions of different carbonate rock throughout the reservoir. Deposition is the process by which sediments settles on a surface and we assume that the depositional properties can be classified in one of two categories grainstone and fines, abbreviated GS and FS, where fines consists of wackestone and packstone. Grainstone, packstone and wackestone are part of the Dunham classification system for carbonate rock, see Dunham (1962). The carbonate rocks present in the reservoir are anhydrite, calcite and dolomite, the proportions of these sum to unity. The aim of this thesis is to predict a diagenetic class given the depositional property and the proportions of carbonate rock. Diagenesis is an alteration of sediment into sedimentary rock. The diagenetic classes are good, moderate, oomoldic and poor, abbreviated G, M, O and P respectively.

Ulvmoen & Omre (2010) presents a methodology for prediction of lithology/fluid characteristics based on prestack seismic data. The study defines the inversion in a Bayesian framework in a 3D target zone which is discretized into vertical profiles and lateral horizons. The likelihood model links the seismic data to the lithology/fluid characteristics and is defined independently for each profile. The prior model is a profile Markov random field with each profile modelled as a Markov chain. The model for the carbonate field will be similar to this, but the exact definition of the likelihood and prior model will be different due to the different natures of the problems.

2 Observations and Notation

For each node in the target zone the proportions of anhydrite, calcite and dolomite, denoted carbonate rocks, is considered known. The proportions are numbers between zero and one and in each node the proportions sum to one. There is also information about the depositional property in each node, either FS or GS.

The observations arise from wells drilled throughout the reservoir. When the wells were drilled, columns were extracted and parts of the columns were analyzed under a microscope to determine the proportions of carbonate rocks as well as diagenetic classes and depositional properties.

The information from this analysis as well as information from a permeability measurement was then utilized in order to estimate these properties for the entire well.

Using observations from the wells the proportions of carbonate rocks and depositional properties for the reservoir were estimated.

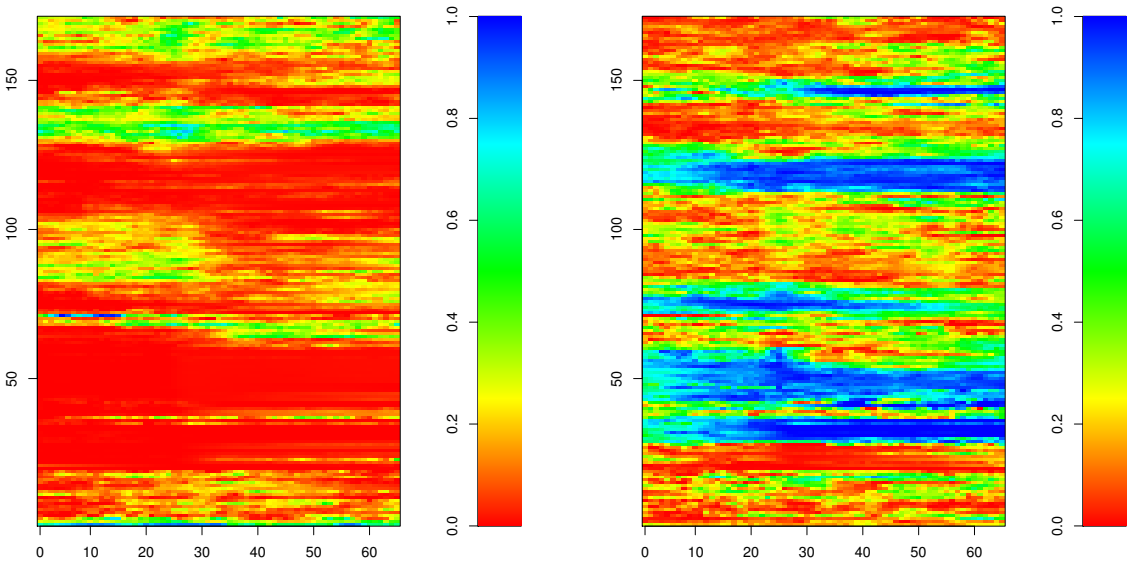
Figure 1 shows a plot of the data in the 2D target zone.

Denote the target zone in 3D by $\mathcal{D}^3$, and let the discretisation of the target zone be by the lattice $\mathcal{L}_\mathcal{D}$ divided into vertical profiles and lateral horizons. Denote a node in the lattice $(\mathbf{x}, t)$. The profiles are discretized upwards by the lattice $\mathcal{L}^t_\mathcal{D}$ where the nodes are labelled from 1 to T. The lateral horizons are discretized by the lattice $\mathcal{L}^\mathbf{x}_\mathcal{D}$. The diagenetic class in node $(\mathbf{x}, t)$ is denoted $\pi_{\mathbf{x},t}$. Further we define the diagenetic classes $\pi_{\mathbf{x},t} \in \{G, M, O, P\}$. The complete set of diagenetic classes in the 2D target zone is represented by $\boldsymbol{\pi} : \{\pi_{\mathbf{x},t}; (\mathbf{x}, t) \in \mathcal{L}_\mathcal{D}\}$.

Denote the proportions of anhydrite, calcite and dolomite $d^{p(1)}_{\mathbf{x},t}$, $d^{p(2)}_{\mathbf{x},t}$ and $d^{p(3)}_{\mathbf{x},t}$ respectively, $d^{p(i)}_{\mathbf{x},t} \in [0, 1]$, $d^{p(1)}_{\mathbf{x},t} + d^{p(2)}_{\mathbf{x},t} + d^{p(3)}_{\mathbf{x},t} = 1$. Let the proportions in node $(\mathbf{x}, t)$ be the vector $\mathbf{d}^p_{\mathbf{x},t}$ with elements $\{d^{p(i)}_{\mathbf{x},t}; i \in \{1, 2, 3\}\}$. Denote the complete set of proportions $\mathbf{d}^p : \{\mathbf{d}^p_{\mathbf{x},t}; (\mathbf{x}, t) \in \mathcal{L}_\mathcal{D}\}$.

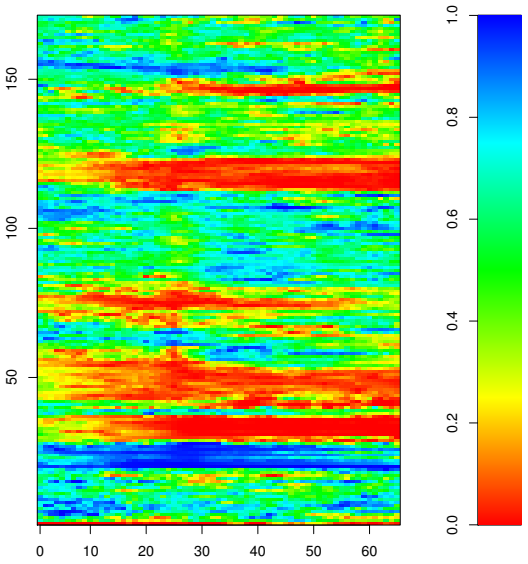
The depositional property in node $(\mathbf{x}, t)$ is denoted $d^d_{\mathbf{x},t}$ with $d^d_{\mathbf{x},t} \in \{GS, FS\}$. Denote the complete set of depositional properties $\mathbf{d}^d : \{d^d_{\mathbf{x},t}; (\mathbf{x}, t) \in \mathcal{L}_\mathcal{D}\}$.

The term $p(\cdot)$ is a general term for probability and can denote univariate probability, multivariate probability, probability mass functions (pmf) for discrete variables and probability density functions (pdf) for continuous variables.

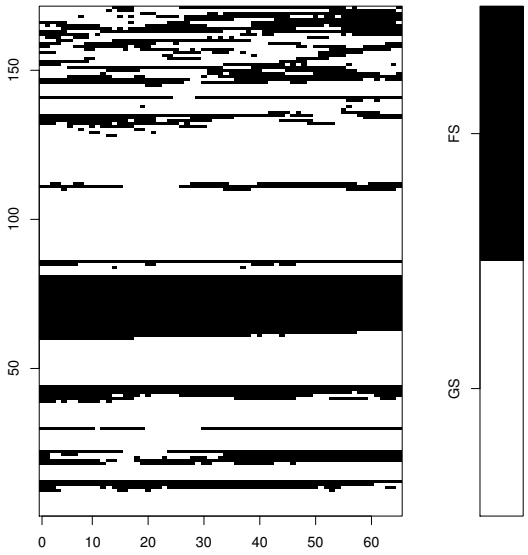


(a) Anhydrite

(b) Calcite



(c) Dolomite



(d) Depositional properties

Figure 1: (a)-(c) shows the proportions of carbonate rocks in the target zone, while (d) shows the depositional properties

3 Theory

In the following we will present some general theory on the topics the gamma distribution, the beta distribution, the multinomial distribution, the Dirichlet distribution, the generalized Dirichlet distribution, Markov chains, hidden Markov models, Markov random fields and Markov Chain Monte Carlo (MCMC) with a Metropolis-Hastings (MH) implementation. The degree of detail varies with what is needed for model development later in the thesis.

3.1 The Gamma Distribution

The gamma distribution, see Gamerman & Lopes (2006), is denoted Gamma(α, β) and has pdf

$$p(x;\alpha,\beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & , x > 0 \\ 0 & , otherwise \end{cases}$$

with $\Gamma(\cdot)$ being the gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt. \tag{1}$$

Sampling from Gamma$(\alpha, 1)$ can be done by considering $\alpha \in (0,1)$, $\alpha = 1$ and $\alpha > 1$ separately. Before presenting the simulation algorithm, Algorithm 3.1, we need to define some quantities.

If α is between 0 and 1 we use rejection sampling and define a function $q(x)$:

$$q(x) = \begin{cases} cx^{\alpha-1} & , x \in (0,1) \\ ce^{-x} & , x \geq 1 \\ 0 & , otherwise \end{cases},$$

where c is a normalizing constant. Observe that $\frac{p(x)}{q(x)} \leq d = \frac{1}{\Gamma(\alpha)c}$. Calculate acceptance probability k:

$$k = \frac{1}{d}\frac{p(x)}{q(x)} = \begin{cases} e^{-x} & , x \in (0,1) \\ x^{\alpha-1} & , x \geq 1 \\ 0 & , otherwise \end{cases}.$$

If α is equal to 1, the pdf

$$p(x;1,1) = \begin{cases} e^{-x} & ,x > 0 \\ 0 & ,otherwise \end{cases}$$

is the same as the pdf of the exponential distribution which is simple to sample from.

If α is greater than 1 we use ratio of uniforms sampling and need to define region

$$C_p = \left\{ (x_1, x_2) | 0 \leq x_1 \leq \sqrt{p^* \left(\frac{x_2}{x_1} \right)} \right\}$$

where

$$p^*(x) = \begin{cases} x^{\alpha-1} e^{-x} & ,x > 0 \\ 0 & ,otherwise \end{cases}.$$

Calculate constants a, b_- and b_+:

$$\begin{aligned} a &= \sqrt{\sup_x p^*(x)} = \sqrt{(\alpha-1)^{\alpha-1} e^{-(\alpha-1)}} \\ b_+ &= \sqrt{\sup_{x \geq 0} x^2 p^*(x)} = 0 \\ b_- &= -\sqrt{\sup_{x \leq 0} x^2 p^*(x)} = \sqrt{(\alpha+1)^{\alpha+1} e^{-(\alpha+1)}} \end{aligned}$$

so that $C_p \subset [0, a] \times [b_-, b_+]$.

Algorithm 3.1 is the algorithm for sampling from the gamma distribution, and it is defined from the expressions above.

Algorithm 3.1 Sampling from the gamma distribution

initiate : α
if $\alpha \in (0,1)$ **then**
  initiate : $p(x)$, $q(x)$, d
  repeat
    $x \sim q(x)$
    calculate $k = \frac{1}{d}\frac{p(x)}{q(x)}$
    $u \sim U[0,1]$
  until $u \leq k$
  return x
end if
if $\alpha = 1$ **then**
  $x \sim \text{Exp}(1)$
  return x
end if
if $\alpha > 1$ **then**
  initiate : C_p, a, b_-, b_+
  repeat
    $x_1 \sim U[0,a]$
    $x_2 \sim U[b_-, b_+]$
  until $(x_1, x_2) \in C_p$
  return $\frac{x_2}{x_1}$
end if

3.2 The Beta Distribution

The beta distribution with parameters $(a,b) > 0$, see Ross (2007), is denoted $\text{Beta}(a,b)$ and has pdf

$$p(x;a,b) = \begin{cases} \frac{1}{\text{B}(a,b)} x^{a-1}(1-x)^{b-1} & ,0 \leq x \leq 1 \\ 0 & ,otherwise \end{cases}.$$

The normalizing constant is defined by $\text{B}(\cdot,\cdot)$, which is the beta function with two parameters defined as

$$\text{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \tag{2}$$

with $\Gamma(\cdot)$ being the gamma function defined in Expression 1.

The beta function is also defined for a parameter vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)$:

$$\text{B}(\boldsymbol{\alpha}) = \frac{\prod_i^n \Gamma(\alpha_i)}{\Gamma(\sum_i^n \alpha_i)}. \tag{3}$$

3.3 The Multinomial Distribution

The k-dimensional multinomial distribution with parameter vector $\mathbf{x} > 0$, see Walpole, Myers, Myers & Ye (2007), is denoted $\text{Mult}_k(\mathbf{x})$ and has pdf

$$p(\boldsymbol{\beta};\mathbf{x}) = \begin{cases} \binom{n}{\beta_1,\beta_2,...,\beta_k} x_1^{\beta_1} x_2^{\beta_2} \cdots x_k^{\beta_k} & \begin{array}{l} x_1,...,x_k > 0 \\ x_1 + \cdots + x_k = 1 \\ \beta_1,...,\beta_k \geq 0 \\ \beta_1 + \cdots + \beta_k = n \end{array} \\ 0 & ,otherwise \end{cases},$$

where $\boldsymbol{\beta}$ is a vector of integers.

3.4 The Dirichlet Distribution

The k-dimensional Dirichlet distribution with parameter vector $\boldsymbol{\alpha} > 0$, see Connor & Mosimann (1969), is denoted $\text{Dir}_k(\boldsymbol{\alpha})$ and has pdf

$$p(\mathbf{x};\boldsymbol{\alpha}) = \begin{cases} \frac{1}{\text{B}(\boldsymbol{\alpha})} \prod_{i=1}^k x_i^{\alpha_i - 1} & \begin{array}{l} x_1,...,x_k > 0 \\ x_1 + \cdots + x_k = 1 \end{array} \\ 0 & ,otherwise \end{cases} \tag{4}$$

where $\text{B}(\boldsymbol{\alpha})$ is the beta function defined in Expression 3.

If the random vector $\mathbf{x} = (x_1,...,x_k)$ is Dirichlet distributed, the expected value of random variable x_i is

$$\text{E}(x_i) = \frac{\alpha_i}{\alpha_0}$$

where $\alpha_0 = \sum_i \alpha_i$.

The Dirichlet distribution is conjugate to the multinomial distribution. Assume $\mathbf{x} = (x_1,...,x_k)$ is a vector of random variables and that we are given the vector $\boldsymbol{\beta} = (\beta_1,...,\beta_k)$. We are interested in the distribution of $\mathbf{x}$ given $\boldsymbol{\beta}$, namely $p(\mathbf{x}|\boldsymbol{\beta})$, denoted the posterior distribution

$$p(\mathbf{x}|\boldsymbol{\beta}) = const \times p(\boldsymbol{\beta}|\mathbf{x})p(\mathbf{x})$$

where $p(\boldsymbol{\beta}|\mathbf{x})$ is the likelihood function, $p(\mathbf{x})$ is the prior distribution and the constant is a normalizing constant. Further assume that the prior is Dirichlet distributed, $\mathbf{x} \sim \text{Dir}_k(\boldsymbol{\alpha})$ and that the likelihood has a multinomial distribution $[\boldsymbol{\beta}|\mathbf{x}] \sim \text{Mult}_k(\mathbf{x})$. The posterior distribution then becomes

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\beta}) &= const \times p(\boldsymbol{\beta}|\mathbf{x})p(\mathbf{x}) \\ &= const \times \prod_{i=1}^{k} x_i^{\alpha_i - 1} x_i^{\beta_i} \\ &= const \times \prod_{i=1}^{k} x_i^{\alpha_i + \beta_i - 1} \\ &\sim \text{Dir}_k(\boldsymbol{\alpha} + \boldsymbol{\beta}), \end{aligned}$$

hence the posterior distribution is also Dirichlet distributed.

The parameters of the Dirichlet distribution can be estimated through a maximum likelihood method as presented in Wicker, Muller, Kalathur & Poch (2008). Given a data sample $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_N)$, $\mathbf{x}_i = (x_{i1}, ..., x_{ik})$, the maximum likelihood approach aims to maximize the log-likelihood function given by

$$l(\alpha_1, ..., \alpha_k|\mathbf{x}_1, ..., \mathbf{x}_N) = N \ln \Gamma(\alpha_0) - N \sum_{j=1}^{k} \ln \Gamma(\alpha_j) + \sum_{j=1}^{k} (\alpha_j - 1) \sum_{i=1}^{N} \ln x_{ij}. \tag{5}$$

Let $\mathbf{H} = \{h_{ij}; i, j = 1, ..., k\}$ denote the $k \times k$ Hessian of $l(\alpha_1, ..., \alpha_k; x_1, ..., x_k)$ and define a vector $\mathbf{g} = (g_1, ..., g_k)$ with elements given by

$$\begin{aligned} g_j &= \frac{\partial}{\partial \alpha_j} l(\alpha_1, ..., \alpha_k|x_1, ..., x_k) \\ &= N\Psi(\alpha_0) - N\Psi(\alpha_j) + \sum_{i=1}^{N} \ln x_{ij} \end{aligned}$$

where $\Psi(\cdot)$ is the digamma function defined by

$$\Psi(x) = \frac{d}{dx} \ln \Gamma(x).$$

Expression 5 is maximized by the Newton-Raphson method which gives the iteration

$$\boldsymbol{\alpha}^{n+1} = \boldsymbol{\alpha}^n + \mathbf{H}^{-1}\mathbf{g} \tag{6}$$

with elements h_{ij}^{-1} of the inverse Hessian given by

$$h_{ij}^{-1} = \frac{1}{N\Psi_1(\alpha_i)}\delta_{ij} + \frac{1}{N\Psi_1(\alpha_i)\Psi_1(\alpha_j)}\frac{\Psi_1(\alpha_0)}{1 - \Psi_1(\alpha_0)\sum_{l=1}^{k}\frac{1}{\Psi_1(\alpha_l)}}$$

where δ_{ij} is the Kronecker delta and $\Psi_1(\cdot)$ is the trigamma function defined by

$$\Psi_1(x) = \frac{d^2}{dx^2}\ln\Gamma(x).$$

An efficient rate of convergence of the iterations can be achieved by choosing an appropriate starting value. Wicker et al. (2008) presents a method for estimating the starting value called maximum likelihood approximation. Define the constant f_j by

$$f_j = \frac{\alpha_j}{\alpha_0}.$$

Estimate the starting value of f_j, f_j^0 by

$$f_j^0 = \frac{1}{N}\sum_{i=1}^{N} x_{ij}.$$

By maximum likelihood approximation the starting value of α_0, α_0^0 is given by

$$\alpha_0^0 = \frac{N(k-1)\gamma}{N\sum_{j=1}^{k} f_j \ln f_j - \sum_{j=1}^{k} f_j \sum_{i=1}^{N} \ln x_{ij}}$$

where the constant γ is defined as $\gamma = -\Psi(1)$. Further, the starting value of $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^0$ is a vector with elements given by

$$\alpha_j^0 = \alpha_0^0 f_j^0. \tag{7}$$

Sampling from the Dirichlet distribution is simple by using samples from the gamma distribution with $\beta = 1$, see Algorithm 3.2.

Algorithm 3.2 Sampling from the Dirichlet distribution

1: **initiate** : $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)$
2: **for** $i = 1, ..., k$ **do**
3: $\quad y_i \sim \text{Gamma}(\alpha_i, 1)$
4: **end for**
5: **for** $i = 1, ..., k$ **do**
6: $\quad x_i = \frac{y_i}{\sum_{j=1}^{k} y_j}$
7: **end for**
8: **return** $(x_1, ..., x_k)$

3.5 The Generalized Dirichlet Distribution

The generalized Dirichlet distribution has a more general covariance structure than the Dirichlet distribution, often making it more applicable, see Connor & Mosimann (1969).

Consider a random vector $\mathbf{x} = (x_1, ..., x_k) > 0$ where the random variables sum to unity. Define a new random vector $\mathbf{z} = (z_1, ..., z_{k-1})$ where the elements are given by $z_1 = x_1$, $z_j = x_j/(1 - x_1 - \cdots x_{j-1})$; $j = 2, ..., k-1$. Assume the z_j; $j = 1, ..., k-1$ are mutually independent and beta distributed with parameters α_j and β_j, $z_j \sim \text{Beta}(\alpha_j, \beta_j)$; $j = 1, ..., k-1$.

Since the z_j; $j = 1, ..., k-1$ are mutually independent their joint distribution is easily calculated, and by simple transformation of variables the pdf of the k-dimensional generalized Dirichlet distribution with parameter vectors $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{k-1}) > 0$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_{k-1}) > 0$, denoted $\text{GDir}_k(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is derived:

$$p(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} \prod_{i=1}^{k-1} \frac{1}{\text{B}(\alpha_i, \beta_i)} x_i^{\alpha_i - 1} (1 - \sum_{j=1}^{i} x_j)^{\gamma_i} & , \quad x_1, ..., x_k > 0 \\ & \quad x_1 + \cdots + x_k = 1 \\ 0 & , otherwise \end{cases}$$

where

$$\gamma_i = \begin{cases} \beta_i - \alpha_{i+1} - \beta_{i+1} & , i = 1, ..., k-2 \\ \beta_i - 1 & , i = k-1 \end{cases}$$

and $\text{B}(\cdot, \cdot)$ is the beta function defined in Expression 2.

The parameters of the generalized Dirichlet distribution can be estimated from a data sample $\mathbf{x} = (\mathbf{x}_1, ..., \mathbf{x}_N)$, $\mathbf{x}_i = (x_{i1}, ..., x_{ik})$ by defining a new sample $\mathbf{z} = (\mathbf{z}_1, ..., \mathbf{z}_N)$, $\mathbf{z}_i = (z_{i1}, ..., z_{i(k-1)})$ where $z_{i1} = x_{i1}$ and $z_{ij} = x_{ij}/(1 - x_{i1} - \cdots - x_{i(j-1)})$; $j = 2, ..., k-1$. For each $j = 1, ..., k-1$,

define data sample $\mathbf{z}^j = (z_{1j}, ..., z_{Nj})$ and assume these to be observations of the random variable z^j, $z^j \sim \text{Beta}(\alpha_j, \beta_j)$. The beta distribution is a special case of the Dirichlet distribution, $\text{Beta}(\alpha_j, \beta_j) = \text{Dir}_2(\alpha_j, \beta_j)$, hence the parameters α_j and β_j can be estimated by the iteration presented in Expression 6.

Wong (1998) describes an algorithm for generating samples from the generalized Dirichlet distribution, this is presented in Algorithm 3.3.

Algorithm 3.3 Sampling from the generalized Dirichlet distribution

initiate :
  $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_{k-1})$
  $\boldsymbol{\beta} = (\beta_1, ..., \beta_{k-1})$
$x_1 \sim \text{Beta}(\alpha_1, \beta_1)$
for $i = 2, ..., k-1$ **do**
  $z_i \sim \text{Beta}(\alpha_i, \beta_i)$
  $x_i = z_i \left(1 - \sum_{j=1}^{i-1} x_j\right)$
end for
$x_k = 1 - \sum_{j=1}^{k-1} x_j$
return $(x_1, ..., x_k)$

3.6 Markov Chains

A Markov chain, see Ross (2007), is a discrete time stochastic process $\boldsymbol{\pi} : \{\pi_t; t \in \mathcal{L}_{\mathcal{D}}^t\}$, $\pi_t \in \Omega$, which fulfils the Markov property

$$p(\pi_t|\pi_{t-1}, ..., \pi_1) = p(\pi_t|\pi_{t-1}).$$

A joint probability can always be expressed as a product of conditional probabilities

$$p(\boldsymbol{\pi}) = p(\pi_T|\pi_{T-1}, ..., \pi_1) \times \cdots \times p(\pi_2|\pi_1)p(\pi_1),$$

and due to the Markov property, the Markov chain can be written

$$\begin{aligned} p(\boldsymbol{\pi}) &= p(\pi_T|\pi_{T-1}) \times \cdots \times p(\pi_2|\pi_1)p(\pi_1) \\ &= \prod_t p(\pi_t|\pi_{t-1}) \end{aligned}$$

where $p(\pi_1|\pi_0) = p(\pi_1)$ for notational convenience. From this it follows that the Markov chain is fully described by the distribution $p(\pi_1)$ and the transition probabilities $p(\pi_t|\pi_{t-1})$. The

transition probabilities for a fixed t are arranged into a transition matrix

$$\mathbf{P}^t = \{p(\pi_t|\pi_{t-1}); \pi_t, \pi_{t-1} \in \Omega\},$$

where the options for π_{t-1} are represented by the rows of the matrix and the options for π_t are represented by the columns of the matrix. The number of elements of the matrix is the number of possible classes squared, and the elements of a row must sum to one:

$$\sum_{\pi_t} p(\pi_t|\pi_{t-1}) = 1.$$

A Markov chain is homogeneous if $\mathbf{P}^t = \mathbf{P}$ for all transitions in the chain, otherwise the Markov chain is inhomogeneous.

If it is possible to access any state in Ω in a finite number of steps if initially in any given state in Ω, the Markov chain is irreducible.

The period of a state is the greatest common divisor of the number of steps in all the ways of returning to a state if initially in that given state. If the period of the state is one, the state is said to be aperiodic. If all the states in Ω are aperiodic, the Markov chain is aperiodic.

A state is recurrent if starting in a specific state the chain will with probability one return to that state, and it is positive recurrent if it returns in a finite amount of time. All recurrent states in a Markov chain with a finite state space are positive recurrent. The Markov chain is positive recurrent if all the states of the Markov chain are positive recurrent.

A homogeneous Markov chain has a unique limiting distribution $\mathbf{p} : \{p(\pi); \pi \in \Omega\}$ if the chain is irreducible, aperiodic and positive recurrent. If so, the limiting distribution

$$p(\pi) = \lim_{k\to\infty} p(\pi_k|\pi_1); \quad \forall \pi \in \Omega \tag{8}$$

can be calculated by

$$\begin{array}{l} \mathbf{p} = \mathbf{P}^T\mathbf{p} \\ \sum_\pi p(\pi) = 1 \end{array}. \tag{9}$$

3.7 Hidden Markov Models

Consider the Markov chain $\boldsymbol{\pi} : \{\pi_t; t \in \mathcal{L}_{\mathcal{D}}^t\}$ with $\pi_t \in \Omega$. Assume that the state π_t is not visible, but an observation o_t depending on π_t is available. Denote the set of observations $\mathbf{o} : \{o_t; t \in \mathcal{L}_{\mathcal{D}}^t\}$. This process is a hidden Markov model, see Scott (2002). The dependency structure is illustrated in Figure 2.

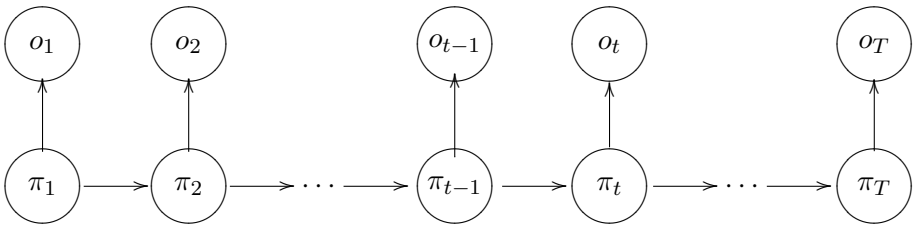


Figure 2: Illustration of the hidden Markov model

We are interested in exploring the Markov process given the observations. In a Bayesian setting this is known as the posterior distribution denoted $p(\boldsymbol{\pi}|\mathbf{o})$. Using Bayes rule the posterior distribution can be written

$$p(\boldsymbol{\pi}|\mathbf{o}) = const \times p(\mathbf{o}|\boldsymbol{\pi})p(\boldsymbol{\pi}),$$

where $p(\mathbf{o}|\boldsymbol{\pi})$ is the likelihood function, $p(\boldsymbol{\pi})$ is the prior distribution and the constant is a normalizing constant which can be difficult to calculate.

Based on the dependency structure the likelihood function factorizes and can be written

$$p(\mathbf{o}|\boldsymbol{\pi}) = \prod_t p(o_t|\pi_t).$$

The prior distribution is a Markov chain and hence can be written

$$p(\boldsymbol{\pi}) = \prod_t p(\pi_t|\pi_{t-1})$$

where $p(\pi_1|\pi_0) = p(\pi_1)$ for notational convenience.

The posterior model of interest is

$$p(\boldsymbol{\pi}|\mathbf{o}) = const \times \prod_t p(o_t|\pi_t)p(\pi_t|\pi_{t-1}).$$

It can be shown that the posterior model also follows a Markov chain, hence it can be written

$$p(\boldsymbol{\pi}|\mathbf{o}) = \prod_t p(\pi_t|\pi_{t-1}, \mathbf{o}), \tag{10}$$

where the factors $p(\pi_t|\pi_{t-1}, \mathbf{o})$ need to be determined.

The posterior distribution can be assessed by the Forward-Backward Algorithm.

Define forward probabilities

$$p_f(\pi_1, ..., \pi_t) = p(\pi_1, ..., \pi_t|o_1, ..., o_t)$$

and backward probabilities

$$p_b(\pi_t, ..., \pi_T) = p(\pi_t, ..., \pi_T|o_1, ..., o_T).$$

The forward probabilities are calculated recursively, initiated by calculating

$$p_f(\pi_1) = p(\pi_1|o_1) = const \times p(d_1|\pi_1)p(\pi_1)$$

where the constant is a normalizing constant given by

$$\sum_{\pi_1} p_f(\pi_1) = 1.$$

For $t = 2, ..., T$ the joint forward probabilities $p_f(\pi_{t-1}, \pi_t)$ are calculated by

$$\begin{aligned} p_f(\pi_{t-1}, \pi_t) &= p(\pi_{t-1}, \pi_t|o_1, ..., o_t) \\ &= const \times p(o_t|\pi_t)p(\pi_t|\pi_{t-1})p(\pi_{t-1}|o_1, ..., o_{t-1}) \\ &= const \times p(o_t|\pi_t)p(\pi_t|\pi_{t-1})p_f(\pi_{t-1}) \end{aligned}$$

where the constant can be calculated by

$$\sum_{\pi_{t-1}} \sum_{\pi_t} p_f(\pi_{t-1}, \pi_t) = 1.$$

Further the marginal forward probability $p_f(\pi_t)$ can be calculated by the expression

$$p_f(\pi_t) = \sum_{\pi_{t-1}} p_f(\pi_{t-1}, \pi_t).$$

Calculating the backward probabilities is initiated by noting that

$$p_b(\pi_T) = p_f(\pi_T).$$

For $t = T, ..., 2$, calculate the backward probabilities $p_b(\pi_{t-1}|\pi_t)$ by

$$\begin{aligned} p_b(\pi_{t-1}|\pi_t) &= p(\pi_{t-1}|\pi_t, o_1, ..., o_T) \\ &= p(\pi_{t-1}|\pi_t, o_1, ..., o_t) \\ &= \frac{p(\pi_{t-1}, \pi_t|o_1, ..., o_t)}{p(\pi_t|o_1, ..., o_t)} \\ &= \frac{p_f(\pi_{t-1}, \pi_t)}{p_f(\pi_t)} \end{aligned}$$

i.e. the backward probabilities can be calculated by the forward probabilities.

The marginal backward probability $p_b(\pi_{t-1})$ is calculated by

$$p_b(\pi_{t-1}) = \sum_{\pi_t} p_b(\pi_{t-1}|\pi_t) p_b(\pi_t).$$

Algorithm 3.4 Forward-Backward Algorithm

initiate
  $p_f(\pi_1) = const \times p(o_1|\pi_1)p(\pi_1)$
  $const \rightarrow \sum_{\pi_1} p_f(\pi_1) = 1$
for $t = 2$ to T **do**
  $p_f(\pi_{t-1}, \pi_t) = const \times p(o_1|\pi_1)p(\pi_t|\pi_{t-1})$
  $const \rightarrow \sum_{\pi_{t-1}} \sum_{\pi_t} p_f(\pi_{t-1}, \pi_t) = 1$
  $p_f(\pi_t) = \sum_{\pi_{t-1}} p_f(\pi_{t-1}, \pi_t)$
end for
$p_b(\pi_T) = p_f(\pi_T)$
for $t = T$ to 2 **do**
  $p_b(\pi_{t-1}|\pi_t) = \frac{p_f(\pi_{t-1}, \pi_t)}{p_f(\pi_t)}$
  $p_b(\pi_{t-1}) = \sum_{\pi_t} p_b(\pi_{t-1}|\pi_t) p_b(\pi_t)$
end for

Algorithm 3.4 is the Forward-Backward Algorithm.

From Expression 10 we see that we need the probability $p_b(\pi_t|\pi_{t-1})$, and not $p_b(\pi_{t-1}|\pi_t)$ as

is calculated by the backward recursion. However, the Markov chain is easily reversed by

$$\begin{aligned} p(\pi_t|\pi_{t-1}, \mathbf{o}) &= p_b(\pi_t|\pi_{t-1}) \\ &= \frac{p_b(\pi_t, \pi_{t-1})}{p_b(\pi_{t-1})} \\ &= \frac{p_b(\pi_{t-1}|\pi_t)p_b(\pi_t)}{p_b(\pi_{t-1})}, \end{aligned}$$

hence $p_b(\pi_t|\pi_{t-1})$ is expressed by the already calculated backward probabilities.

We can now easily simulate from the posterior distribution $p(\boldsymbol{\pi}|\mathbf{o})$, Algorithm 3.5 creates a realisation $\pi^s : \{\pi_t^s; t \in \mathcal{L}_{\mathcal{D}}^t\} \sim p(\boldsymbol{\pi}|\mathbf{o})$.

Algorithm 3.5 Simulation Algorithm

1: **initiate**
2: $\quad \pi_1^s \sim p_b(\pi_1)$
3: **for** $t = 2$ to T **do**
4: $\quad \pi_t^s \sim p_b(\pi_t|\pi_{t-1}^s)$
5: **end for**

A location wise MAP estimate $\hat{\boldsymbol{\pi}}$ is given by

$$\hat{\boldsymbol{\pi}} : \left\{ \arg\max_{\pi_t} \{p_b(\pi_t)\}; t \in \mathcal{L}_{\mathcal{D}}^t \right\} \tag{11}$$

3.8 Markov Random Fields

Consider a 2D random field $\mathcal{D}^2$, let the discretisation of $\mathcal{D}^2$ be by the lattice $\mathcal{L}_{\mathcal{D}}^{\mathbf{x}}$. The state of node $\mathbf{x}$ is denoted $\pi_{\mathbf{x}}$, $\pi_{\mathbf{x}} \in \Omega$. The complete set of states is denoted $\boldsymbol{\pi} : \{\pi_{\mathbf{x}}; \mathbf{x} \in \mathcal{L}_{\mathcal{D}}^{\mathbf{x}}\}$. Define $\boldsymbol{\pi}_{-\mathbf{x}} : \{\pi_{\mathbf{y}}; \mathbf{y} \in \mathcal{L}_{\mathcal{D}}^{\mathbf{x}}, \mathbf{y} \neq \mathbf{x}\}$ as the set of states excluding the state in node $\mathbf{x}$.

Each node $\mathbf{x}$ has a set of neighbours known as the neighbourhood around $\mathbf{x}$, denote this neighbourhood $\delta(\mathbf{x})$. If the neighbourhood of $\mathbf{x}$ consists of the four closest nodes to $\mathbf{x}$, the neighbourhood is denoted a first order neighbourhood. A Markov random field is a random field where a state in node $\mathbf{x}$ given all other states in the random field only depends on the states of the nodes in the neighbourhood of $\mathbf{x}$, hence the full conditional distribution can be written

$$p(\pi_{\mathbf{x}}|\pi_{-\mathbf{x}}) = p(\pi_{\mathbf{x}}|\pi_{\mathbf{y}}; \mathbf{y} \in \delta(\mathbf{x})); \quad \forall \mathbf{x} \in \mathcal{L}_{\mathcal{D}}^{\mathbf{x}}.$$

By the Hammersley-Clifford theorem the Markov random field is fully specified by the con-

ditional distributions, see Besag (1974).

3.9 Profile Markov Random Fields

Now consider the 3D random field $\mathcal{D}^3$ discretized by the lattice $\mathcal{L}_\mathcal{D}$ divided into vertical profiles and lateral horizons. Denote a node in the lattice $(\mathbf{x}, t)$ and the state in this node $\pi_{\mathbf{x},t}$. The profiles are discretized by the lattice $\mathcal{L}^t_\mathcal{D}$ and the lateral horizons are discretized by the lattice $\mathcal{L}^x_\mathcal{D}$.

Similarly to the Markov random field, a profile Markov random field is defined by

$$p(\boldsymbol{\pi}_\mathbf{x}|\boldsymbol{\pi}_{-\mathbf{x}}) = p(\boldsymbol{\pi}_\mathbf{x}|\boldsymbol{\pi}_{-\mathbf{y}}; \mathbf{y} \in \delta(\mathbf{x})); \quad \forall \mathbf{x} \in \mathcal{L}^\mathbf{x}_\mathcal{D}$$

where $\boldsymbol{\pi}_\mathbf{x} : \{\pi_{\mathbf{x},t}; t \in \mathcal{L}^t_\mathcal{D}\}$ is a profile for arbitrary $\mathbf{x}$, $\boldsymbol{\pi}_{-\mathbf{x}} : \{\pi_\mathbf{y}; \mathbf{y} \in \mathcal{L}^\mathbf{x}_\mathcal{D}, \mathbf{y} \neq \mathbf{x}\}$ and $\delta(\mathbf{x})$ is a profile neighbourhood around $\mathbf{x}$. If the profile neighbourhood consists of the four closet profiles to profile $\mathbf{x}$, the neighbourhood is denoted a first order profile neighbourhood.

The profile Markov random field is fully specified by the conditional distributions.

3.10 Markov Chain Monte Carlo

Consider a pdf $p(x)$, $x \in \Omega$, on the form

$$p(x) = const \times h(x) \tag{12}$$

where $h(x)$ is a non-negative function and the constant is difficult to calculate. Markov Chain Monte Carlo (MCMC) is method for creating realisations from $p(x)$ by constructing a Markov chain with limiting distribution $p(x)$ followed by simulating the Markov chain. The transition probabilities of a Markov chain are not uniquely defined by the limiting distribution, hence there are several ways of constructing the Markov chain. Hastings (1970) presents a method for constructing the Markov chain and the implementation is known as the Metropolis-Hastings (MH) algorithm.

Suggest a proposal distribution $q(y|x)$ summing to unity for fixed x, and accept this distribution with acceptance probability $\alpha(y|x)$. The transition probabilities of the Markov chain

are

$$\begin{aligned} p(y|x) &= q(y|x)\alpha(y|x) && , y \neq x \\ p(x|x) &= 1 - \textstyle\sum_{y \neq x} q(y|x)\alpha(y|x) && , otherwise \end{aligned} \tag{13}$$

A sufficient condition for the existence of the limiting distribution defined in Expression 8 is the detailed balance given by

$$p(x)p(y|x) = p(y)p(x|y); \quad \forall x, y \in \Omega. \tag{14}$$

Insert Expression 13 into Expression 14 which gives

$$p(x)q(y|x)\alpha(y|x) = p(y)q(x|y)\alpha(x|y) \tag{15}$$

Define expression $r(y|x)$ as

$$r(y|x) = p(x)q(y|x)\alpha(y|x),$$

and due to Expression 15 we have

$$r(y|x) = r(x|y).$$

Since $\alpha(\cdot|\cdot)$ is an acceptance probability it cannot be larger than one, hence

$$\begin{aligned} \alpha(y|x) = \tfrac{r(y|x)}{p(x)q(y|x)} \leq 1 \quad &\Rightarrow r(y|x) \leq p(x)q(y|x) \\ \alpha(x|y) = \tfrac{r(x|y)}{p(y)q(x|y)} \leq 1 \quad &\Rightarrow r(x|y) \leq p(y)q(x|y) \end{aligned} .$$

In Hastings (1970) it is shown that $r(\cdot|\cdot)$ should be as large as possible, thus

$$r(y|x) = \min\left\{p(x)q(y|x), p(y)q(x|y)\right\}$$

which gives the acceptance probability

$$\begin{aligned} \alpha(y|x) &= \tfrac{1}{p(x)q(y|x)} \min\left\{p(x)q(y|x), p(y)q(x|y)\right\} \\ &= \min\left\{1, \tfrac{p(y)q(x|y)}{p(x)q(y|x)}\right\} \end{aligned} . \tag{16}$$

The constant in Expression 12 which was difficult to calculate disappears from the expression for the acceptance probability as we divide $p(y)/p(x)$, and does not appear anywhere else in the MH-method.

Assume $\pi_{t-1} = x$ in the Markov chain. Suggest $\pi_t = y$, where y is sampled from the proposal distribution with pdf $q(y|x)$. Accept $\pi_t = y$ with the acceptance probability from Expression 16. If y is not accepted the Markov chain remains in x.

We have not specified an initial distribution for the Markov chain, and common practice is to choose a random starting point and iterate until the Markov chain forgets its origin, i.e. until it converges. The time until convergence is denoted the burn-in period, and we only use samples produced after the burn-in period as realisations from $p(x)$.

Algorithm 3.6 is the MH algorithm.

Algorithm 3.6 The MH Algorithm

initiate :
  x
repeat
  propose $y \sim q(y|x)$
  compute $\alpha(y|x)$
  draw $y \sim U[0, 1]$
  if $u \leq \alpha(y|x)$ **then**
    $x = y$
  else
    $x = x$
  end if
until sufficient amount of samples produced

4 Stochastic Model in 1D

The problem to be addressed is to predict diagenetic classes, $\boldsymbol{\pi}$, given proportions of carbonate rocks, $\mathbf{d}^p$, and depositional properties, $\mathbf{d}^d$. The problem is solved in a Bayesian setting where the complete solution is the posterior model expressed as

$$p(\boldsymbol{\pi}|\mathbf{d}^p, \mathbf{d}^d) = const \times p(\mathbf{d}^p|\boldsymbol{\pi}, \mathbf{d}^d)p(\boldsymbol{\pi}|\mathbf{d}^d)$$

where the expression $p(\mathbf{d}^p|\boldsymbol{\pi}, \mathbf{d}^d)$ is the likelihood model, the expression $p(\boldsymbol{\pi}|\mathbf{d}^d)$ is the prior model and the constant is a normalizing constant which computation can be very computer demanding.

The model will initially be developed in a 1D-setting for a profile discretized by the grid $\mathcal{L}^t_{\mathcal{D}}$, followed by an expansion into 3D. In the 1D-setting the point $\mathbf{x}$ is omitted from the notation, hence a node will be denoted t, the diagenetic class in that node π_t and so on.

4.1 Likelihood Model in 1D

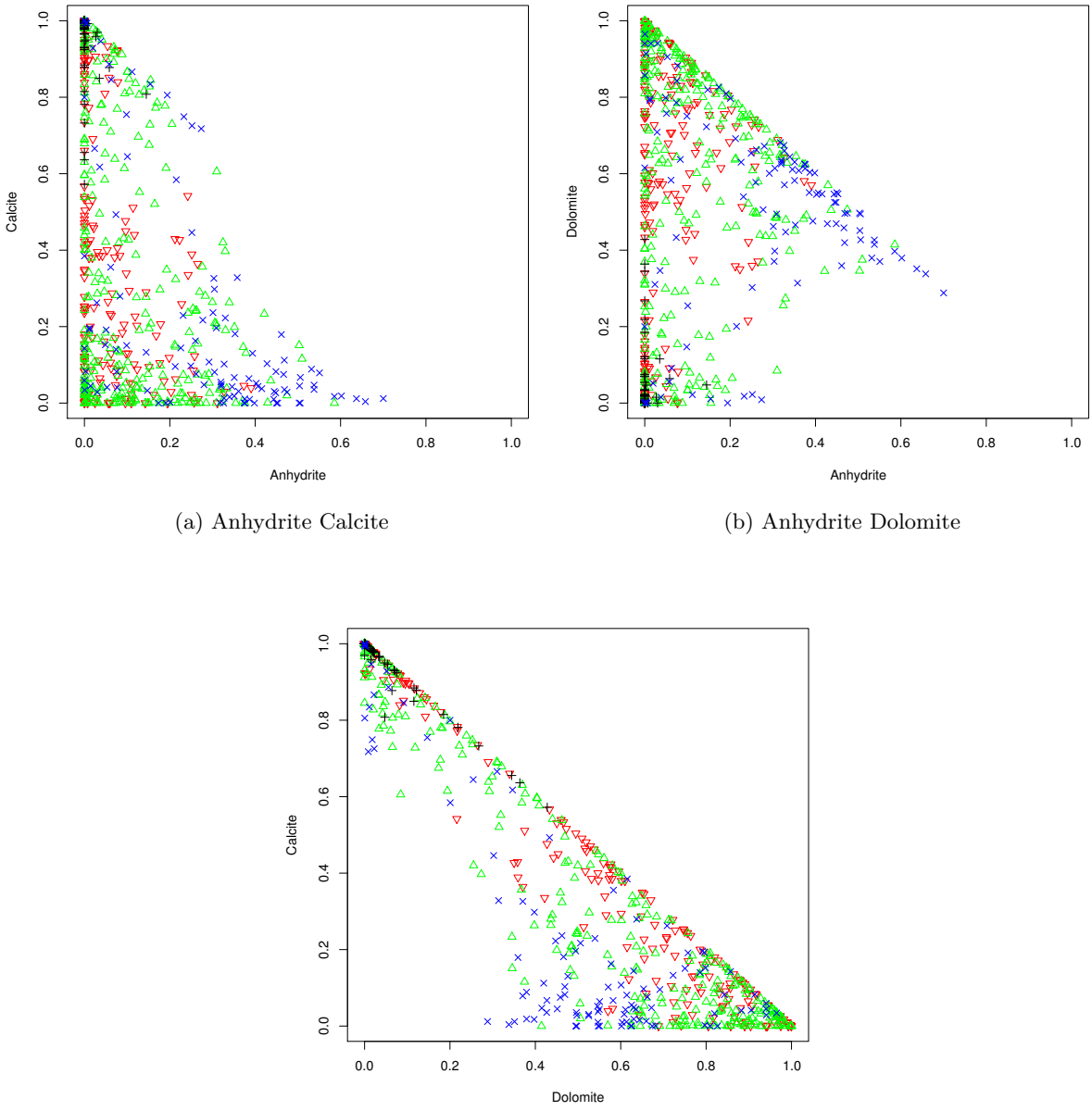
The likelihood model describes the likelihood of the proportions of carbonate rocks given the diagenetic classes and depositional properties. To build the model we will use data from the wells.

For the likelihood model we assume independence between the nodes such that it can be written in product form:

$$p(\mathbf{d}^p|\boldsymbol{\pi}, \mathbf{d}^d) = \prod_t p(\mathbf{d}^p_t|\boldsymbol{\pi}, \mathbf{d}^d) = \prod_t p(\mathbf{d}^p_t|\pi_t, d^d_t).$$

The distribution in each node is estimated from the observations. The proportions of carbonate rocks are real numbers on the interval between zero and one which sum to one.

Figure 3 and Figure 4 show plots of the observations with proportions on the axes. The colours represent the four diagenetic classes, where red represents good, green represents moderate, black represents oomoldic and poor is represented by blue. There will be one probability distribution for each given combination of diagenetic class and depositional property, eight in total. Give the parameters of the distributions superfixes π and d^d to clarify which diagenetic class and depositional property are given.

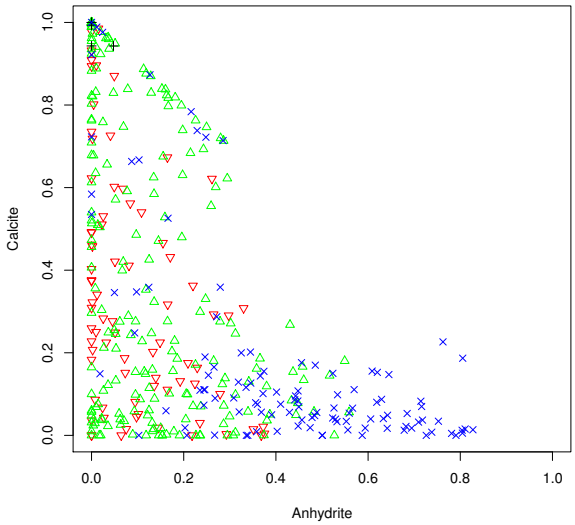


(a) Anhydrite Calcite

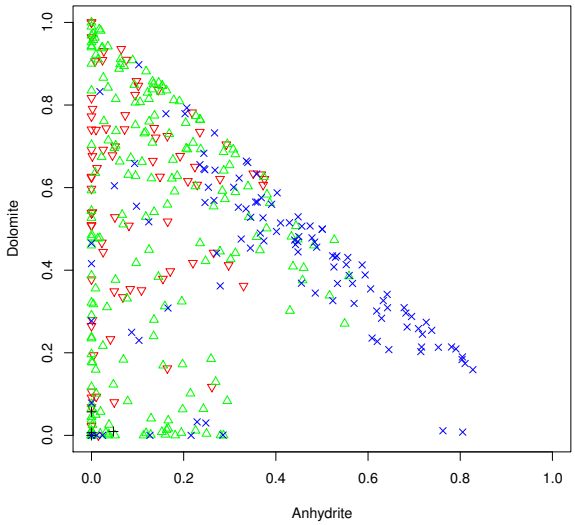
(b) Anhydrite Dolomite

(c) Dolomite Calcite

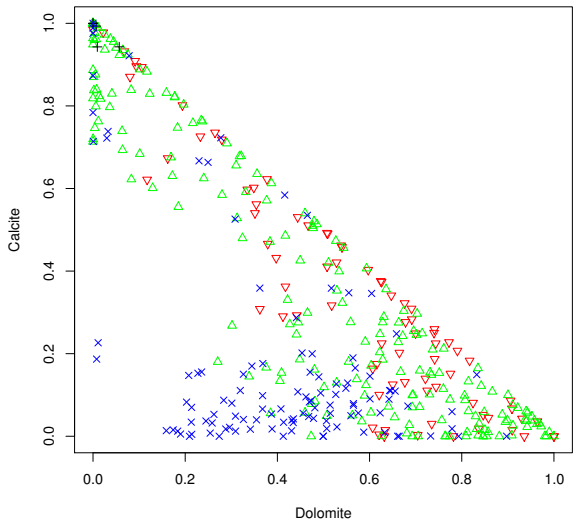
Figure 3: Plot of the observations in depositional property GS. The colours represent the diagenetic classes, G: red, M: green, O: black, P: blue



(a) Anhydrite Calcite



(b) Anhydrite Dolomite



(c) Dolomite Calcite

Figure 4: Plot of the observations in depositional property FS. The colours represent the diagenetic classes, G: red, M: green, O: black, P: blue

The proportions summing to unity could indicate Dirichlet distributed or generalized Dirichlet distributed observations, however, we allow zero proportions for which the mentioned distributions have probability zero. We choose to adjust the data such that a small value is added to the zero proportions and correspondingly subtracted from the larger proportions. The ramifications of this adjustment are considered minor due the uncertainty in the data.

If the Dirichlet distribution and the generalized Dirichlet distribution fit the data equally well, we will use the Dirichlet distribution as simpler distributions are preferable. To check the fit we estimate the parameters for the Dirichlet distribution and the generalized Dirichlet distribution from the observations for each combination of diagenetic class and depositional property, and compare the observations to samples drawn from these distributions.

The parameters of the Dirichlet distribution, $\boldsymbol{\alpha}^{\pi,d^d} = \left(\alpha_1^{\pi,d^d}, \alpha_2^{\pi,d^d}, \alpha_3^{\pi,d^d}\right)$, are estimated from the adjusted observations by the maximum likelihood approach using the iteration presented in Expression 6 with starting value calculated by Expression 7.

The estimated parameters of the Dirichlet distribution are:

$$\begin{aligned}
\boldsymbol{\alpha}^{G,GS} &= (0.13, 0.31, 0.35) \\
\boldsymbol{\alpha}^{M,GS} &= (0.16, 0.25, 0.34) \\
\boldsymbol{\alpha}^{O,GS} &= (0.12, 8.19, 0.16) \\
\boldsymbol{\alpha}^{P,GS} &= (0.38, 0.28, 0.52)
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\alpha}^{G,FS} &= (0.18, 0.41, 0.77) \\
\boldsymbol{\alpha}^{M,FS} &= (0.19, 0.33, 0.37) \\
\boldsymbol{\alpha}^{O,FS} &= (0.15, 16.14, 0.24) \\
\boldsymbol{\alpha}^{P,FS} &= (0.48, 0.31, 0.53)
\end{aligned}$$

The parameters of the generalized Dirichlet distribution,

$$\left(\boldsymbol{\alpha}^{\pi,d^d}, \boldsymbol{\beta}^{\pi,d^d}\right) = \left(\alpha_1^{\pi,d^d}, \alpha_2^{\pi,d^d}, \beta_1^{\pi,d^d}, \beta_2^{\pi,d^d}\right),$$

are estimated from the adjusted observations by the procedure described in Section 3.5.

The estimated parameters of the generalized Dirichlet distribution are:

$$\begin{aligned}
\left(\boldsymbol{\alpha}^{G,GS}, \boldsymbol{\beta}^{G,GS}\right) &= (0.18, 0.26, 3.34, 0.28) \\
\left(\boldsymbol{\alpha}^{M,GS}, \boldsymbol{\beta}^{M,GS}\right) &= (0.23, 0.21, 2.57, 0.26) \\
\left(\boldsymbol{\alpha}^{O,GS}, \boldsymbol{\beta}^{O,GS}\right) &= (0.14, 5.01, 42.50, 0.14) \\
\left(\boldsymbol{\alpha}^{P,GS}, \boldsymbol{\beta}^{P,GS}\right) &= (0.49, 0.23, 1.52, 0.40)
\end{aligned}$$

$$\begin{aligned}
\left(\boldsymbol{\alpha}^{G,FS}, \boldsymbol{\beta}^{G,FS}\right) &= (0.22, 0.36, 2.79, 0.63) \\
\left(\boldsymbol{\alpha}^{M,FS}, \boldsymbol{\beta}^{M,FS}\right) &= (0.25, 0.27, 2.09, 0.29) \\
\left(\boldsymbol{\alpha}^{M,FS}, \boldsymbol{\beta}^{M,FS}\right) &= (0.15, 16.70, 15.60, 0.25) \\
\left(\boldsymbol{\alpha}^{P,FS}, \boldsymbol{\beta}^{P,FS}\right) &= (0.53, 0.28, 1.03, 0.47)
\end{aligned}$$

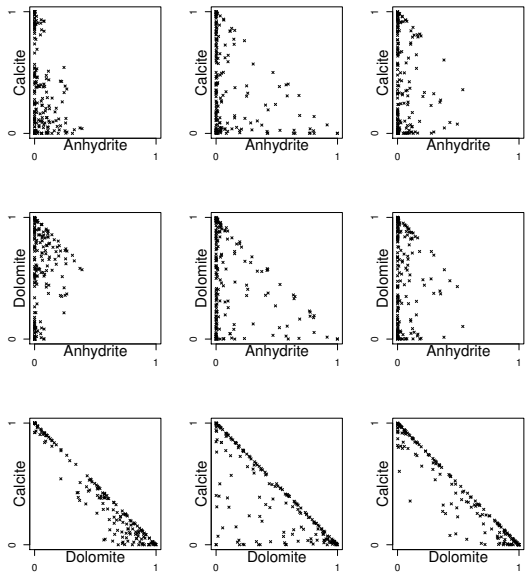
For each combination of diagenetic class and depositional property, samples are drawn from the estimated Dirichlet and generalized Dirichlet distribution by Algorithm 3.2 and Algorithm 3.3 respectively. The number of samples drawn are the same as the number of observations for the given diagenetic class and depositional property.

Figure 5 through Figure 12 are plots comparing the observed values to those sampled from the estimated distributions. For diagenetic class G, M and O the generalized Dirichlet distribution is clearly a better fit than the Dirichlet distribution, while for diagenetic class P neither distribution fits particularly well. Overall the generalized Dirichlet distribution provides the best fit, hence this is chosen for the likelihood model in a node:

$$p(\mathbf{d}_t^p|\pi_t, d_t^d) = \begin{cases} \frac{1}{\mathrm{B}(\alpha_1,\beta_1)}\frac{1}{\mathrm{B}(\alpha_2,\beta_2)} & \times {d_t^{p(1)}}^{\alpha_1-1}\left(1-d_t^{p(1)}\right)^{\beta_1-\alpha_2-\beta_2} \\ & \times {d_t^{p(2)}}^{\alpha_2-1}{d_t^{p(3)}}^{\beta_2-1} \end{cases} \begin{array}{l} , \\ \end{array} \begin{array}{l} d_t^{p(1)}, d_t^{p(2)}, d_t^{p(3)} > 0 \\ d_t^{p(1)} + d_t^{p(2)} + d_t^{p(3)} = 1 \end{array}$$

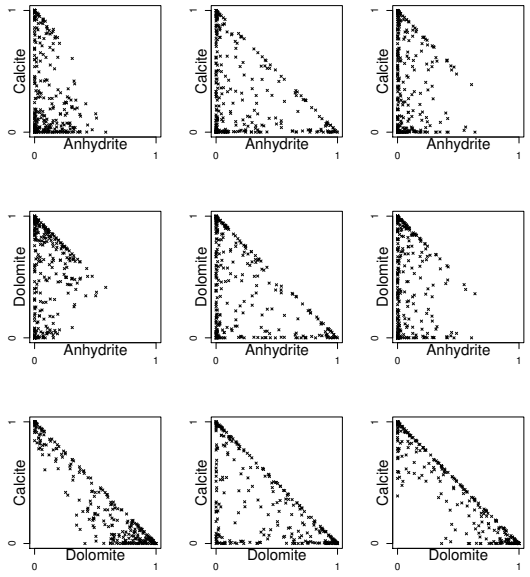
$$0 \qquad , otherwise$$

where the superfixes of the parameters are implied.



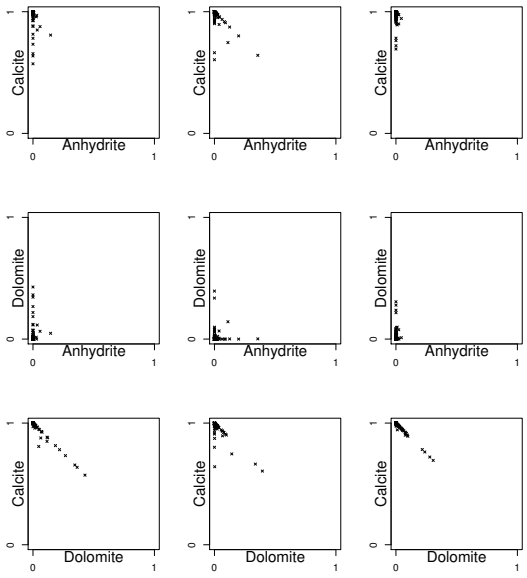
(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 5: Plot comparing distributions for the observations in diagenetic class G, depositional property GS



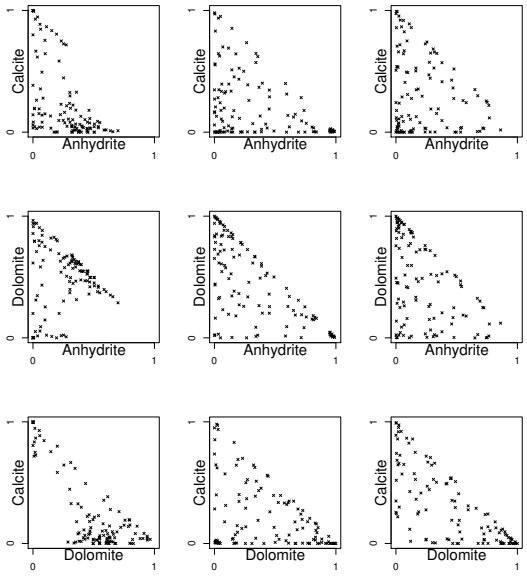
(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 6: Plot comparing distributions for the observations in diagenetic class M, depositional property GS



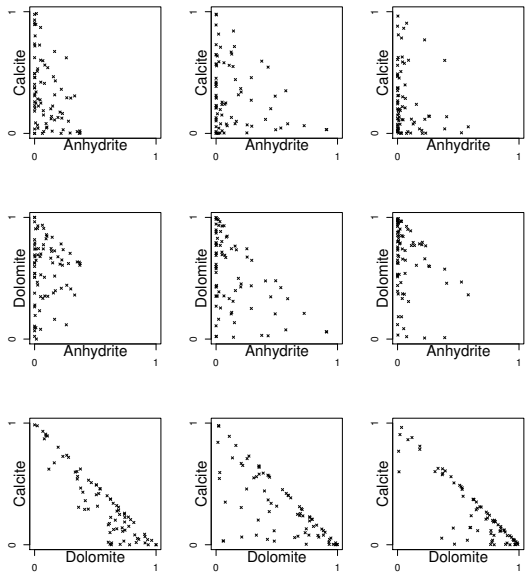
(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 7: Plot comparing distributions for the observations in diagenetic class O, depositional property GS



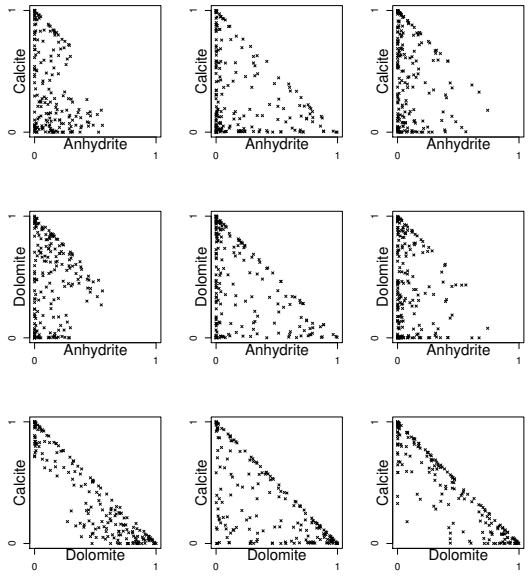
(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 8: Plot comparing distributions for the observations in diagenetic class P, depositional property GS



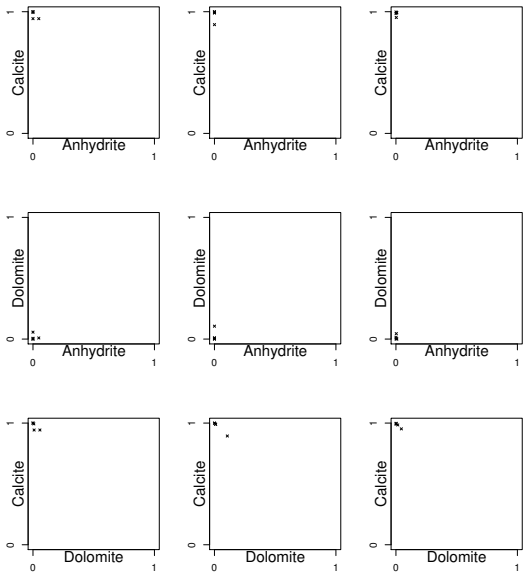
(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 9: Plot comparing distributions for the observations in diagenetic class G, depositional property FS



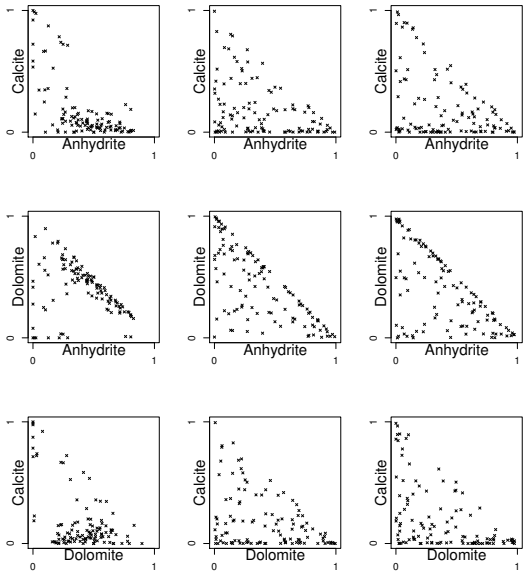
(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 10: Plot comparing distributions for the observations in diagenetic class M, depositional property FS



(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 11: Plot comparing distributions for the observations in diagenetic class O, depositional property FS



(a) Observations (b) Dirichlet (c) Generalized Dirichlet

Figure 12: Plot comparing distributions for the observations in diagenetic class P, depositional property FS

4.2 Prior Model in 1D

The prior model contains prior knowledge about the diagenetic classes, hence it does not rely on the observations of carbonate rock.

We define a Markov chain with state space $\Omega = \{G, M, O, P\}$ upwards through the vertical profile. This means that the prior model can be written

$$p(\boldsymbol{\pi}|\mathbf{d}^d) = \prod_t p(\pi_t|\pi_{t-1}, d_t^d, d_{t-1}^d)$$

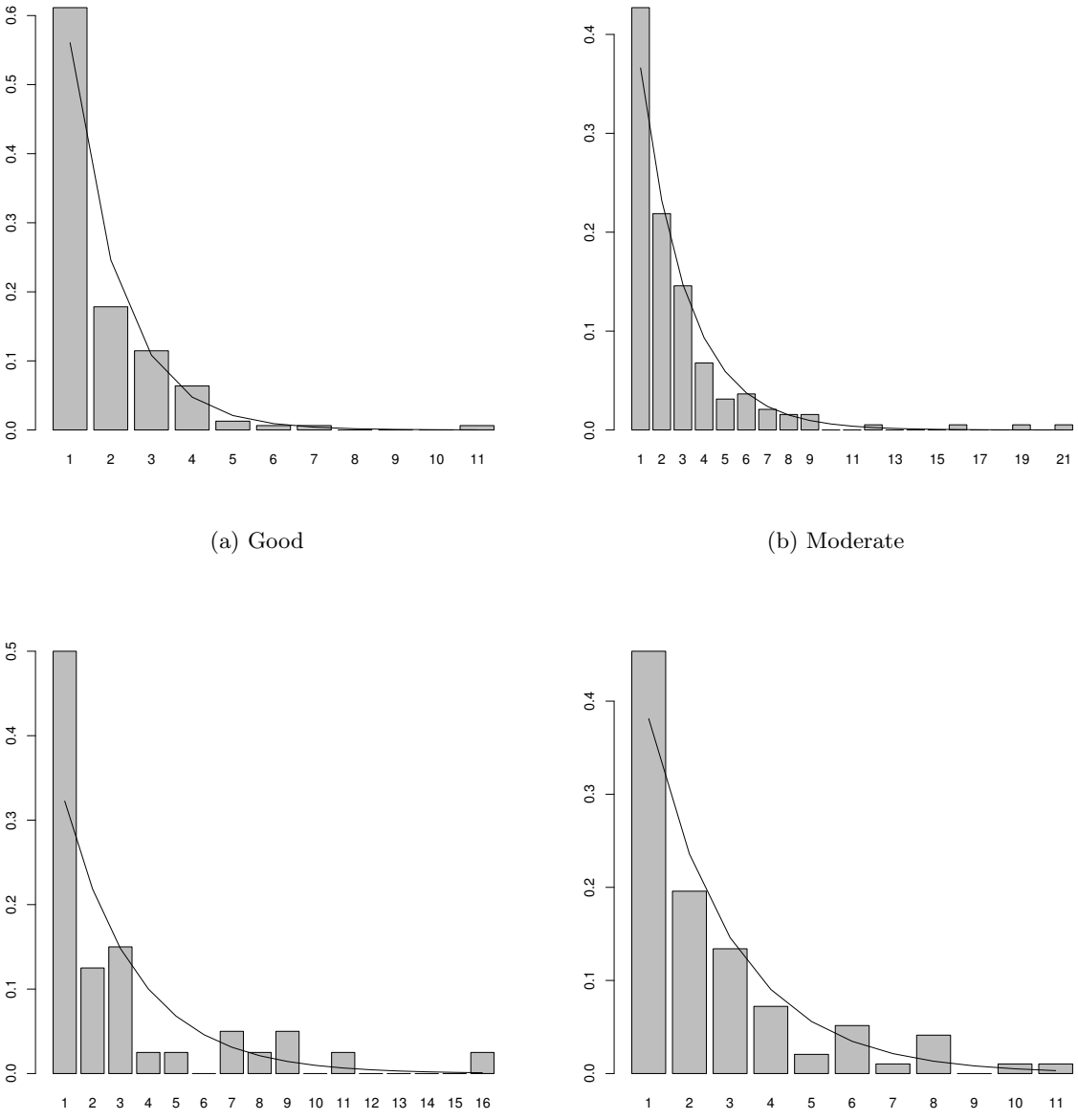
where $p(\pi_1|\pi_0, d_1^d, d_0^d) = p(\pi_1|d_1^d)$ for notational convenience.

To substantiate the use of a Markov chain as the prior model we look at the transitions between diagenetic classes from the well data. We create a general transition matrix by counting the number of transitions between diagenetic classes upwards, disregarding the depositional properties. The general transition matrix, denoted $\mathbf{P}$, becomes

$$\mathbf{P} = \begin{matrix} G \\ M \\ O \\ P \end{matrix} \begin{bmatrix} 0.44 & 0.39 & 0.06 & 0.12 \\ 0.21 & 0.63 & 0.04 & 0.11 \\ 0.15 & 0.15 & 0.68 & 0.02 \\ 0.12 & 0.25 & 0.01 & 0.62 \end{bmatrix},$$

where the element in row one, column one is denoted $p(G|G)$ and represents the probability of going from diagenetic class good in node $t-1$, $\pi_{t-1} = G$, to diagenetic class good in node t, $\pi_t = G$. The element in row one, column two is denoted $p(M|G)$ and so on. If the profile follows a Markov chain, the length of intervals of consecutive equal diagenetic classes should follow geometric distributions with parameters one minus the diagonal elements of the transition matrix, e.g. the probability of just one consecutive diagenetic class good is $1 - p(G|G) = 1 - 0.44 = 0.56$, while the probability of two consecutive goods is $(1 - p(G|G)) \times p(G|G) = (1 - 0.44) \times 0.44 = 0.25$. Hence, if we count the length of all intervals of consecutive equal goods we expect 56% of the intervals to have length 1 and 25% of the intervals to have length 2.

Figure 13 shows the length of the intervals of consecutive equal classes compared to the theoretical geometric distribution, and we see that the fit is very good which indicates a Markov prior model.



(a) Good

(b) Moderate

(c) Oomoldic

(d) Poor

Figure 13: Plot comparing length of intervals to the geometric distribution

The transition probabilities between node $t-1$ and node t can be expressed as a transition matrix $\mathbf{P}_{d_{t-1}^d, d_t^d}$ where the elements are the conditional pmfs $p(\pi_t|\pi_{t-1}, d_{t-1}^d, d_t^d)$. The transition matrices are dependent on t through the dependency on d_{t-1}^d and d_t^d, hence the Markov chain is inhomogeneous.

The transition matrices are different depending on the depositional properties in the nodes between which the transition occurs. There are four possible combinations of depositional properties, moving from GS to GS, from FS to FS, from FS to GS and from GS to FS.

We will consider two approaches to estimating the transition matrices. The first approach considers the transition probabilities a global property of the profile, hence the transition probabilities are independent of the value of t, though still dependent on t through the dependency on the depositional properties.

The second approach considers the transition probabilities a more local property, hence the vertical positions of two nodes affects the transition probabilities between them, hence the transition probabilities will depend on the value of t as well as the depositional properties. Figure 14 shows the diagenetic classes of wells in the reservoir. We see that the diagenetic classes are not distributed equally vertically through the reservoir, and the distribution is somewhat similar in the wells, e.g. the proportion of diagenetic class P is larger for $t > 70$ than $t < 70$ for all wells and the majority of diagenetic class O is located around $t \approx 50$. Geologically these observations are reasonable as the deposition process causes lateral geological layers, and properties may vary somewhat between the layers. Node t is within the same geological layer for all wells.

If the prior model were estimated from only one well the local approach could overfit the model, especially if there were no significant differences between distributions of diagenetic classes in different geological layers. Since there are observations from seven wells the cross well calculation ensures that the local approach essentially behaves as the global approach if vertical differences are not present.

The transition matrices will be estimated by the same method for both approaches. For the first approach, all well observations will be used for the estimation for any given t, while for the second approach only well observations in the interval $t \in [t-n, t+n]$ will be used. If the interval falls outside the observation region, the interval shrinks accordingly, hence if $t = 1$, the interval is reduced to $[1, 1+n]$.

Using the global estimation approach there will be four transition matrices in total, while for

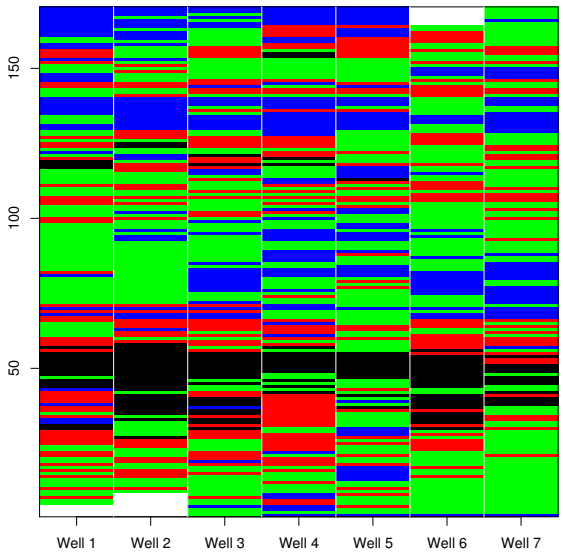


Figure 14: Diagenetic properties in the wells. G: red, M: green, O: black, P: blue, the white areas are undefined

the local approach there will be four transition matrices for each t. If both approaches perform equally well the global approach will be preferred since it is faster and simpler.

4.2.1 Estimating the Transition Matrices

The transition matrices are estimated by putting a Dirichlet prior on the rows of the transition matrix. The parameters of the Dirichlet prior are the number of each diagenetic class in the depositional property to which the chain is transitioning. Using the global approach these are counted using all well observations, using the local approach these are only counted in the specified interval. Denote the proportions of each diagenetic class a_G, a_M, a_O and a_P, hence the prior for a row is

$$p(\mathbf{p}_{\pi_{t-1}}) \sim \text{Dir}_4(a_G, a_M, a_O, a_P).$$

Define a multinomial likelihood for the row

$$p(n_G, n_M, n_0, n_P|\mathbf{p}_{\pi_{t-1}}) \sim \text{Mult}_4(\mathbf{p}_{\pi_{t-1}})$$

where $n_G^{\pi_{t-1}}$, $n_M^{\pi_{t-1}}$, $n_O^{\pi_{t-1}}$ and $n_P^{\pi_{t-1}}$ is the number of transitions to good, moderate, oomoldic and poor when transitioning between the given depositional properties, given that the previous

state is π_{t-1}. Again we only use the observations specified by the approach.

The Dirichlet distribution is the conjugate prior for the multinomial distribution, hence the posterior distribution is also Dirichlet:

$$p(\mathbf{p}_{\pi_{t-1}}|n_G^{\pi_{t-1}}, n_M^{\pi_{t-1}}, n_0^{\pi_{t-1}}, n_P^{\pi_{t-1}}) \sim \mathrm{Dir}_4(a_G + n_G^{\pi_{t-1}}, a_M + n_M^{\pi_{t-1}}, a_O + n_O^{\pi_{t-1}}, a_P + n_P^{\pi_{t-1}}).$$

To estimate the transition matrix we take the expected value of the Dirichlet distribution for each row, for each element this means the parameter divided by the sum of the parameters in the row. Hence the elements of the row become

$$\mathbf{p}_{\pi_{i-1}} = \frac{\begin{bmatrix} a_G + n_G^{\pi_{i-1}} & a_M + n_M^{\pi_{i-1}} & a_O + n_O^{\pi_{i-1}} & a_P + n_P^{\pi_{i-1}} \end{bmatrix}}{a_G + n_G^{\pi_{i-1}} + a_M + n_M^{\pi_{i-1}} + a_O + n_O^{\pi_{i-1}} + a_P + n_P^{\pi_{i-1}}}$$

The transition matrices using the global approach are estimated to:

$$\mathbf{P}_{GS,GS} = \begin{bmatrix} 0.31 & 0.40 & 0.15 & 0.15 \\ 0.26 & 0.46 & 0.14 & 0.14 \\ 0.24 & 0.37 & 0.24 & 0.14 \\ 0.25 & 0.40 & 0.14 & 0.21 \end{bmatrix}, \mathbf{P}_{FS,FS} = \begin{bmatrix} 0.24 & 0.48 & 0.01 & 0.27 \\ 0.18 & 0.58 & 0.01 & 0.23 \\ 0.21 & 0.50 & 0.01 & 0.28 \\ 0.20 & 0.47 & 0.01 & 0.32 \end{bmatrix}$$

$$\mathbf{P}_{FS,GS} = \begin{bmatrix} 0.27 & 0.41 & 0.16 & 0.16 \\ 0.26 & 0.42 & 0.15 & 0.17 \\ 0.26 & 0.41 & 0.16 & 0.16 \\ 0.26 & 0.41 & 0.16 & 0.18 \end{bmatrix}, \mathbf{P}_{GS,FS} = \begin{bmatrix} 0.22 & 0.50 & 0.01 & 0.28 \\ 0.21 & 0.50 & 0.01 & 0.28 \\ 0.21 & 0.50 & 0.01 & 0.28 \\ 0.20 & 0.48 & 0.01 & 0.31 \end{bmatrix}.$$

The local approach produces too many transition matrices to list all.

4.3 Posterior Model in 1D

The posterior model, uniquely defined by the likelihood and prior model, is given by

$$p(\boldsymbol{\pi}|\mathbf{d}^p,\mathbf{d}^d) = \begin{cases} const & \times \prod_t p(\mathbf{d}_t^p|\pi_t,d_t^d) & d_t^{p(1)},d_t^{p(2)},d_t^{p(3)}>0 \\ & \times \prod_t p(\pi_t|\pi_{t-1},d_t^d,d_{t-1}^d) & d_t^{p(1)}+d_t^{p(2)}+d_t^{p(3)}=1 \\ 0 & & ,otherwise \end{cases},$$

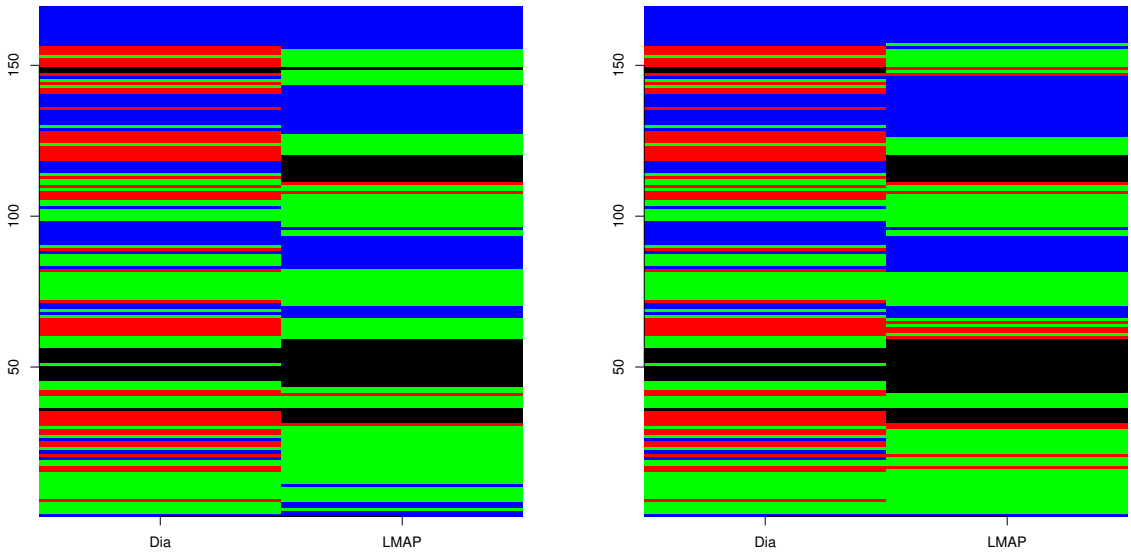
where calculation of the constant is very computer demanding since it is defined by $\sum_{\boldsymbol{\pi}} p(\boldsymbol{\pi}|\mathbf{d}^p,\mathbf{d}^p)$, i.e. the sum over all combinations of diagenetic classes in the profile.

Since the likelihood model factorizes and the prior model is a Markov chain, the posterior model is a hidden Markov chain and can be assessed by the Forward-Backward algorithm.

4.4 Results and Discussion

The model is tested on a well from which the observations were not used to develop the model, this is denoted the reference well and the diagenetic classes from the reference well are denoted π^R. The observed proportions, $\mathbf{d}^p$, and the observed depositional properties, $\mathbf{d}^d$, are used to calculate the posterior probability of the diagenetic classes, and these probabilities will be compared to the observed diagenetic classes π^R.

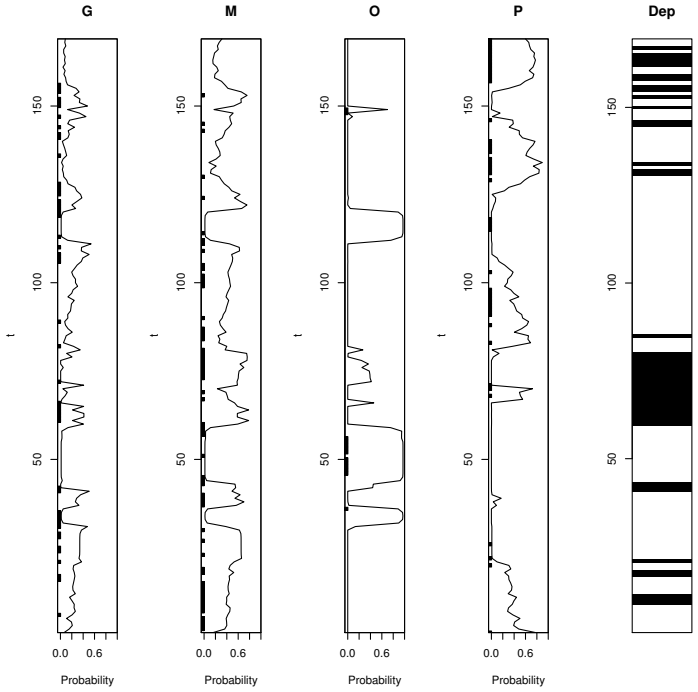
Figure 15 compares the location wise MAP estimate of the diagenetic classes defined in Expression 11 to the reference diagenetic classes for the global and local approach. We see that overall the location wise MAP appears smoother with longer class intervals than the reference classes as is expected since the MAP-operator tends to reduce heterogeneity. The two approaches produce similar location wise MAPs, but there are some noteworthy differences. The local approach captures more of the heterogeneity in the well than the global approach. In some cases the LMAP of the global approach matches the reference classes better than the local approach, e.g. at node $t \approx 150$ the LMAP of the global approach matches the correct O, while the LMAP of the local approach is M and at node $t \approx 45$ the global LMAP is the correct G, while the local LMAP predicts M. The LMAP of the local approach also matches some reference classes the LMAP of the global approach does not, e.g. the local LMAP matches G around $t \approx 150$, $t \approx 60$ and $t \approx 25$ and especially towards the bottom the local approach correctly predicts M where the global approach predicts P.



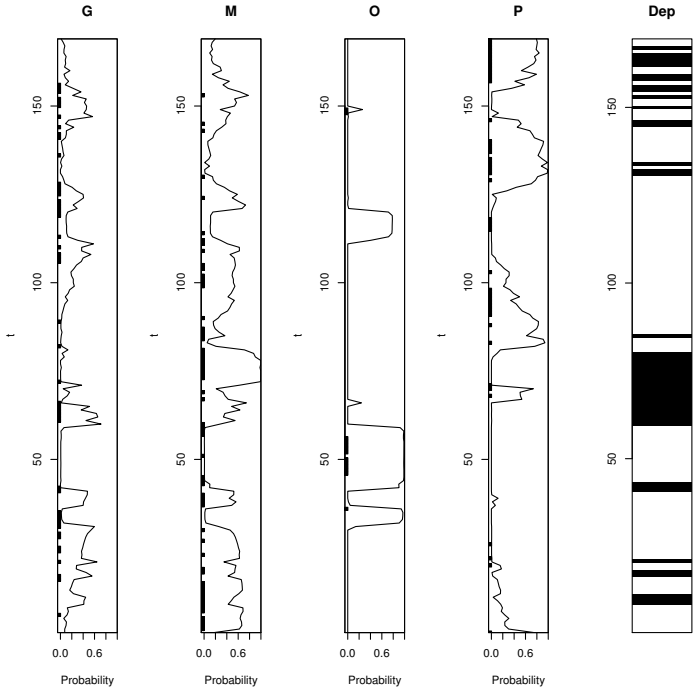
(a) Global (b) Local

Figure 15: Plot comparing the reference diagenetic class to the LMAP. G: red, M: green, O: black, P: blue

The LMAP only gives which class is most probable from the posterior distribution, but there is no information about whether the predicted diagenetic class in a node is significantly more probable than the other classes of if the difference is marginal. Figure 16 show the marginal posterior distribution, $p(\pi_t|\mathbf{d}^p, \mathbf{d}^d)$ for G, M, O and P compared with the reference classes marked on the left axis, as well as the depositional properties for the global and local estimation approach respectively. Again there are similarities between the two approaches, and at node $t \approx 45$ where the LMAP of the global approach matches the reference class the difference between the distributions is marginal. However, at node $t \approx 150$ the difference between the two approaches is significant for class O. Generally the local approach is better when is comes to the marginal posterior distribution of O, both have a large probability for O at node $t \approx 120$, but for the local approach the probability is smaller than the global approach, also at $t \approx 70$ the global approach has a relatively large probability for O, while this is correctly close to zero for the local approach. The marginal posterior probability for G at node $t \approx 60$ is significantly larger for the local approach than the global approach, and towards the bottom where there is almost no class P the global approach has a large marginal posterior probability for P while it is smaller for the local approach.



(a) Global



(b) Local

Figure 16: Plot showing the marginal posterior distribution, $p(\pi_t|\mathbf{d}^p, \mathbf{d}^d)$ for each diagenetic class, the reference diagenetic class is marked on the left axis. The plot to the right is the depositional property, GS: white, FS: black

	Global	Local
δ_G	0.25	0.29
δ_M	0.44	0.49
δ_O	0.87	0.85
δ_P	0.53	0.58
δ	0.44	0.49

Table 1: Probability of correct classification

Denote the number of diagenetic class π in a profile $N(\pi)$, where $N(G) + N(M) + N(O) + N(P) = T$. Define a probability of correct classification δ for the whole profile

$$\delta = \frac{1}{T} \sum_t p(\pi_t = \pi_t^R | \mathbf{d}^p, \mathbf{d}^d)$$

and for diagenetic class

$$\delta_\pi = \frac{1}{N(\pi)} \sum_{\pi_t^R = \pi} p(\pi_t = \pi_t^R | \mathbf{d}^p, \mathbf{d}^d).$$

Ideally the probability of correct classification δ and hence δ_π should equal unity, which would mean 100% correct classification. Table 1 shows the probability of correct classification for both the global and local approach. For both approaches we see that probability of classification is below average for G and above average for O. The local approach mostly outperforms the global approach except for O, this is probably due to the classification at point $t \approx 150$. Since the local approach mostly outperforms the global approach the former is chosen for the model.

Diagenetic class G is often misclassified as M as apparent from Figure 15b, but from Figure 16b we see that in many cases where this happens they have nearly the same marginal posterior probability. Around $t = 120$ diagenetic class O is very probable compared to the other classes, which is a clear misclassification. The reason for this is that O is only found where there is a majority of calcite while the other diagenetic classes are more widely distributed, hence the likelihood distribution of O is much steeper than the others. If there is a majority of calcite, as around $t = 120$ and also around $t = 50$, the probability of O is much larger than the other diagenetic classes since all of their distributions integrate to one.

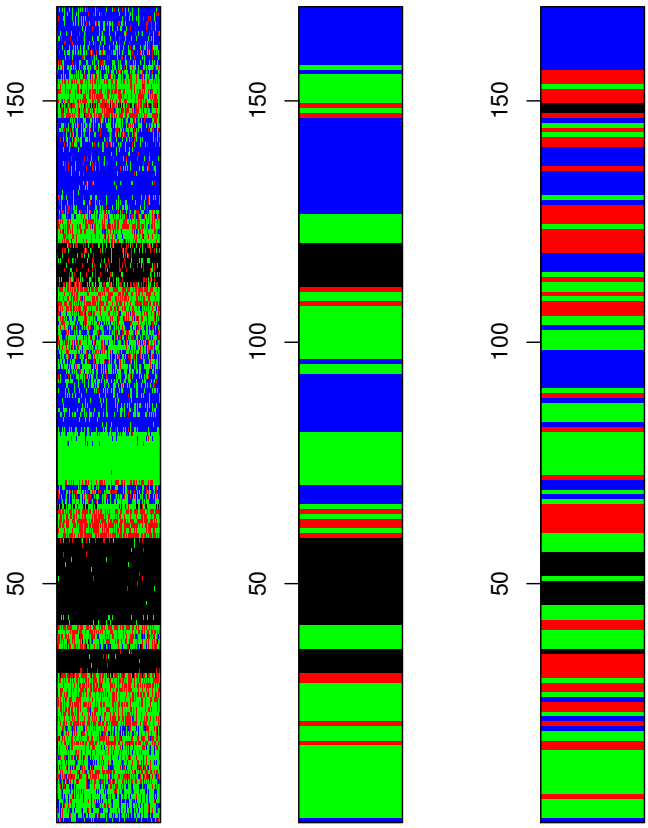
Figure 17 shows 200 independent simulations from the posterior distribution compared with the LMAP solution and the true diagenetic classes. The nodes where the probability of a certain diagenetic class is large appears evenly coloured, while nodes where the diagenetic class is more

π	Prop(π^R)	90% CI for Prop(π)
G	0.29	[0.14,0.23]
M	0.36	[0.31,0.43]
O	0.08	[0.15,0.19]
P	0.29	[0.22,0.31]

Table 2: Reference proportions of the diagenetic classes compared to 90% CI for estimated proportions of diagenetic classes

uncertain appear more blurred. We see that even though the heterogeneity of the well is lost in the LMAP solution it is present in the simulations.

The proportions of diagenetic class π seems to vary between the reference diagenetic classes of the well and the LMAP. Denote the proportion of diagenetic class π Prop(π) defined by Prop(π) $= N(\pi)/T$. We wish to check whether the true diagenetic proportions are within 90% confidence intervals (CI) of the proportions predicted by the posterior distribution. Create 90% CIs by producing 1000 realisations from the posterior distribution, calculating the proportions of diagenetic classes for each of the realisations. The 90% CI for a diagenetic class π is estimated by removing the 5% highest and 5% lowest simulated proportions. Table 2 shows the reference proportions of diagenetic classes compared to the 90% CIs. We see that for digenetic classes M and P the reference proportions are within the 90% CIs, while for classes G and O the reference proportions fall outside the 90% CI.



(a) Simulations (b) LMAP (c) Diagenetic class

Figure 17: (a) 200 independent realisations from $p(\boldsymbol{\pi}|\mathbf{d}^p, \mathbf{d}^d)$. (b) Location wise Maximum a Posteriori (c) True diagenetic classes. G: red, M: green, O: black, P: blue

5 Model Expansion

In the expansion of the model from 1D to 3D we still assume the likelihood model factorizes and can be written

$$p(\mathbf{d}^p|\boldsymbol{\pi}, \mathbf{d}^d) = \prod_{(\mathbf{x},t)} p(\mathbf{d}^p_{\mathbf{x},t}|\pi_{\mathbf{x},t}, d^d_{\mathbf{x},t})$$

where $p(\mathbf{d}^p_{\mathbf{x},t}|\pi_{\mathbf{x},t}, d^d_{\mathbf{x},t})$ is generalized Dirichlet distributed as before.

A priori we expect the diagenetic classes to be laterally continuous. To model the continuity we define the prior model to be a profile Markov random field defined by

$$p(\boldsymbol{\pi}_{\mathbf{x}}|\mathbf{d}^d_{\mathbf{x}}, \boldsymbol{\pi}_{-\mathbf{x}}) = p(\boldsymbol{\pi}_{\mathbf{x}}|\mathbf{d}^p_{\mathbf{x},t}, \boldsymbol{\pi}_{\mathbf{y}}; \mathbf{y} \in \delta(\mathbf{x})); \quad \forall \mathbf{x} \in \mathcal{L}^{\mathbf{x}}_{\mathcal{D}}$$

where all variables are as defined in Section 3.9 and $\delta(\mathbf{x})$ is specifically a first order profile neighbourhood around profile $\mathbf{x}$.

As previously stated, the profile Markov random field is fully specified by the full conditional distributions. Similarly to the prior model in the 1D-setting, let the full conditional distribution of the profile be modelled by an inhomogeneous Markov chain

$$p(\boldsymbol{\pi}_{\mathbf{x}}|\mathbf{d}^d_{\mathbf{x}}, \boldsymbol{\pi}_{\mathbf{y}}; \mathbf{y} \in \delta(\mathbf{x})) = \prod_t p(\pi_{\mathbf{x},t}|\pi_{\mathbf{x},t-1}, d^d_{\mathbf{x},t}, d^d_{\mathbf{x},t-1}, \pi_{\mathbf{y},t}; \mathbf{y} \in \delta(\mathbf{x})); \quad \forall \mathbf{x} \in \mathcal{L}^{\mathbf{x}}_{\mathcal{D}}$$

with $p(\pi_{\mathbf{x},1}|\pi_{\mathbf{x},0}, d^d_{\mathbf{x},1}, d^d_{\mathbf{x},0}, \pi_{\mathbf{y},t}; \mathbf{y} \in \delta(\mathbf{x})) = p(\pi_{\mathbf{x},1}|d^d_{\mathbf{x},1}, \pi_{\mathbf{y},t}; \mathbf{y} \in \delta(\mathbf{x}))$ for notational convenience.

The transition probabilities between node $(\mathbf{x}, t-1)$ and node $(\mathbf{x}, t)$ can be expressed as a transition matrix $\mathbf{P}_{d^d_{\mathbf{x},t-1}, d^d_{\mathbf{x},t}}(\pi_{\mathbf{y},t}; \mathbf{y} \in \delta(\mathbf{x}))$ where the elements are the conditional probability mass functions $p(\pi_{\mathbf{x},t}|\pi_{\mathbf{x},t-1}, d^d_{\mathbf{x},t}, d^d_{\mathbf{x},t-1}, \pi_{\mathbf{y},t}; \mathbf{y} \in \delta(\mathbf{x}))$.

5.1 Lateral Continuity

Several ways of incorporating lateral continuity have been suggested. Rimstad & Omre (2010) introduce a correction term $V(\pi_{\mathbf{x},t}, \delta(\mathbf{x}), \beta)$ by which the transition probabilities are multiplied, hence the expression for the transition probabilities becomes

$$p(\pi_{\mathbf{x},t}|\pi_{\mathbf{x},t-1}, d^d_{\mathbf{x},t}, d^d_{\mathbf{x},t-1}, \pi_{\mathbf{y},t}; \mathbf{y} \in \delta(\mathbf{x})) = const \times p(\pi_{\mathbf{x},t}|\pi_{\mathbf{x},t-1}, d^d_{\mathbf{x},t}, d^d_{\mathbf{x},t-1}) V(\pi_{\mathbf{x},t}, \delta(\mathbf{x}), \beta)$$

where the correction term is defined by

$$V(\pi_{\mathbf{x},t}, \delta(\mathbf{x}), \beta) = \exp\left\{\beta \sum_{\mathbf{y}\in\delta(\mathbf{x})} \mathrm{I}(\pi_{\mathbf{y},t} = \pi_{\mathbf{x},t})\right\}$$

where $\mathrm{I}(\cdot)$ is an indicator function and β is a coupling parameter.

Ulvmoen, Omre & Buland (2010) create basic transition matrices where transitions probabilities to classes in the neighbourhood are larger than transition probabilities to classes not in the neighbourhood. If the neighbours were $\{G, G, O, P\}$ the transition matrix could be

$$\mathbf{P} = \begin{matrix} G \\ M \\ O \\ P \end{matrix} \begin{bmatrix} 0.4998 & 0.0004 & 0.2499 & 0.2499 \\ 0.4998 & 0.0004 & 0.2499 & 0.2499 \\ 0.4998 & 0.0004 & 0.2499 & 0.2499 \\ 0.4998 & 0.0004 & 0.2499 & 0.2499 \end{bmatrix},$$

where there is a larger probability of transitioning to the neighbouring classes. The transition matrices are then adjusted such that their associated limiting distributions are consistent with the expected proportions of the classes in the reservoir.

We will use an approach similar to the latter approach, only reversed. Use the estimated matrices in Section 4.2 as suggested transition matrices given depositional properties $d^d_{\mathbf{x},t-1}$ and $d^d_{\mathbf{x},t}$. To incorporate lateral continuity into the prior model adjust the suggested transition matrix towards an associated limiting distribution with higher probability for neighbouring classes.

Methods for adjusting the transition matrices will be presented in the following section.

6 Adaption of Transition Matrices

This section will discuss adaption of transition matrices of a Markov chain to a given associated limiting distribution. The challenge is: Consider a reference proportion vector $\boldsymbol{\zeta}$ and a reference transition matrix $\mathbf{P}$ of a Markov chain with state space Ω, find a transition matrix $\mathbf{P}^*$ similar to $\mathbf{P}$, and the associated limiting distribution $\mathbf{p}^*$ of $\mathbf{P}^*$ similar to $\boldsymbol{\zeta}$. $\mathbf{P}$ has associated limiting distribution $\mathbf{p}$, Ω has n elements.

Define a difference function between matrices:

$$\delta_1(\mathbf{P}, \mathbf{P}^*) = \frac{1}{n^2} \sum_{y \in \Omega} \sum_{x \in \Omega} (p(y|x) - p^*(y|x))^2 \tag{17}$$

and vectors:

$$\delta_2(\boldsymbol{\zeta}, \mathbf{p}^*) = \frac{1}{n} \sum_{x \in \Omega} (\zeta(x) - p^*(x))^2 \tag{18}$$

where division by n^2 in Expression 17 and n in Expression 18 ensure that $\delta_1(\cdot, \cdot)$ and $\delta_2(\cdot, \cdot)$ represent average change in an element.

Define a loss function

$$L_\lambda(\mathbf{P}^*; \mathbf{P}, \boldsymbol{\zeta}) = \lambda \delta_1(\mathbf{P}, \mathbf{P}^*) + (1 - \lambda)\delta_2(\boldsymbol{\zeta}, \mathbf{p}^*)$$

which describes deviance from both $\mathbf{P}$ and $\boldsymbol{\zeta}$. The parameter $\lambda \in [0, 1]$ defines a trade-off between changes in $\mathbf{P}$ and $\boldsymbol{\zeta}$.

We will present two approaches aimed at finding the $\mathbf{P}^*$ which minimizes $L_\lambda(\mathbf{P}^*; \mathbf{P}, \boldsymbol{\zeta})$.

6.1 The Iteration Algorithm

The iteration algorithm is based on the adjustment presented in Ulvmoen et al. (2010). The objective is to minimize $L_0(\mathbf{P}^*; \mathbf{P}, \boldsymbol{\zeta})$ by adjusting $\mathbf{P}$ gradually until we reach a Markov chain with limiting distribution $\mathbf{p}^*$ close to $\boldsymbol{\zeta}$.

Define a set of adjustment factors

$$\mathbf{s} = \{s(x); x \in \Omega\}$$

and transition matrix

$$\mathbf{P}_{\mathbf{s}}^{*} = \left\{ \alpha_{\mathbf{s}}(x) p(y|x) \frac{s(y)}{p(y)}; x, y \in \Omega \right\} \tag{19}$$

where

$$\alpha_{\mathbf{s}}(x) = \left[\sum_{y \in \Omega} p(y|x) \frac{s(y)}{p(y)} \right]^{-1}.$$

Let the limiting distribution associated with $\mathbf{P}_{\mathbf{s}}^{*}$ be denoted $\mathbf{p}_{\mathbf{s}}^{*}$. We wish to find an $\mathbf{s}$ such that $\delta_2(\boldsymbol{\zeta}, \mathbf{p}_{\mathbf{s}}^{*}) = 0$. By looking at Expression 19 it is easy to see that if $s(x)$ grows larger, $p_{\mathbf{s}}^{*}(x)$ will grow larger, hence if $\zeta(x) > p_{\mathbf{s}}^{*}(x)$, we wish to increase the value of $s(x)$ and similarly if $\zeta(x) < p_{\mathbf{s}}^{*}(x)$ we wish to decrease the value of $s(x)$. Thus we want to adjust $s(x)$ in the direction of $\zeta(x) - p_{\mathbf{s}}^{*}(x)$, so to update $s(x)$ we add $k(\zeta(x) - p_{\mathbf{s}}^{*}(x))$, k a constant, to the old value:

$$\mathbf{s}^{(i+1)} = \mathbf{s}^{(i)} + k \left(\boldsymbol{\zeta} - \mathbf{p}_{\mathbf{s}^{(i)}}^{*} \right).$$

The constant k determines the size of the steps and should be chosen carefully. If k is too large the steps will be so large that the algorithm will never converge towards the desired value but keep jumping over it, while if k is too small the steps will be so small that convergence is very slow.

Algorithm 6.1 performs iterations until $\delta_2 \left(\boldsymbol{\zeta}, \mathbf{p}_{\mathbf{s}^{(i)}}^{*} \right) \leq \epsilon$, then $\mathbf{P}_{\mathbf{s}}^{*}$ will be a matrix with associated limiting distribution $\mathbf{p}_{\mathbf{s}}^{*}$ close to the reference proportions $\boldsymbol{\zeta}$.

Algorithm 6.1 Iteration Algorithm

1: **initiate** :
2: $\mathbf{s}^{(0)} = \boldsymbol{\zeta}, k, \epsilon$
3: **repeat**
4: $\mathbf{s}^{(i+1)} = \mathbf{s}^{(i)} + k \left(\boldsymbol{\zeta} - \mathbf{p}_{\mathbf{s}^{(i)}}^{*} \right)$
5: compute $\mathbf{P}_{\mathbf{s}^{(i+1)}}^{*} \rightarrow \mathbf{p}_{\mathbf{s}^{(i+1)}}^{*}$
6: compute $\delta_2 \left(\boldsymbol{\zeta}, \mathbf{p}_{\mathbf{s}^{(i)}}^{*} \right)$
7: **until** $\delta_2 \left(\boldsymbol{\zeta}, \mathbf{p}_{\mathbf{s}^{(i)}}^{*} \right) \leq \epsilon$

6.2 The 'Bayesian' Algorithm

View transition matrix $\mathbf{P}^{*}$ as a random matrix with pdf $p(\mathbf{P}^{*})$.

We want to estimate a transition matrix given the limiting distribution, i.e.

$$\mathrm{E}(\mathbf{P}^*|\boldsymbol{\zeta}). \tag{20}$$

The pdf of $[\mathbf{P}^*|\boldsymbol{\zeta}]$ is

$$p(\mathbf{P}^*|\boldsymbol{\zeta}) = const \times p(\boldsymbol{\zeta}|\mathbf{P}^*)p(\mathbf{P}^*). \tag{21}$$

Conditional expression $p(\boldsymbol{\zeta}|\mathbf{P}^*)$ is the likelihood function, linking $\boldsymbol{\zeta}$ to $\mathbf{P}^*$. The density $p(\mathbf{P}^*)$ is the prior model and should incorporate prior knowledge about $\mathbf{P}^*$.

In special cases $p(\mathbf{P}^*|\boldsymbol{\zeta})$ can be computed analytically, but in general evaluating the constant will be very computer damanding.

6.2.1 Prior model

For the prior model $p(\mathbf{P}^*)$, assign the rows of the transition matrix $\mathbf{P}^*$ independent Dirichlet distributions with parameters identical to the corresponding row of the reference matrix $\mathbf{P}$:

$$\mathbf{p}_x^* \sim \mathrm{Dir}_n(\mathbf{p}_x), \tag{22}$$

where $\mathbf{p}_x^* : \{p^*(y|x); y \in \Omega\}; \forall x \in \Omega$ and $\mathbf{p}_x : \{p(y|x); y \in \Omega\}; \forall x \in \Omega$. Hence the pdf is

$$p(\mathbf{P}^*) = \prod_i p(\mathbf{p}_x^*; \mathbf{p}_x),$$

where $p(\mathbf{p}_x^*; \mathbf{p}_x)$ is the Dirichlet pdf as defined in Expression 4.

A problem with this approach is that we may have $p(y|x) = 0$ or $p^*(y|x) = 0$, for which the pdf of the Dirichlet distribution is zero. However, we want 0 to be a possible value, so if $p(y|x) = 0$ or $p^*(y|x) = 0$, simply disregard this element and consider the rest of the row as Dirichlet distributed.

It is possible to introduce a scale factor $s_x > 1$ for each row to adjust the steepness of the distribution. Multiply the parameters of the Dirichlet distribution with the scale factor, hence the rows of the transition matrix $\mathbf{P}^*$ have distributions

$$\mathbf{p}_x^* \sim \mathrm{Dir}_n(s_x\mathbf{p}_x) \tag{23}$$

and the pdf would then become

$$p(\mathbf{P}^*) = \prod_i p(\mathbf{p}_x^*; s_x\mathbf{p}_x).$$

6.2.2 Likelihood model

The likelihood model $p(\boldsymbol{\zeta}|\mathbf{P}^*)$ links the reference proportions $\boldsymbol{\zeta}$ to the transition matrix $\mathbf{P}^*$.

Assume that $\boldsymbol{\zeta}$ is the expected value of the associated limiting distribution of $\mathbf{P}^*$, then Expression 9 can be expressed as a multivariate linear regression problem:

$$\mathbf{z} = \mathbf{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}$$

where $\mathbf{z}$ is a vector of length $(n+1)$:

$$\mathbf{z} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$\mathbf{Z}$ is a matrix of dimension $((n+1) \times n)$:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{P}^{*T} - \mathrm{I} \\ 1 \cdots 1 \end{bmatrix},$$

$\boldsymbol{\zeta}$ is the limiting distribution and the error term is normally distributed:

$$\boldsymbol{\varepsilon} \sim N_{n+1}(\mathbf{0}, \sigma^2\mathbf{I}).$$

The maximum likelihood estimator, see Johnson & Wichern (2007), of $\boldsymbol{\zeta}$ will in fact be the real associated limiting distribution $\mathbf{p}^*$, so

$$\mathbf{p}^* = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}\mathbf{y}.$$

Moreover, $\mathbf{p}^*$ is normally distributed:

$$\mathbf{p}^* \sim N_n(\boldsymbol{\zeta}, \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}), \tag{24}$$

or equally:

$$[\boldsymbol{\zeta}|\mathbf{P}^*] \sim N_n(\mathbf{p}^*, \sigma^2(\mathbf{Z}^T\mathbf{Z})^{-1}) = N_n(\mathbf{p}^*, \boldsymbol{\Sigma}). \tag{25}$$

The value of the standard deviation σ determines the deviance of the limiting distribution $\mathbf{p}^*$ from the reference proportions $\boldsymbol{\zeta}$, the smaller the value of σ the less deviance.

6.2.3 MH Implementation

Evaluating the constant in Expression 21 is very computer demanding, hence we will use the MH algorithm to estimate Expression 20.

Suggest a new value for $\mathbf{P}^*$, $\tilde{\mathbf{P}}^*$, from a freely chosen proposal distribution with pdf $q(\tilde{\mathbf{P}}^*|\mathbf{P}^*)$. The efficiency of the MH algorithm will depend on this choice. Common choises are independent proposal, where we suggest a new value for $\mathbf{P}^*$ independently of the current value, and random walk proposal, where we suggest the new value from a distribution which has the old value of $\mathbf{P}^*$ as its expected value.

Suggesting a new value for $\mathbf{P}^*$, $\tilde{\mathbf{P}}^*$, by independent proposal for each row:

$$[\tilde{\mathbf{p}}_x^*|\mathbf{p}_x^*] \sim \mathrm{Dir}_n(\boldsymbol{\zeta}).$$

Hence the proposal distribution has pdf

$$q(\tilde{\mathbf{P}}^*|\mathbf{P}^*) = q(\tilde{\mathbf{P}}^*) = \prod_{x\in\Omega} p(\tilde{\mathbf{p}}_x^*; \boldsymbol{\zeta}),$$

$p(\tilde{\mathbf{p}}^*; \mathbf{p}^*)$ as defined in Expression 4.

The acceptance probability from Expression 16 is:

$$\alpha(\tilde{\mathbf{P}}^*|\mathbf{P}^*) = \min\left\{1, \frac{p(\boldsymbol{\zeta}|\tilde{\mathbf{P}}^*)p(\tilde{\mathbf{P}}^*)q(\mathbf{P}^*|\tilde{\mathbf{P}}^*)}{p(\boldsymbol{\zeta}|\mathbf{P}^*)p(\mathbf{P}^*)q(\tilde{\mathbf{P}}^*|\mathbf{P}^*)}\right\}.$$

With the suggested proposal distribution this becomes:

$$\alpha(\tilde{\mathbf{P}}^*|\mathbf{P}^*) = \min\left\{1, \frac{\frac{1}{|\boldsymbol{\Sigma}_{\tilde{\mathbf{p}}^*}|^{1/2}} e^{-\frac{1}{2}(\tilde{\mathbf{p}}^*-\boldsymbol{\zeta})^T\boldsymbol{\Sigma}_{\tilde{\mathbf{p}}^*}^{-1}(\tilde{\mathbf{p}}^*-\boldsymbol{\zeta})} \prod_x \prod_y \tilde{p}^*(y|x)^{p(y|x)-1} \prod_x \prod_y p^*(y|x)^{\zeta(y)-1}}{\frac{1}{|\boldsymbol{\Sigma}_{\mathbf{p}^*}|^{1/2}} e^{-\frac{1}{2}(\mathbf{p}^*-\boldsymbol{\zeta})^T\boldsymbol{\Sigma}_{\mathbf{p}^*}^{-1}(\mathbf{p}^*-\boldsymbol{\zeta})} \prod_x \prod_y p^*(y|x)^{p(y|x)-1} \prod_x \prod_y \tilde{p}^*(y|x)^{\zeta(y)-1}}\right\}.$$

Use the average of the realisations produced after the burn-in period to estimate $\mathrm{E}(\mathbf{P}^*|\boldsymbol{\zeta})$.

6.3 Discussion

The iteration algorithm is a very fast algorithm which allows $\delta_2(\boldsymbol{\zeta}, \mathbf{p}^*)$ to become arbitrarily small, though we have no control over $\delta_1(\mathbf{P}, \mathbf{P}^*)$. The 'Bayesian' algorithm produces better results than the iteration algorithm, especially for large λ, but it is not very efficient since we have to obtain convergence in the MH algorithm for thereafter to acquire the appropriate number of samples to estimate $\mathrm{E}(\mathbf{P}^*|\boldsymbol{\zeta})$.

We will use the iteration algorithm for model implementation since the adjustment has to be performed for a large number of transition matrices. For the local approach to creating transition matrices there are four different transition matrices for every $t \in \mathcal{L}_{\mathcal{D}}^t$ and adjustment by the 'Bayesian' algorithm would be very time consuming. However, if the transition matrices were created by the global approach there would be four transition matrices in total, hence for a small number of different limiting distributions $\boldsymbol{\zeta}$ the 'Bayesian' algorithm could be used for the adjustment.

7 Model Implementation in 2D

Figure 1 shows the observations for which we will now estimate the diagenetic classes. The well we tested on in Section 4.3 is located at $x = 25$, i.e. the 25th column of the well.

There are two sets of observations of carbonate rock for the well, the nodes of the two sets are in the same location but are estimated with different support. The support of the well observations used to assess the model in Section 4 is the same as for the other wells, while the support of the second set of observations is the same as for the rest of the 2D target zone. To make a fair comparison of the 1D model and the 2D model both models will be applied to observations with the same support, i.e. the 1D model will be implemented with observations from the 25th column of the 2D reservoir observations and compared with results from the 2D implementation.

The likelihood model in 2D factorizes and is given by

$$p(\mathbf{d}^p|\boldsymbol{\pi}, \mathbf{d}^d) = \prod_{(x,t)} p(\mathbf{d}^p_{x,t}|\pi_{x,t}, d^d_{x,t}).$$

The prior model is a profile Markov random field defined by

$$p(\boldsymbol{\pi}_x|\mathbf{d}^d_x, \boldsymbol{\pi}_{-x}) = p(\boldsymbol{\pi}_x|\mathbf{d}^p_{x,t}, \boldsymbol{\pi}_y; y \in \delta(x))$$

where $\delta(x)$ is a first order profile neighbourhood, which in the 2D case means the two closest profiles.

The transition matrices will be adjusted towards a limiting distribution with a larger probability for the neighbouring diagenetic classes. Start with a basic vector $(0.05, 0.05, 0.05, 0.05)$ and add 0.4 to each element representing a neighbouring class, e.g. if the neighbouring classes were $\delta(x) = \{G, M\}$, the limiting distribution would be $\boldsymbol{\zeta} = (0.45, 0.45, 0.05, 0.05)$.

The reservoir will be explored by a block Gibbs method, an MCMC method where the acceptance probability is always one. Profiles are picked one by one left to right and simulated given their full conditional distributions. The profiles are assessed by the Forward-Backward algorithm with transition matrices as presented in Section 5.1 and simulated by Algorithm 3.5. When the algorithm has picked all the profiles in the reservoir we have performed one sweep.

7.1 Convergence

To find the burn-in period of the algorithm we do four runs with different starting configurations. The first configuration is all nodes given diagenetic class G, the second configuration is all nodes given diagenetic class M, the third configuration is all nodes given diagenetic class O and the fourth configuration is all nodes given diagenetic class P. For each sweep the proportion of diagenetic classes are calculated. The proportions of the four runs will approach the same value when iterating, and when this value is reached we believe the algorithm has converged.

Figure 18 shows the proportions through the sweeps for the four runs with different starting configurations. We see that a burn-in period of 60 sweeps is sufficient.

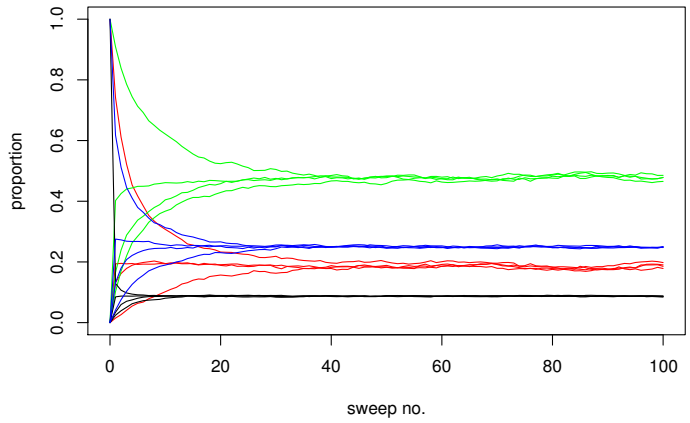


Figure 18: Proportions of diagenetic classes for four runs with different starting configurations. G: red, M: green, O: black, P: blue

7.2 Results and Discussion

We will present results from the 2D implementation and where appropriate compare with results from the 1D implementation. First we do a simulation independent of the reference diagenetic classes in the well, thereafter a simulation conditioning on the reference diagenetic classes.

Figure 19 shows the LMAP for the 2D reservoir predicted by the dominant representative diagenetic class in each node from 1000 realisations. The reference diagenetic classes are displayed at the well location to show the difference in heterogeneity. We see that there is more heterogeneity in the reference observations than the LMAP. This is expected since the MAP-operator tends to reduce heterogeneity.

Figure 20 shows four realisations from the simulation, the reference diagenetic classes are displayed. As expected there is more heterogeneity in the realisations than the LMAP, and the reference diagenetic classes are less prominent in the realisations than the LMAP. Some areas remain the same in all realisations, especially oomoldic and poor seem to be restricted to certain areas where they are present in all four realisations. These areas can also be recognized from the LMAP.

The probability maps in Figure 21 show the probabilities of each diagenetic class calculated from the 1000 realisations from the simulation. We see that for oomoldic and poor there seems to be either a very large or a very small probability for the diagenetic class, which is in accordance with only finding them in restricted areas where they are largely represented. For diagenetic classes good and moderate the probabilities are less extreme, especially in the area with t between 100 and 130.

Figure 22 shows the marginal posterior distribution $p(\pi_t|\mathbf{d}^p, \mathbf{d}^d)$, for the well in the 1D model and $p(\pi_{t,25}|\boldsymbol{\pi}_{-25}, \mathbf{d}^p, \mathbf{d}^d)$, for the well in the 2D model. The marginal distribution for the 2D model is estimated by taking the average of the 1000 realisations. The shape of the marginal posterior distribution for the two models is similar, but the probabilities appear more extreme for the 2D model which is mostly preferable, e.g. for P at $t \approx 65$, M at $t \approx 75$ and G at $t \approx 60$ the probability of picking the correct diagenetic classes has increased. Again we see that for oomoldic and poor the probabilities are either close to zero or one, while diagenetic class good and moderate are less extreme.

Table 3 shows probability of correct classification in the well for the 1D model and the 2D model. We see that overall the 2D model performs marginally better than the 1D model, the improvement is larger for diagenetic classes oomoldic and poor.

Table 4 compares the proportions in the well to 90% CI for well proportions estimated from the 1D and 2D realisations. We see that for diagenetic classes G and O the intervals are closer to the real value for the 1D model, but the intervals are also larger for the 1D model with is in accordance with what we saw from the marginal posterior distributions.

Figure 23 shows 200 realisations for the well from the 2D reservoir, the LMAP for the well from the 2D reservoir and the reference diagenetic classes. Areas where the probability of a certain diagenetic class is large appear as uniform in colour in Figure 23a, while less certain areas appear blurry.

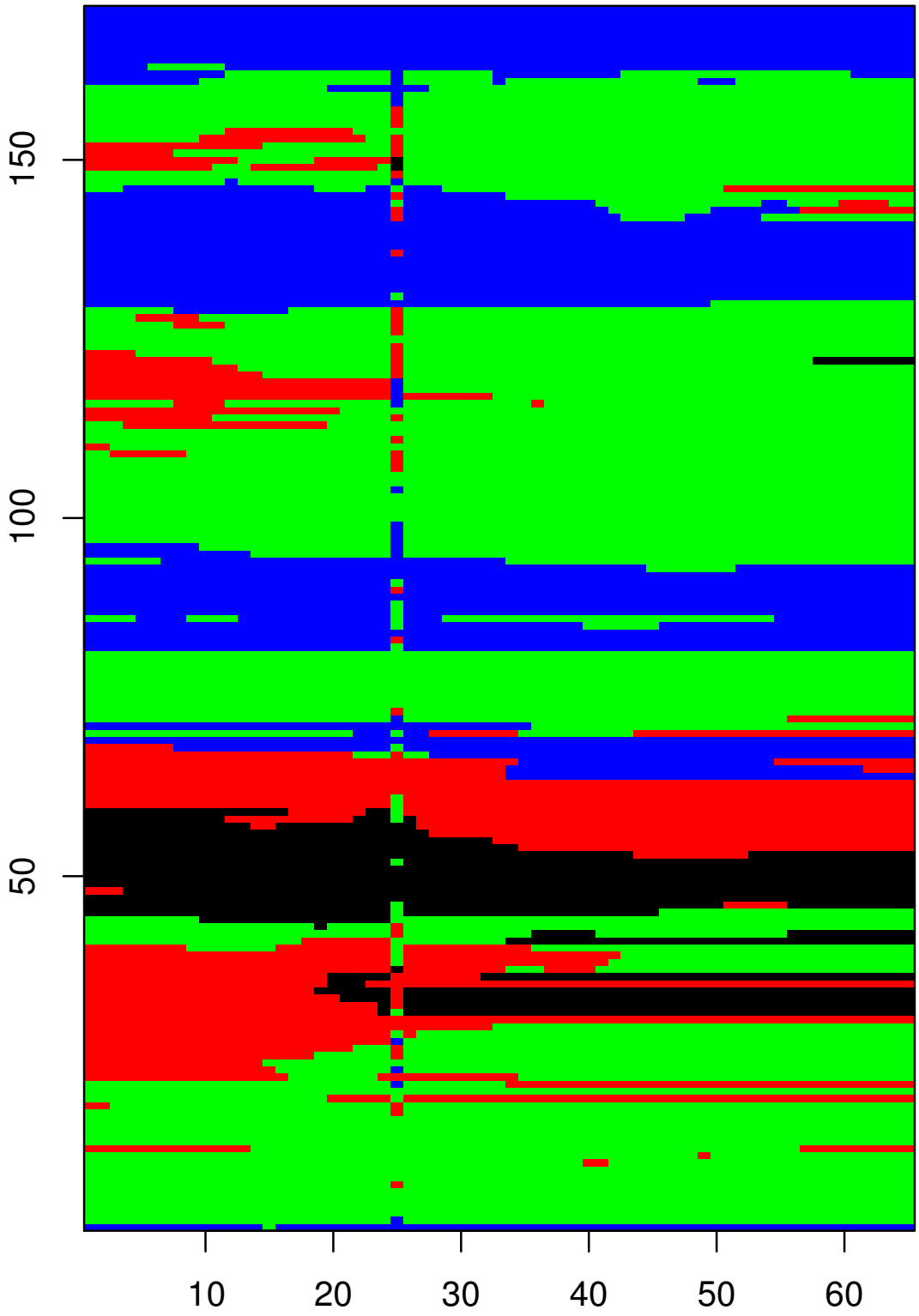


Figure 19: LMAP for the 2D reservoir not conditional on well diagenetic classes, reference well diagenetic classes displayed at column 25

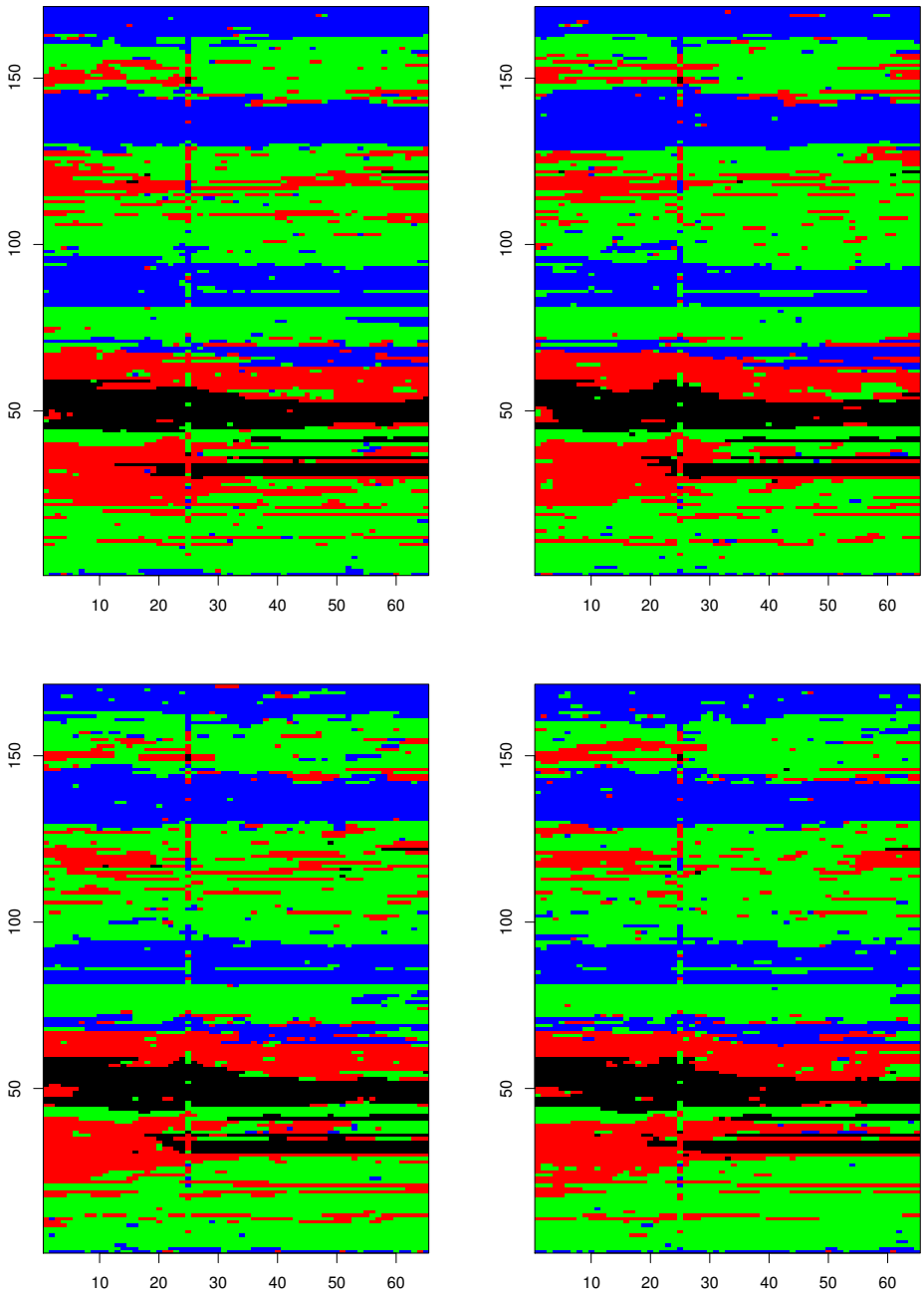
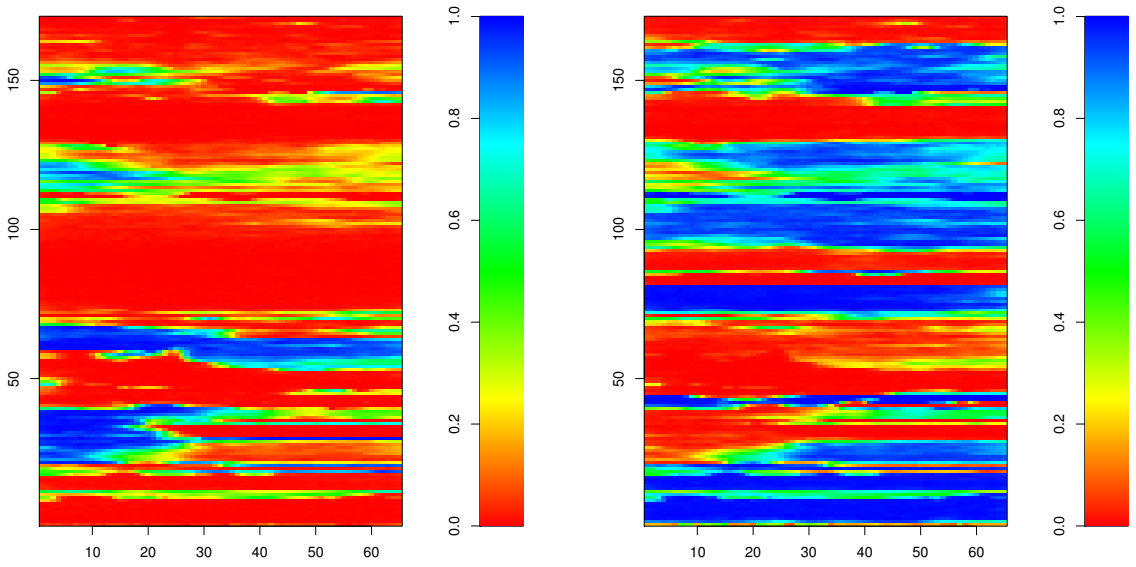
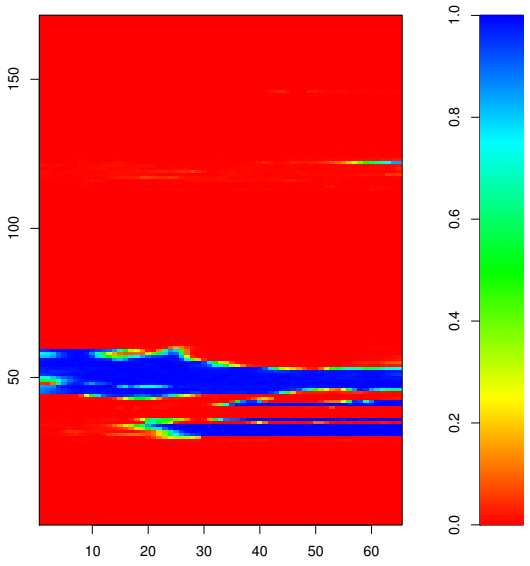


Figure 20: Realisations from the posterior model for the 2D reservoir not conditional on well diagenetic classes, reference well diagenetic classes displayed at column 25

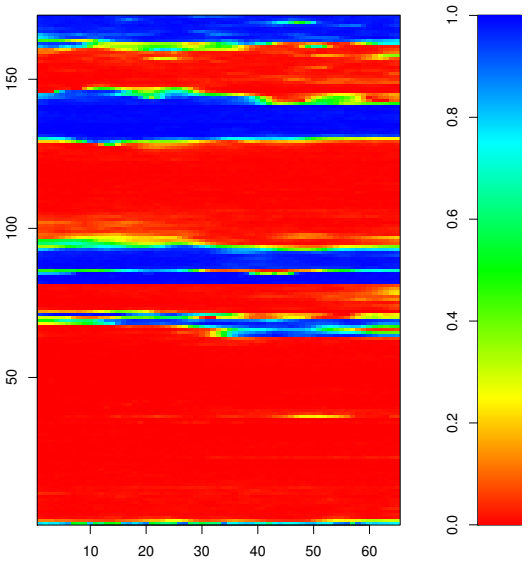


(a) Good

(b) Moderate



(c) Oomoldic



(d) Poor

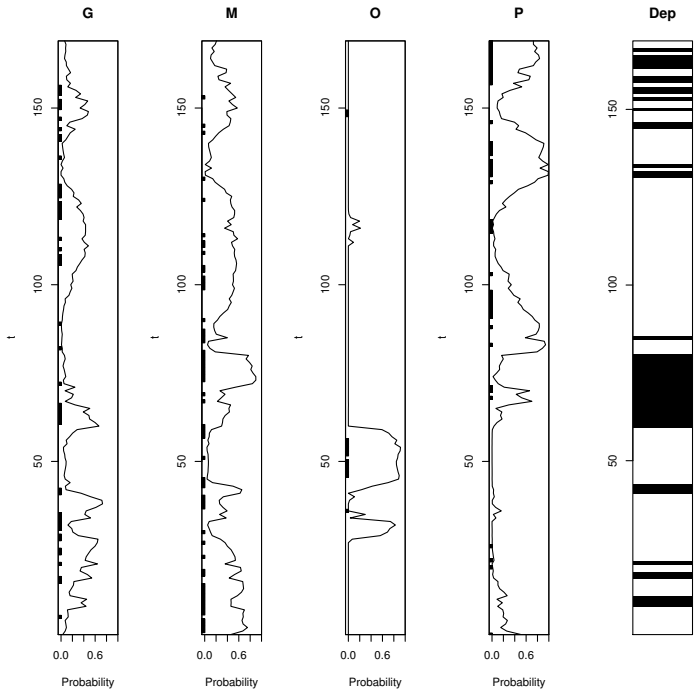
Figure 21: The probability of each diagenetic class for the 2D reservoir not conditional on well diagenetic classes

	1D	2D
δ_G	0.31	0.24
δ_M	0.46	0.56
δ_O	0.67	0.76
δ_P	0.57	0.61
δ	0.47	0.50

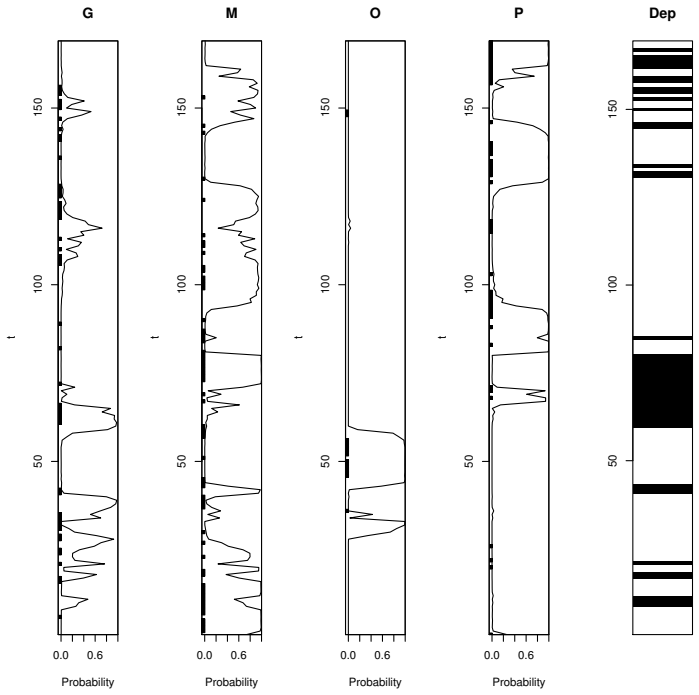
Table 3: Probability of correct classification

π	Prop(π^R)	90% CI for Prop(π) 1D	90% CI for Prop(π) 2D
G	0.29	[0.17,0.27]	[0.12,0.23]
M	0.36	[0.32,0.42]	[0.34,0.44]
O	0.08	[0.08,0.13]	[0.11,0.13]
P	0.29	[0.25,0.36]	[0.26,0.34]

Table 4: Reference proportions of the diagenetic classes compared to 90% CI for estimated proportions of diagenetic classes

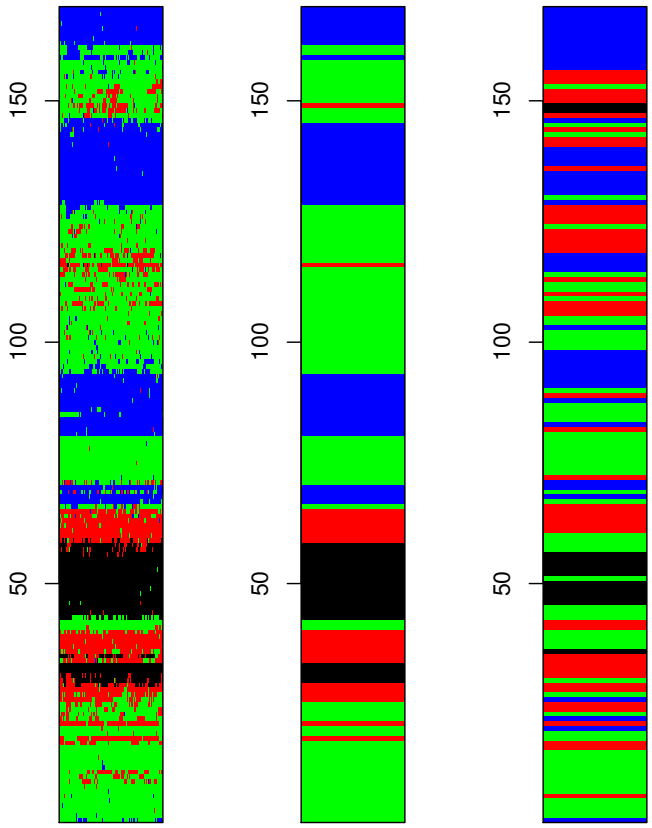


(a) 1D



(b) 2D

Figure 22: Plot showing the marginal posterior distribution for each diagenetic class, the reference diagenetic class is marked on the left axis. The plot to the right is the depositional properties in the well, GS: white, FS: black



(a) Simulations (b) LMAP (c) Diagenetic class

Figure 23: (a) 200 independent realisations from $p(\boldsymbol{\pi}_{25}|\boldsymbol{\pi}_{-25}, \mathbf{d}^p, \mathbf{d}^d)$. (b) Locationwise Maximum a Posteriori (c) Reference diagenetic classes in the well. G: red, M: green, O: black, P: blue

We also do a run where the model is conditioned on well diagenetic classes. The reference observations $\boldsymbol{\pi}^R$, $\pi^R \in \{G, M, O, P\}$, are considered exact observations of the diagenetic classes along the well profile. Let the posterior model be

$$p(\boldsymbol{\pi}|\boldsymbol{\pi}^R, \mathbf{d}^p, \mathbf{d}^d) = const \times p(\mathbf{d}^p|\boldsymbol{\pi}, \mathbf{d}^d)p(\boldsymbol{\pi}^R|\boldsymbol{\pi})p(\boldsymbol{\pi})$$

where $p(\boldsymbol{\pi}^R|\boldsymbol{\pi})$ is the well likelihood model defined by

$$p(\pi^R_{x,t}|\pi_{x,t}) = \delta\left(\pi^R_{x,t}, \pi_{x,t}\right)$$

where $\delta(\cdot,\cdot)$ is the Kronecker delta. The well likelihood is only defined at the well location.

Figure 24 shows the LMAP of the 2D model conditioned on well diagenetic classes. The effect of the conditioning well is mostly present in the areas where the probabilities of diagenetic classes are less extreme, i.e. for t between 100 and 130.

The four realisations in Figure 25 show that when the model is conditioned on the well diagenetic classes the difference between the well location and the rest of the reservoir is almost unnoticeable.

Figure 26 shows the probability maps for the run conditioned on well diagenetic classes. If a well diagenetic class is present in an area where the probability of that given class is close to zero the lateral continuity might have a small effect, e.g. for M at $t = 51$ there is a small increase in the probability for M at node $(51, 24)$ and node $(51, 26)$ compared with Figure 21, but the effect is so small that it does not impact the LMAP.

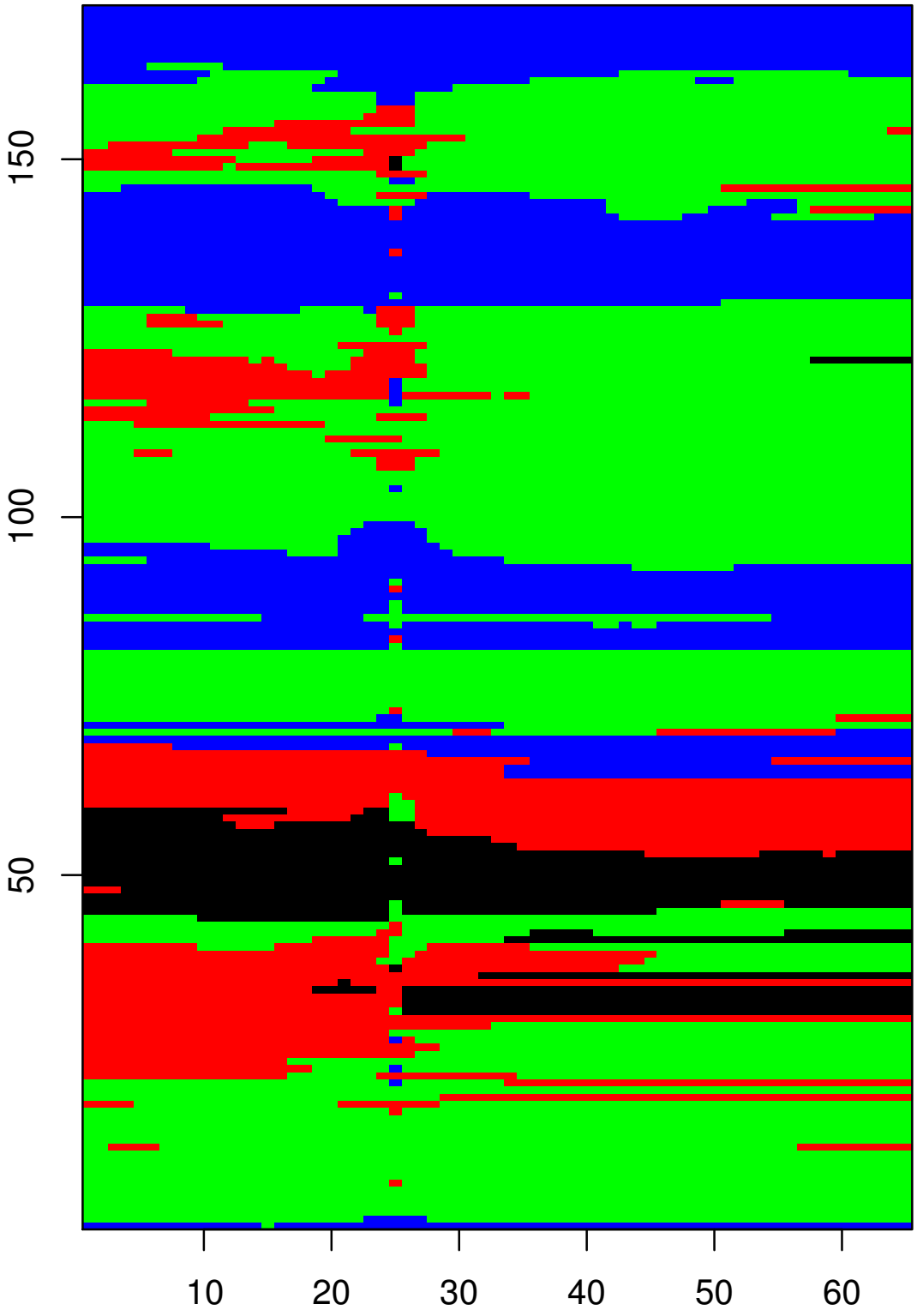


Figure 24: LMAP for the 2D reservoir conditioned on well diagenetic classes

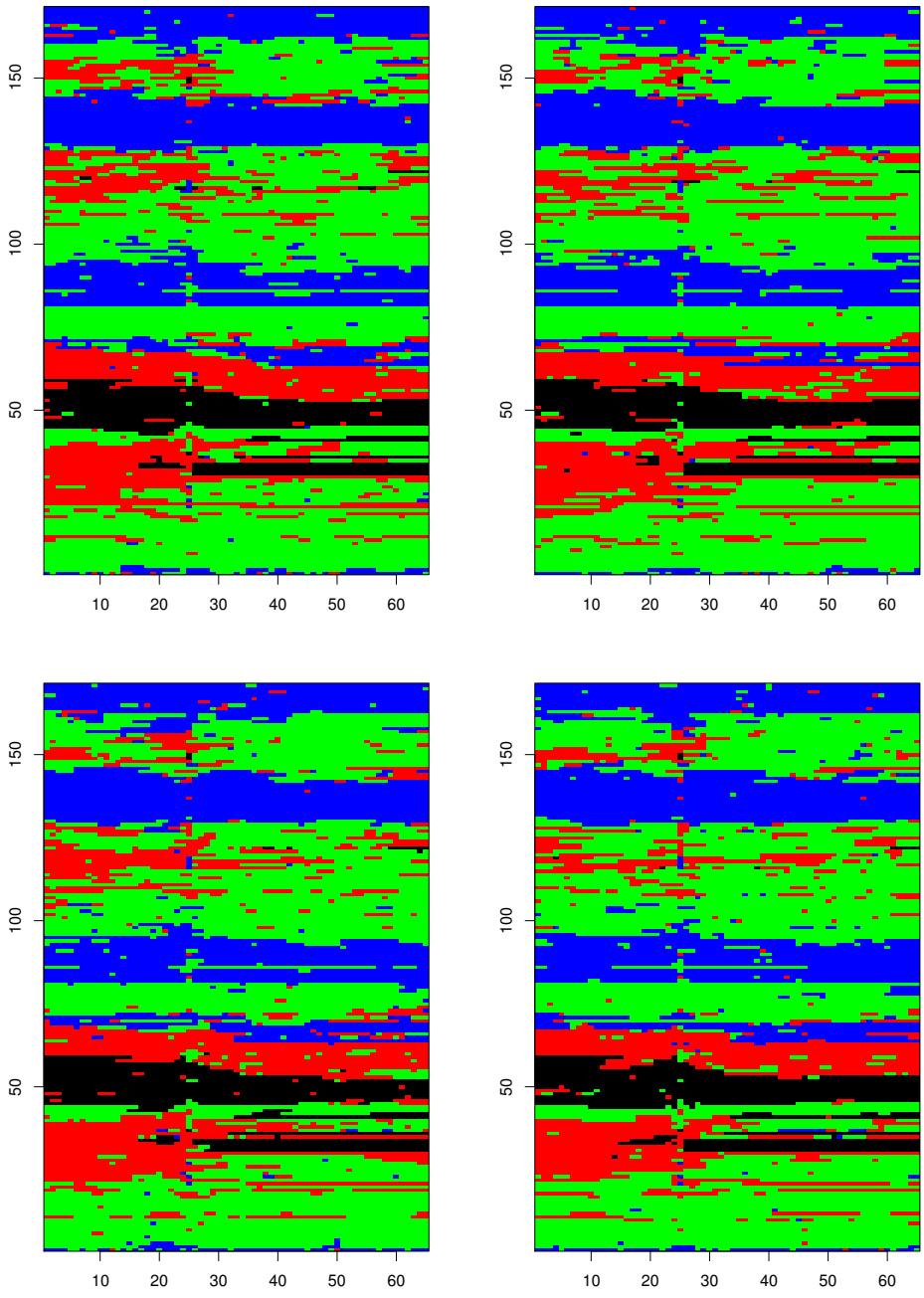
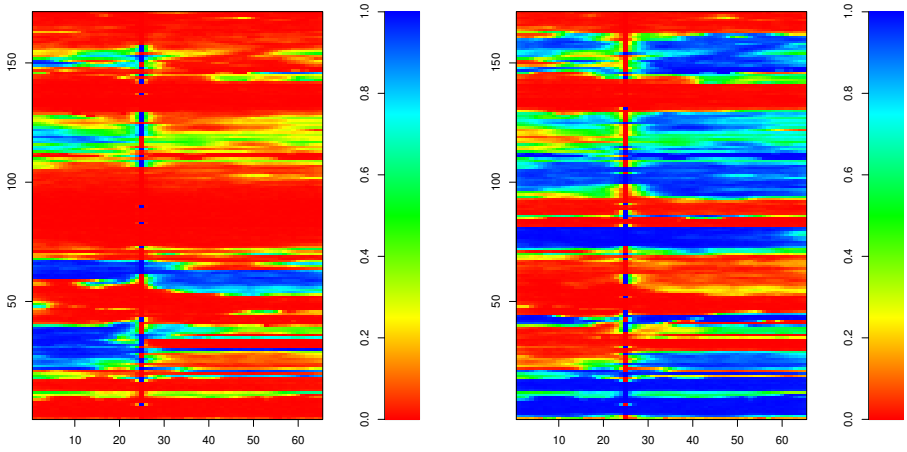


Figure 25: Realisations from the posterior model for the 2D reservoir conditioned on well diagenetic classes



(a) Good

(b) Moderate

(c) Oomoldic

(d) Poor

Figure 26: The probability of each diagenetic class for the 2D reservoir conditioned on well diagenetic classes

8 Conclusion

In this thesis Bayesian inversion is performed on real observations to predict the diagenetic classes of a carbonate reservoir where the proportions of carbonate rock and depositional properties are known. Experiences from the thesis are:

- A good statistical formulation of the problem is developed in 3D. The model captures most known effects such as lateral continuity. Reliable parameter estimates for the Markov chain and the generalized Dirichlet distribution reproduce the 1D well observations.
- Results from a study on real data show that simulations reproduce the heterogeneity of the real observations. Originally the model conditions on reservoir observations, but conditioning on well observations is easily incorporated and we can simulate the representative 2D diagenetic classes for both models.
- Future work should look into the origin of conditioning. Implementation of the model for a 3D reservoir should also be studied as the computing time is quite fast for the 2D model.

References

Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems', *Journal of the Royal Statistical Society: Series B (Methodological)* **36**, 192–236.

Connor, R. J. & Mosimann, J. E. (1969), 'Concepts of independence for proportions with a generalization of the Dirichlet distribution', *Journal of the American Statistical Association* **64**, 194–206.

Dunham, R. J. (1962), 'Classification of carbonate rocks according to depositional texture', *American Association of Petroleum Geologists* **1**, 108–121.

Gamerman, D. & Lopes, H. F. (2006), *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*, second edn, Chapman & Hall.

Hastings, W. K. (1970), 'Monte carlo sampling methods using Markov chains and their applications', **57**, 97–109.

Johnson, R. A. & Wichern, D. W. (2007), *Applied Multivariate Statistical Analysis*, Sixth edn, Pearson Prentice Hall.

Rimstad, K. & Omre, H. (2010), 'Impact of rock physics depth trends and Markov random fields on hierarchical Bayesian lithology/fluid prediction', *To appear in Geophysics* .

Ross, S. M. (2007), *Introduction to Probability Models*, ninth edn, Academic Press.

Scott, S. L. (2002), 'Bayesian methods for hidden Markov models: Recursive computing in the 21st century', *Journal of the American Statistical Association* **97**, 337–351.

Ulvmoen, M. & Omre, H. (2010), 'Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations: Part 1 - methodology', *Geophysics* **75**, R21–R35.

Ulvmoen, M., Omre, H. & Buland, A. (2010), 'Improved resolution in Bayesian lithology/fluid inversion from prestack seismic data and well observations: Part 2 - real case study', *Geophysics* **75**, B73–B82.

Walpole, Myers, Myers & Ye (2007), *Probability & Statistics for Engineers and Scientists*, Eigth edn, Pearson Prentice Hall.

Wicker, N., Muller, J., Kalathur, R. K. R. & Poch, O. (2008), 'A maximum likelihood approximation method for Dirichlet's parameter estimation', *Computational Statistics & Data Analysis* **52**, 1315–1322.

Wong, T.-T. (1998), 'Generalized Dirichlet distribution in Bayesian analysis', *Applied Mathematics and Computation* **97**, 165–181.