# Uncertainty and statistics
# Part IV

Håkon Tjelmeland

CGF and Department of Mathematical Sciences
Norwegian University of Science and Technology

Friday November 12th 2021

# Recall the typical problem setup

- Have a model for stochastic variables, $x_1, \dots, x_n, y_1, \dots, y_m$
  - we have a formula for $p(x_1, \dots, x_n, y_1, \dots, y_m)$
  - we have observed $y_1, \dots, y_m$
  - distribution of interest

$$p(x_1, \dots, x_n | y_1, \dots, y_m) = \frac{p(x_1, \dots, x_n, y_1, \dots, y_m)}{p(y_1, \dots, y_m)}$$

where

$$p(y_1, \dots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \dots, x_n, y_1, \dots, y_m) dx_1 \cdots dx_n$$

# Recall the typical problem setup

- Have a model for stochastic variables, $x_1, \ldots, x_n, y_1, \ldots, y_m$
  - we have a formula for $p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we have observed $y_1, \ldots, y_m$
  - distribution of interest

    $$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

    where

    $$p(y_1, \ldots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \ldots, x_n, y_1, \ldots, y_m) dx_1 \cdots dx_n$$

  - common to write

    $$p(x_1, \ldots, x_n | y_1, \ldots, y_m) \propto p(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

# Recall the typical problem setup

- Have a model for stochastic variables, $x_1, \ldots, x_n, y_1, \ldots, y_m$
  - we have a formula for $p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we have observed $y_1, \ldots, y_m$
  - distribution of interest

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) = \frac{p(x_1, \ldots, x_n, y_1, \ldots, y_m)}{p(y_1, \ldots, y_m)}$$

where

$$p(y_1, \ldots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \ldots, x_n, y_1, \ldots, y_m) dx_1 \cdots dx_n$$

  - common to write

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) \propto p(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

- Focus today:
  - Monte Carlo sampling

$$(x_1, \ldots, x_n) \sim p(x_1, \ldots, x_n | y_1, \ldots, y_m)$$

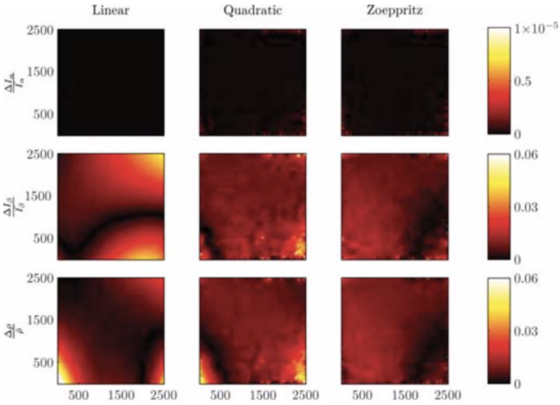
  - MCMC, Metropolis–Hastings algorithm, Gibbs sampling

# Plan for today

- ⋆ Introduce two example situations
  - Rabben, T.E., Tjelmeland, H. and Ursin, B. (2008). Non-linear Bayesian joint inversion of seismic reflection coefficients, Geophysical Journal International.
  - Tjelmeland, H., Luo, X. and Fjeldstad, T. (2019). A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image, Geophysical Prospecting.
- ⋆ The Metropolis–Hastings algorithm
  - the algorithm, proposal distribution
  - toy examples, intuition
  - random walk proposals, Gibbs proposals
  - what is converging?
  - reversible jump Metropolis–Hastings
- ⋆ Metropolis–Hastings for the two example situations
  - proposal distribution
  - results

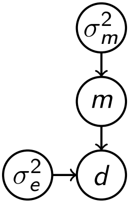
# Inversion of seismic reflection coefficients

- ⋆ Reference: Rabben, T.E., Tjelmeland, H. and Ursin, B. (2008). Non-linear Bayesian joint inversion of seismic reflection coefficients, Geophysical Journal International.



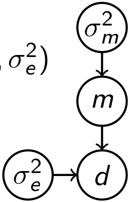
# Inversion of seismic reflection coefficients

- Variables in the problem:
  - $m = \{m_{ij}; i = 1, \ldots, n_y, j = 1, \ldots, n_x\}, m_{ij} = \left\{ \frac{\Delta I_\alpha}{\bar{I}_\alpha}, \frac{\Delta I_\beta}{\bar{I}_\beta}, \frac{\Delta \rho}{\bar{\rho}} \right\}_{ij}$
  - $d = \{d_{ij}, i = 1, \ldots, n_y, j = 1, \ldots, n_x\}, d_{ij} = \{r_{PP}(\theta), r_{PS}(\theta)\}_{ij}$ for $\theta = 0^o(\text{only PP}), 20^o, 40^o, 55^o$
  - $\sigma_m^2$: variance of $m_{ij}$
  - $\sigma_e^2$: variance of $d_{ij}|m$
- Structure of stochastic model




$$p(m, d, \sigma_m^2, \sigma_e^2) = p(\sigma_m^2)p(\sigma_e^2)p(m|\sigma_m^2)p(d|m, \sigma_e^2)$$

# Inversion of seismic reflection coefficients



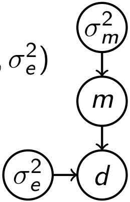

- Recall: $p(m, d, \sigma_m^2, \sigma_e^2) = p(\sigma_m^2)p(\sigma_e^2)p(m|\sigma_m^2)p(d|m, \sigma_e^2)$

- Model components
  - $m|\sigma_m^2 \sim \mathrm{N}(\mu_m, \sigma_m^2 S_m)$
  - $d|m, \sigma_e^2 \sim \mathrm{N}(f(m), \sigma_e^2 S_e)$

$$d = f(m) + e, e \sim \mathrm{N}(0, \sigma_e^2 S_e)$$

  - $\sigma_m^2 \sim \mathrm{IG}(\alpha_m, \beta_m)$
  - $\sigma_e^2 \sim \mathrm{IG}(\alpha_e, \beta_e)$

# Inversion of seismic reflection coefficients




- Recall: $p(m, d, \sigma_m^2, \sigma_e^2) = p(\sigma_m^2)p(\sigma_e^2)p(m|\sigma_m^2)p(d|m, \sigma_e^2)$

- Model components
  - $m|\sigma_m^2 \sim \mathrm{N}(\mu_m, \sigma_m^2 S_m)$
  - $d|m, \sigma_e^2 \sim \mathrm{N}(f(m), \sigma_e^2 S_e)$

$$d = f(m) + e, e \sim \mathrm{N}(0, \sigma_e^2 S_e)$$

  - $\sigma_m^2 \sim \mathrm{IG}(\alpha_m, \beta_m)$
  - $\sigma_e^2 \sim \mathrm{IG}(\alpha_e, \beta_e)$

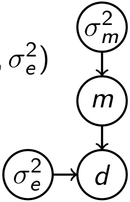
- Use three versions for $f(m)$
  - Zoeppritz equations
  - quadratic approximation of the Zoeppritz equations
  - linear approximation of the Zoeppritz equations

# Quadratic approximation of the Zoeppritz equations

$$
\begin{aligned}
r_{\mathrm{PP}} = {} & \frac{1}{2\cos^2\theta_p}\frac{\Delta I_\alpha}{\bar{I}_\alpha} - 4\sin^2\theta_s\frac{\Delta I_\beta}{\bar{I}_\beta} - \frac{1}{2}\tan^2\theta_p\left(1-4\gamma^2\cos^2\theta_p\right)\frac{\Delta\rho}{\bar{\rho}} \\
& + \tan\theta_p\tan\theta_s\Bigg\{4\gamma^2(1-(1+\gamma^2)\sin^2\theta_p)\left(\frac{\Delta I_\beta}{\bar{I}_\beta}\right)^2 \\
& -4\gamma^2\left[1-\left(\frac{3}{2}+\gamma^2\right)\sin^2\theta_p\right]\left(\frac{\Delta I_\beta}{\bar{I}_\beta}\frac{\Delta\rho}{\bar{\rho}}\right) \\
& + \left[\gamma^2(1-(2+\gamma^2)\sin^2\theta_p) - \frac{1}{4}\right]\left(\frac{\Delta\rho}{\bar{\rho}}\right)^2\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
r_{\mathrm{PS}} = {} & \sqrt{\tan\theta_p\tan\theta_s}\Bigg\{\Bigg[\left[(1-\cos\theta_s(\cos\theta_s+\gamma\cos\theta_p))\left(2\frac{\Delta I_\beta}{\bar{I}_\beta}-\frac{\Delta\rho}{\bar{\rho}}\right)-\frac{1}{2}\frac{\Delta\rho}{\bar{\rho}}\right] \\
& +\frac{1}{2}\left[(1-\cos\theta_s(\cos\theta_s-\gamma\cos\theta_p))\left(2\frac{\Delta I_\beta}{\bar{I}_\beta}-\frac{\Delta\rho}{\bar{\rho}}\right)-\frac{1}{2}\frac{\Delta\rho}{\bar{\rho}}\right] \\
& \times\left[\frac{1}{2\cos^2\theta_p}\frac{\Delta I_\alpha}{\bar{I}_\alpha}+\left(\frac{1}{2\cos^2\theta_s}-8\sin^2\theta_s\right)\frac{\Delta I_\beta}{\bar{I}_\beta}\right. \\
& \left.+\left(4\sin^2\theta_s-\frac{1}{2}(\tan^2\theta_p+\tan^2\theta_s)\right)\frac{\Delta\rho}{\bar{\rho}}\right]\Bigg]\Bigg\},
\end{aligned}
$$

# Inversion of seismic reflection coefficients




- Recall: $p(m, d, \sigma_m^2, \sigma_e^2) = p(\sigma_m^2)p(\sigma_e^2)p(m|\sigma_m^2)p(d|m, \sigma_e^2)$

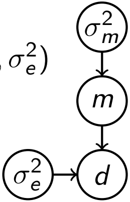
- Model components
  - $m|\sigma_m^2 \sim \mathrm{N}(\mu_m, \sigma_m^2 S_m)$
  - $d|m, \sigma_e^2 \sim \mathrm{N}(f(m), \sigma_e^2 S_e)$

$$d = f(m) + e, e \sim \mathrm{N}(0, \sigma_e^2 S_e)$$

  - $\sigma_m^2 \sim \mathrm{IG}(\alpha_m, \beta_m)$
  - $\sigma_e^2 \sim \mathrm{IG}(\alpha_e, \beta_e)$

- Use three versions of $f(m)$
  - Zoeppritz equations
  - quadratic approximation of the Zoeppritz equations
  - linear approximation of the Zoeppritz equations

# Inversion of seismic reflection coefficients




- Recall: $p(m, d, \sigma_m^2, \sigma_e^2) = p(\sigma_m^2)p(\sigma_e^2)p(m|\sigma_m^2)p(d|m, \sigma_e^2)$

- Model components
  - $m|\sigma_m^2 \sim \mathrm{N}(\mu_m, \sigma_m^2 S_m)$
  - $d|m, \sigma_e^2 \sim \mathrm{N}(f(m), \sigma_e^2 S_e)$

$$d = f(m) + e, e \sim \mathrm{N}(0, \sigma_e^2 S_e)$$

  - $\sigma_m^2 \sim \mathrm{IG}(\alpha_m, \beta_m)$
  - $\sigma_e^2 \sim \mathrm{IG}(\alpha_e, \beta_e)$

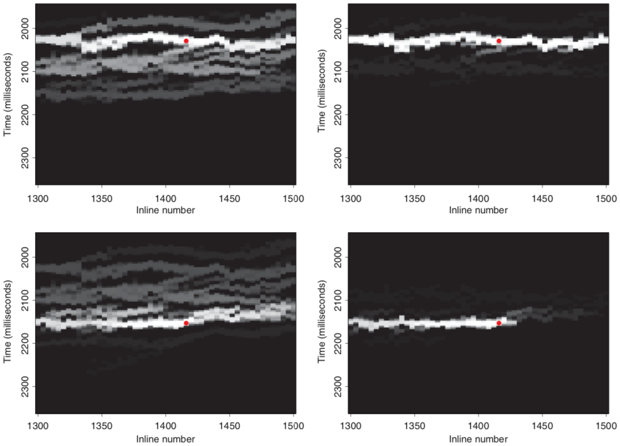
- Use three versions of $f(m)$
  - Zoeppritz equations
  - quadratic approximation of the Zoeppritz equations
  - linear approximation of the Zoeppritz equations

- Distribution of interest

$$p(m, \sigma_m^2, \sigma_e^2|d) \propto p(m, d, \sigma_m^2, \sigma_e^2)$$

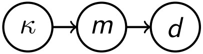
# Lithology/fluid class prediction using a Markov mesh prior

- ⋆ Reference: Tjelmeland, H., Luo, X. and Fjeldstad, T. (2019). A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image, Geophysical Prospecting.



# Lithology/fluid class prediction using a Markov mesh prior

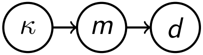
- Variables in the problem:
  - $\kappa = \{\kappa_{ij}; i = 1, \ldots, n_y, j = 1, \ldots, n_x\}$, $\kappa_{ij} \in \{0, 1\}$ (shale, oil sand)
  - $m = \{m_{ij}; i = 1, \ldots, n_y, j = 1, \ldots, n_x\}$, $m_{ij} = (\text{impedance}, v_p/v_s)_{ij}$
  - $d = \{d_{ij}; i = 1, \ldots, n_y, j = 1, \ldots, n_x\}$, $d_{ij} = (\text{near offset}, \text{far offset})_{ij}$
  - $\theta$: parameters vector describing stochastic model for $\kappa$ (fixed)
- Structure of the stochastic model




$$p(\kappa, m, d|\theta) = p(\kappa|\theta)p(m|\kappa)p(d|m)$$

# Lithology/fluid class prediction using a Markov mesh prior

- Recall: $p(\kappa, m, d|\theta) = p(\kappa|\theta)p(m|\kappa)p(d|m)$

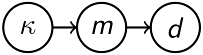



- Model components
  - $\kappa|\theta \sim$ Markov mesh model
  - $m_{ij}|\kappa \sim \mathrm{N}(\mu_{\kappa_{ij}}, \Sigma_{\kappa_{ij}})$
  - $d|m \sim \mathrm{N}(WADm, \Sigma_{\varepsilon})$

$$d = WADm + \varepsilon, \varepsilon \sim \mathrm{N}(0, \Sigma_{\varepsilon})$$

# Lithology/fluid class prediction using a Markov mesh prior

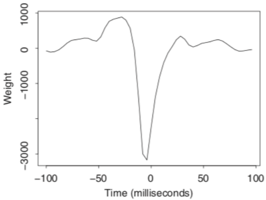
- ⋆ Recall: $p(\kappa, m, d|\theta) = p(\kappa|\theta)p(m|\kappa)p(d|m)$



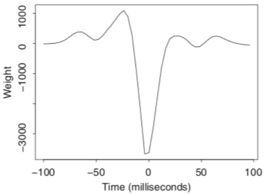

- ⋆ Model components
  - $\kappa|\theta \sim$ Markov mesh model
  - $m_{ij}|\kappa \sim \mathrm{N}(\mu_{\kappa_{ij}}, \Sigma_{\kappa_{ij}})$
  - $d|m \sim \mathrm{N}(WADm, \Sigma_{\varepsilon})$

$$d = WADm + \varepsilon, \varepsilon \sim \mathrm{N}(0, \Sigma_{\varepsilon})$$

wavelet for near offset




wavelet for far offset



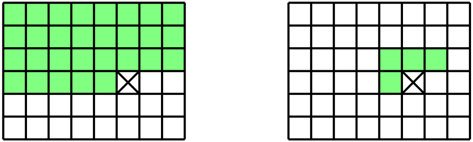
# Markov mesh model

⋆ Order the nodes in lexicographical order from 1 to $n = n_x \cdot n_y$

$$p(\kappa|\theta) = \prod_{i=1}^{n} p(\kappa_i|\kappa_{<i}, \theta)$$

⋆ Conditional independence (Markov) assumption

$$p(\kappa|\kappa_{<i}, \theta) = p(\kappa_i|\kappa_{\nu_i}, \theta)$$



⋆ Thus

$$p(\kappa|\theta) = \prod_{i=1}^{n} p(\kappa_i|\kappa_{\nu_i}, \theta)$$

# Markov mesh model

- ⋆ Recall:

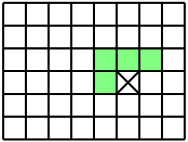
$$p(\kappa|\theta) = \prod_{i=1}^{n} p(\kappa_i|\kappa_{\nu_i}, \theta)$$

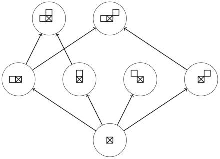
- ⋆ Reformulation of $p(\kappa_i|\kappa_{\nu_i}, \theta)$

$$\theta_i(\kappa_{\nu_i}) = \ln\left(\frac{p(\kappa_i = 1|\kappa_{\nu_i}, \theta)}{1 - p(\kappa_i = 1|\kappa_{\nu_i}, \theta)}\right)$$

$$\theta_i(\kappa_{\nu_i}) = \sum_{\lambda \subseteq \kappa_{\nu_i}} \beta(\lambda)$$

- ⋆ Limit the number of model parameters by setting some $\beta(\lambda) = 0$

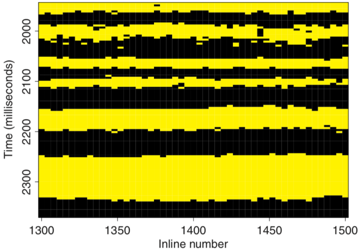




- ⋆ $\theta = (\nu_i; \beta(\lambda), \nu_i \subseteq \nu_i)$

# How to specify $\theta$?

- ⋆ Values in $\theta$ does not have a simple interpretation
  - difficult to set values consistent with prior knowledge
- ⋆ Use a training image
  - hand drawn or from area with similar geological origin
  - contain information about typical spatial structures

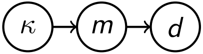



- ⋆ Let $\theta$ be a Monte Carlo sample from

$$p(\theta|\text{training image}) \propto p(\theta)p(\text{training image}|\theta)$$

# Lithology/fluid class prediction using a Markov mesh prior

- Recall: $p(\kappa, m, d|\theta) = p(\kappa|\theta)p(m|\kappa)p(d|m)$

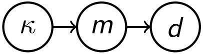


- Model components
  - $\kappa|\theta \sim$ Markov mesh model
  - $m_{ij}|\kappa \sim \mathrm{N}(\mu_{\kappa_{ij}}, \Sigma_{\kappa_{ij}})$
  - $d|m \sim \mathrm{N}(WADm, \Sigma_{\varepsilon})$

$$d = WADm + \varepsilon, \varepsilon \sim \mathrm{N}(0, \Sigma_{\varepsilon})$$

# Lithology/fluid class prediction using a Markov mesh prior

- Recall: $p(\kappa, m, d|\theta) = p(\kappa|\theta)p(m|\kappa)p(d|m)$




- Model components
  - $\kappa|\theta \sim$ Markov mesh model
  - $m_{ij}|\kappa \sim \mathrm{N}(\mu_{\kappa_{ij}}, \Sigma_{\kappa_{ij}})$
  - $d|m \sim \mathrm{N}(WADm, \Sigma_\varepsilon)$

$$d = WADm + \varepsilon, \varepsilon \sim \mathrm{N}(0, \Sigma_\varepsilon)$$

- Distribution of interest

$$p(\kappa, m|\theta, d) \propto p(\kappa, m, d|\theta)$$

  - can analyticall integrate over $m$ to find formula for

$$p(\kappa|\theta, d) \propto p(\kappa, d|\theta)$$

# Markov chain Monte Carlo

- ⋆ Have a model for stochastic variables, $x_1, \ldots, x_n, y_1, \ldots, y_m$
  - we have a formula for $p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we have observed $y_1, \ldots, y_m$
  - distribution of interest

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) \propto p(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

- ⋆ Note: The $y$'s are observed numbers

# Markov chain Monte Carlo

- ⋆ Have a model for stochastic variables, $x_1, \ldots, x_n, y_1, \ldots, y_m$
  - we have a formula for $p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we have observed $y_1, \ldots, y_m$
  - distribution of interest

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) \propto p(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

- ⋆ Note: The $y$'s are observed numbers
- ⋆ We now simplify the notation and write

$$p(x) = p(x_1, \ldots, x_n) = c \cdot h(x) \text{ in stead of } p(x_1, \ldots, x_n | y_1, \ldots, y_m)$$

  - we have a formula for $h(x) = p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we don't know the value of $c$

# Markov chain Monte Carlo

- Have a model for stochastic variables, $x_1, \ldots, x_n, y_1, \ldots, y_m$
  - we have a formula for $p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we have observed $y_1, \ldots, y_m$
  - distribution of interest

$$p(x_1, \ldots, x_n | y_1, \ldots, y_m) \propto p(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

- Note: The $y$'s are observed numbers
- We now simplify the notation and write

$$p(x) = p(x_1, \ldots, x_n) = c \cdot h(x) \text{ in stead of } p(x_1, \ldots, x_n | y_1, \ldots, y_m)$$

  - we have a formula for $h(x) = p(x_1, \ldots, x_n, y_1, \ldots, y_m)$
  - we don't know the value of $c$
- So the goal is to generate a Monte Carlo sample

$$x \sim p(x) = c \cdot h(x)$$

# Markov chain Monte Carlo

- ⋆ Want a Monte Carlo sample from

$$x \sim p(x) = c \cdot h(x)$$

- ⋆ Metropolis–Hastings algorithm: Generate $x^1, x^2, \ldots$ as follows
  1. Generate/set $x^1$ somehow
  2. For $k = 2, 3, \ldots, K$
     (a) Generate a potential new values $z^k \sim q(z^k | x^{k-1})$
     (b) Compute acceptance probability

$$\alpha(z^k | x^{k-1}) = \min \left\{ 1, \frac{p(z^k)}{p(x^{k-1})} \cdot \frac{q(x^{k-1} | z^k)}{q(z^k | x^{k-1})} \right\}$$

     (c) Generate $u^k \sim \text{Uniform}[0, 1]$
     (d) If $(u^k \leq \alpha(z^k | x^{k-1})$
         $x^k = z^k$ (accept $z^k$)
         else
         $x^k = x^{k-1}$ (reject $z^k$)
  3. End

# Markov chain Monte Carlo

- Want a Monte Carlo sample from

$$x \sim p(x) = c \cdot h(x)$$

- Metropolis–Hastings algorithm: Generate $x^1, x^2, \ldots$ as follows
  1. Generate/set $x^1$ somehow
  2. For $k = 2, 3, \ldots, K$
     (a) Generate a potential new values $z^k \sim q(z^k | x^{k-1})$
     (b) Compute acceptance probability

$$\alpha(z^k | x^{k-1}) = \min \left\{ 1, \frac{p(z^k)}{p(x^{k-1})} \cdot \frac{q(x^{k-1} | z^k)}{q(z^k | x^{k-1})} \right\}$$

     (c) Generate $u^k \sim \text{Uniform}[0, 1]$
     (d) If $(u^k \leq \alpha(z^k | x^{k-1})$
            $x^k = z^k$ (accept $z^k$)
         else
            $x^k = x^{k-1}$ (reject $z^k$)
  3. End
- For $k$ large enough, $x^k$ is (approximately) from $p(x)$

# Random walk proposal

- ⋆ Target distribution: $p(x)$
- ⋆ Proposal distribution: $z^k \sim q(z^k|x^{k-1})$
- ⋆ Acceptance probability:

$$\alpha(z^k|x^{k-1}) = \min\left\{1, \frac{p(z^k)}{p(x^{k-1})} \cdot \frac{q(x^{k-1}|z^k)}{q(z^k|x^{k-1})}\right\}$$

# Random walk proposal

- ⋆ Target distribution: $p(x)$
- ⋆ Proposal distribution: $z^k \sim q(z^k | x^{k-1})$
- ⋆ Acceptance probability:

$$\alpha(z^k | x^{k-1}) = \min \left\{ 1, \frac{p(z^k)}{p(x^{k-1})} \cdot \frac{q(x^{k-1} | z^k)}{q(z^k | x^{k-1})} \right\}$$

- ⋆ Random walk proposal:

$$z_i^k = x_i^{k-1} + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, \sigma^2)$$

  - formula for the proposal distribution

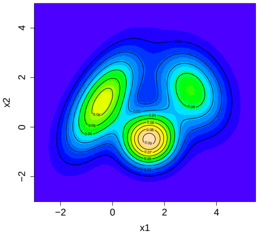
$$q(z^k | x^{k-1}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} (z^k - x^{k-1})^T (z^k - x^{k-1}) \right\}$$

  - acceptance probability

$$\alpha(z^k | x^{k-1}) = \min \left\{ 1, \frac{p(z^k)}{p(x^{k-1})} \right\}$$

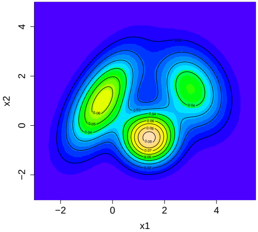
# Toy example: Random walk proposal

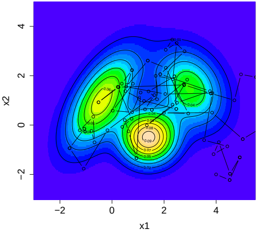
- Target distribution: $p(x)$
- Start with $x^1 = (4, -2)$
- Random walk proposal with $\sigma = 1.0$
- Run $K = 100$ iterations

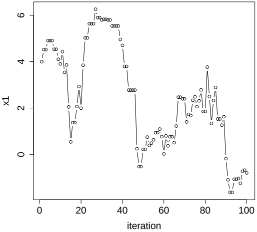


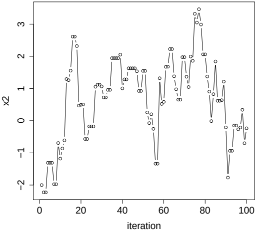
# Toy example: Random walk proposal

- ⋆ Target distribution: $p(x)$
- ⋆ Start with $x^1 = (4, -2)$
- ⋆ Random walk proposal with $\sigma = 1.0$
- ⋆ Run $K = 100$ iterations



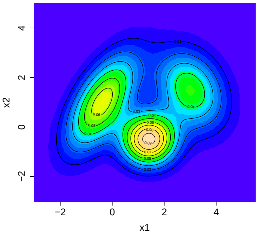



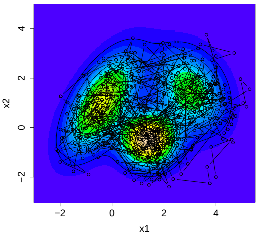



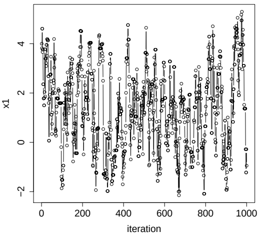



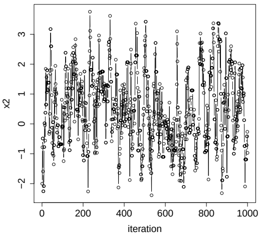
# Toy example: Random walk proposal

- ⋆ Target distribution: $p(x)$
- ⋆ Start with $x^1 = (4, -2)$
- ⋆ Random walk proposal with $\sigma = 1.0$
- ⋆ Run $K = 1000$ iterations



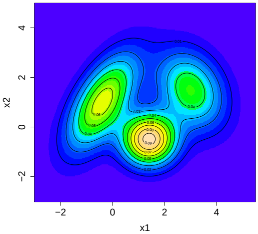



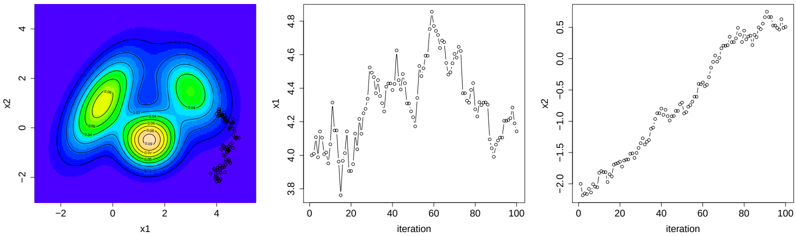



# Toy example: Random walk proposal

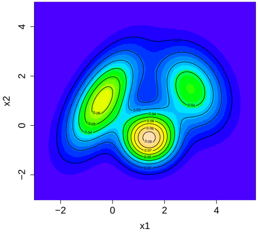
- ⋆ Target distribution: $p(x)$
- ⋆ Start with $x^1 = (4, -2)$
- ⋆ Random walk proposal with $\sigma = 0.1$
- ⋆ Run $K = 100$ iterations

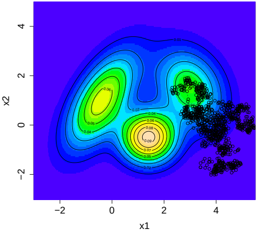


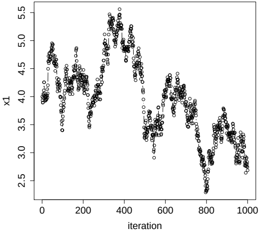



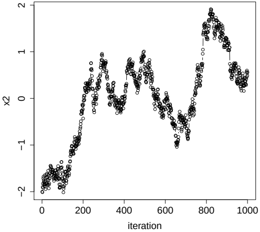
# Toy example: Random walk proposal

- ⋆ Target distribution: $p(x)$
- ⋆ Start with $x^1 = (4, -2)$
- ⋆ Random walk proposal with $\sigma = 0.1$
- ⋆ Run $K = 1000$ iterations



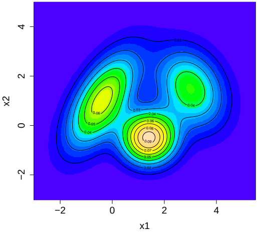



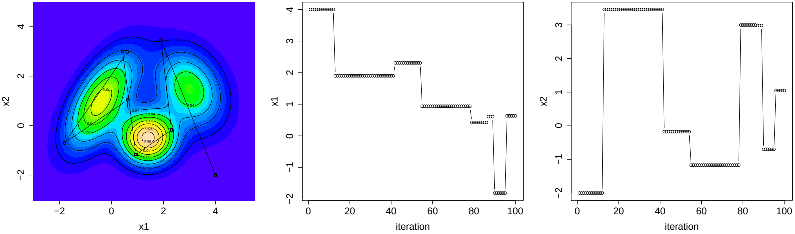



# Toy example: Random walk proposal

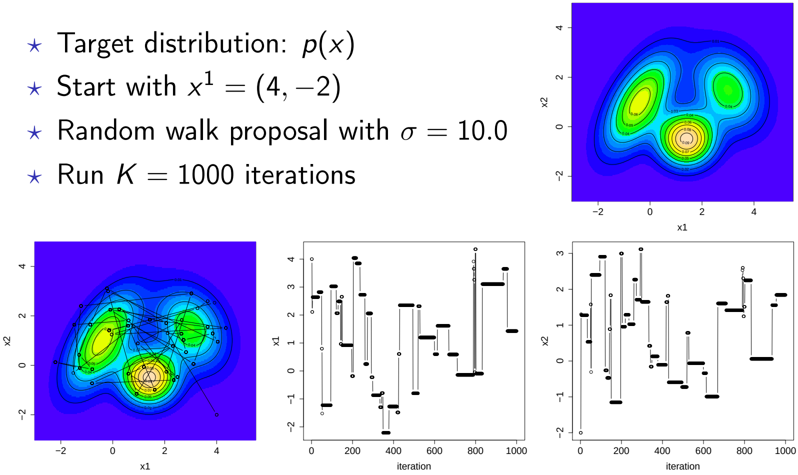
- ⋆ Target distribution: $p(x)$
- ⋆ Start with $x^1 = (4, -2)$
- ⋆ Random walk proposal with $\sigma = 10.0$
- ⋆ Run $K = 100$ iterations



# Toy example: Random walk proposal

- ⋆ Target distribution: $p(x)$
- ⋆ Start with $x^1 = (4, -2)$
- ⋆ Random walk proposal with $\sigma = 10.0$
- ⋆ Run $K = 1000$ iterations



# Gibbs sampling

- ⋆ Target distribution: $p(x)$
- ⋆ Proposal distribution: $z^k \sim q(z^k|x^{k-1})$
- ⋆ Acceptance probability:

$$\alpha(z^k|x^{k-1}) = \min\left\{1, \frac{p(z^k)}{p(x^{k-1})} \cdot \frac{q(x^{k-1}|z^k)}{q(z^k|x^{k-1})}\right\}$$

# Gibbs sampling

- ⋆ Target distribution: $p(x)$
- ⋆ Proposal distribution: $z^k \sim q(z^k|x^{k-1})$
- ⋆ Acceptance probability:

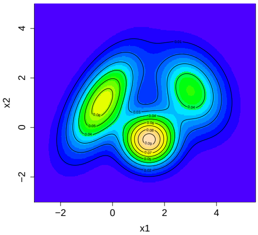
$$\alpha(z^k|x^{k-1}) = \min\left\{1, \frac{p(z^k)}{p(x^{k-1})} \cdot \frac{q(x^{k-1}|z^k)}{q(z^k|x^{k-1})}\right\}$$

- ⋆ Gibbs proposal:
  - sample $i \sim \text{Uniform}\{1, \ldots, n\}$
  - sample $z_i^k \sim p(x_i|x_1^{k-1}, \ldots, x_{i-1}^{k-1}, x_{i+1}^{k-1}, \ldots, x_n^{k-1})$
  - set $z_j^k = x_j^{k-1}$ for $j \neq i$
  - acceptance probability

$$\alpha(z^k|x^{k-1}) = 1$$

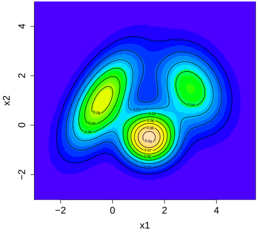
# Toy example: Gibbs sampling

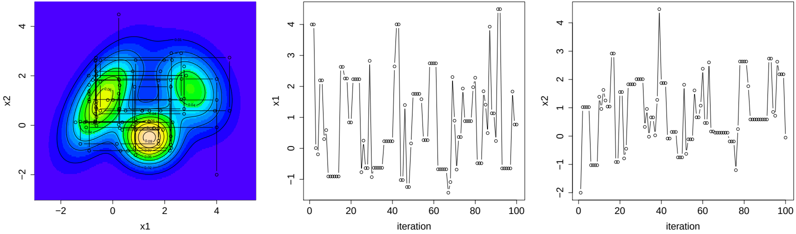
- ⋆ Target distribution: $p(x)$
- ⋆ Run $K = 100$ iterations



# Toy example: Gibbs sampling

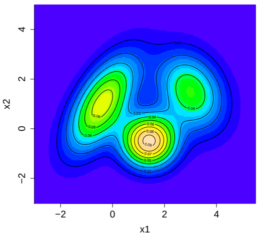
- ⋆ Target distribution: $p(x)$
- ⋆ Run $K = 100$ iterations

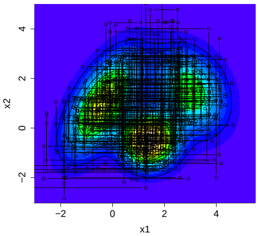


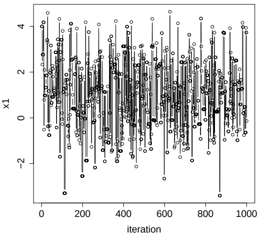



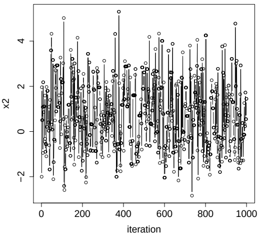
# Toy example: Gibbs sampling

- ⋆ Target distribution: $p(x)$
- ⋆ Run $K = 1000$ iterations



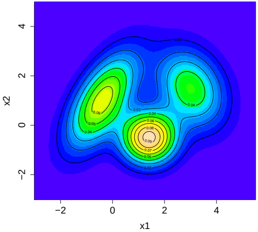



# Toy example: Combine Gibbs and random walk

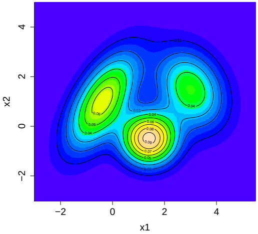
- ⋆ Target distribution: $p(x)$
- ⋆ Gibbs proposal for $x_1$
- ⋆ Random walk proposal for $x_2$

$$z_2^k = x_2^{k-1} + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, 1)$$

- ⋆ Run $K = 100$ iterations



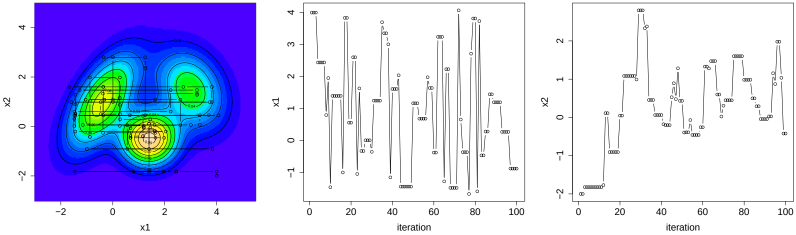
# Toy example: Combine Gibbs and random walk



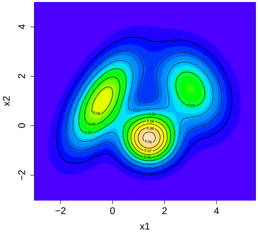

- Target distribution: $p(x)$
- Gibbs proposal for $x_1$
- Random walk proposal for $x_2$

$$z_2^k = x_2^{k-1} + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0, 1)$$

- Run $K = 100$ iterations



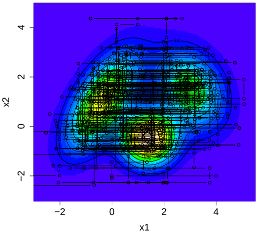
# Toy example: Combine Gibbs and random walk

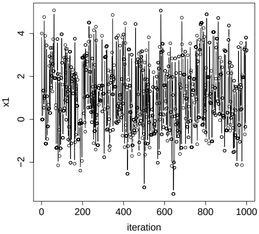


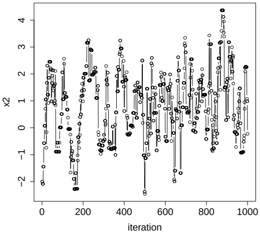

- ⋆ Target distribution: $p(x)$
- ⋆ Gibbs proposal for $x_1$
- ⋆ Random walk proposal for $x_2$

$$z_2^k = x_2^{k-1} + \varepsilon, \quad \varepsilon \sim \mathrm{N}(0,1)$$

- ⋆ Run $K = 1000$ iterations



# Convergence of the Metropolis–Hastings algorithm

- ⋆ Note: It is <u>not</u> $x^k$ that is converging
  - it is the distribution of $x^k$ that is converging
  - we can only look at and plot $x^k$
- ⋆ Common to evaluate convergence by plotting trace plots, i.e. $h(x)$

# Convergence of the Metropolis–Hastings algorithm

- Note: It is <u>not</u> $x^k$ that is converging
  - it is the distribution of $x^k$ that is converging
  - we can only look at and plot $x^k$
- Common to evaluate convergence by plotting trace plots, i.e. $h(x)$
- Experience:
  - for many distributions $p(x)$ simple proposal distribution gives reasonable quick convergence
  - the convergence can be extremely slow — not useful!
  - to choose a good proposal distribution for a given distribution often requires experience
  - typically a proposal distribution modify only a few components in $x^k$

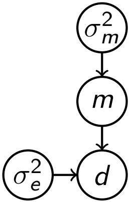
# Convergence of the Metropolis–Hastings algorithm

- Note: It is <u>not</u> $x^k$ that is converging
  - it is the distribution of $x^k$ that is converging
  - we can only look at and plot $x^k$
- Common to evaluate convergence by plotting trace plots, i.e. $h(x)$
- Experience:
  - for many distributions $p(x)$ simple proposal distribution gives reasonable quick convergence
  - the convergence can be extremely slow — not useful!
  - to choose a good proposal distribution for a given distribution often requires experience
  - typically a proposal distribution modify only a few components in $x^k$
- Reversible jump Metropolis–Hastings
  - also the dimension of $x$ is stochastic
  - the number of stochastic variables is a stochastic variable

# Inversion of seismic reflection coefficients

- Reference: Rabben, T.E., Tjelmeland, H. and Ursin, B. (2008). Non-linear Bayesian joint inversion of seismic reflection coefficients, Geophysical Journal International.
- Distribution of interest

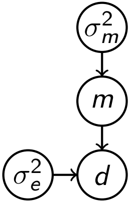
$$p(m, \sigma_m^2, \sigma_e^2 | d) \propto p(m, d, \sigma_m^2, \sigma_e^2)$$



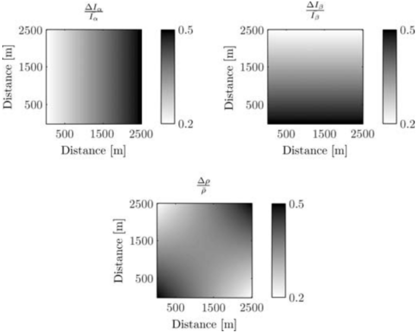
# Inversion of seismic reflection coefficients

- ⋆ Reference: Rabben, T.E., Tjelmeland, H. and Ursin, B. (2008). Non-linear Bayesian joint inversion of seismic reflection coefficients, Geophysical Journal International.
- ⋆ Distribution of interest

$$p(m, \sigma_m^2, \sigma_e^2 | d) \propto p(m, d, \sigma_m^2, \sigma_e^2)$$

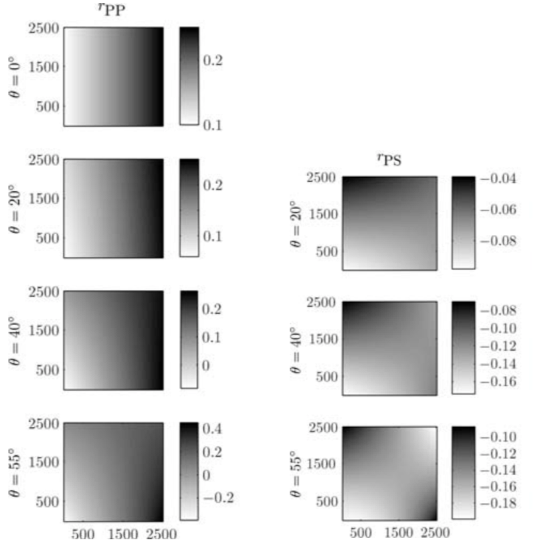



- ⋆ Simulation example: Reference contrasts



# Simulation example: Observations

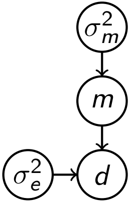
- ⋆ "Observed" PP and PS reflection coefficients using the Zoeppritz model



# Proposal distributions

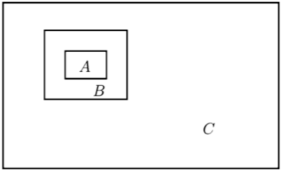
- Target distribution:

$$p(m, \sigma_m^2, \sigma_e^2 | d) \propto p(m, d, \sigma_m^2, \sigma_e^2)$$




- Proposal distributions:
  - Gibbs proposal for $\sigma_m^2$
  - Gibbs proposal for $\sigma_e^2$
  - proposal for $m_A$ based on a linearised model

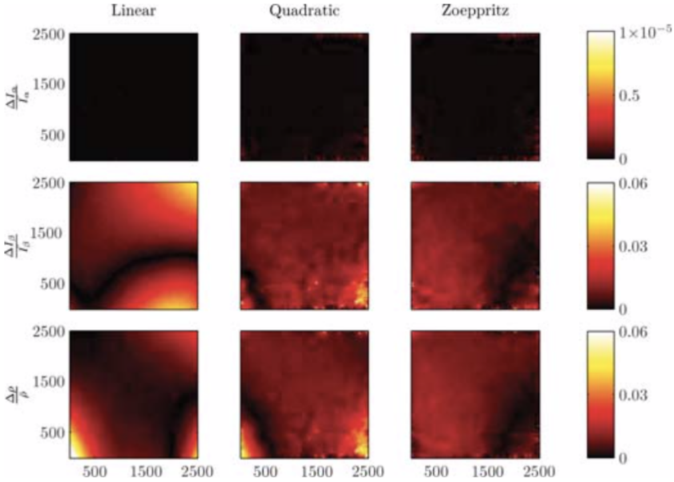
$$p^{\text{lin}}(m_A | m_B, \sigma_m^2, \sigma_e^2)$$




note: the acceptance probability is correcting the error done when proposing from a linearised model and by not conditioning on $m_c$

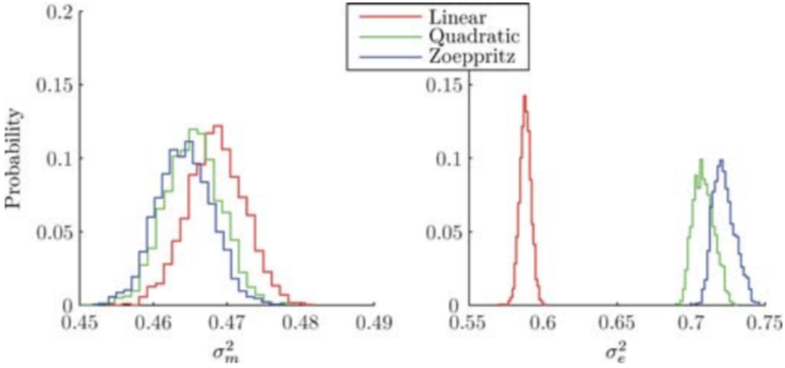
# Inversion of seismic reflection coefficients: Results

- ⋆ Resulting bias in inverted reflection coefficients
  - recall: results for three forward functions



# Inversion of seismic reflection coefficients: Results

- ⋆ Distribution of $\sigma_m^2$ and $\sigma_e^2$
  - recall: results for three forward functions



# Lithology/fluid class prediction using a Markov mesh prior

- ⋆ Reference: Tjelmeland, H., Luo, X. and Fjeldstad, T. (2019). A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image, Geophysical Prospecting.
- ⋆ Distribution of interest

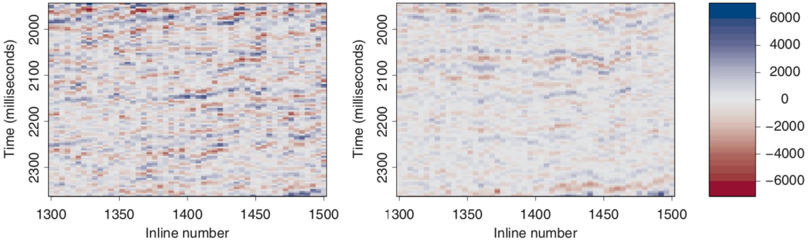
$$p(\kappa|\theta, d) \propto p(\kappa, d|\theta)$$

# Lithology/fluid class prediction using a Markov mesh prior

- Reference: Tjelmeland, H., Luo, X. and Fjeldstad, T. (2019). A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image, Geophysical Prospecting.
- Distribution of interest

$$p(\kappa|\theta, d) \propto p(\kappa, d|\theta)$$

- Near (left) and far (right) offset seismic data

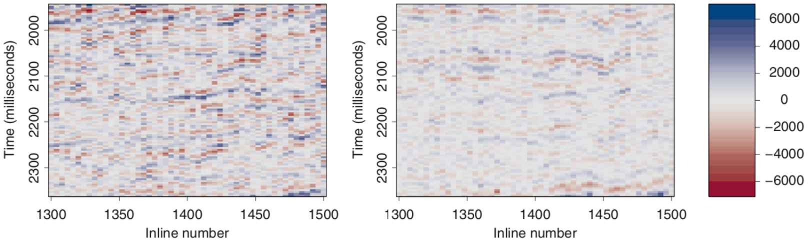


# Lithology/fluid class prediction using a Markov mesh prior

- Reference: Tjelmeland, H., Luo, X. and Fjeldstad, T. (2019). A Bayesian model for lithology/fluid class prediction using a Markov mesh prior fitted from a training image, Geophysical Prospecting.
- Distribution of interest

$$p(\kappa|\theta, d) \propto p(\kappa, d|\theta)$$

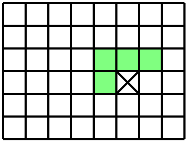
- Near (left) and far (right) offset seismic data

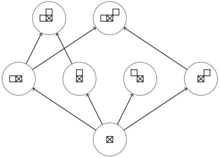



- First: Construction of the prior, $p(\kappa|\theta)$

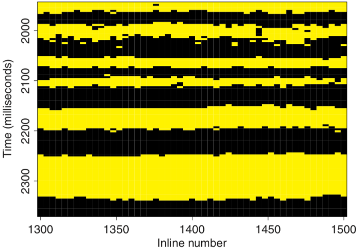
# Prior $p(\kappa)$

- ⋆ Recall: Markov mesh, $\theta = (\nu_i; \beta(\lambda), \nu_i \subseteq \nu_i)$





- ⋆ Training image

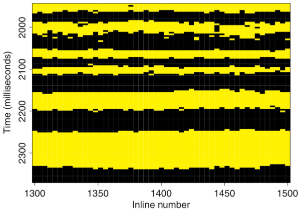



- ⋆ Use reversible jump Metropolis–Hastings to sample from

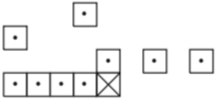
$$p(\theta|\text{training image}) \propto p(\theta)p(\text{training image}|\theta)$$

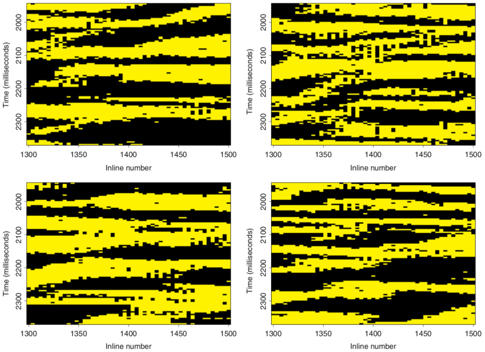
# The fitted prior

- ⋆ Training image




- ⋆ Neighbourhood and Monte Carlo samples



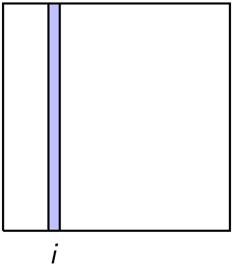


# Proposal distributions

- ⋆ Target distribution:

$$p(\kappa|\theta, d) \propto p(\kappa, d|\theta)$$

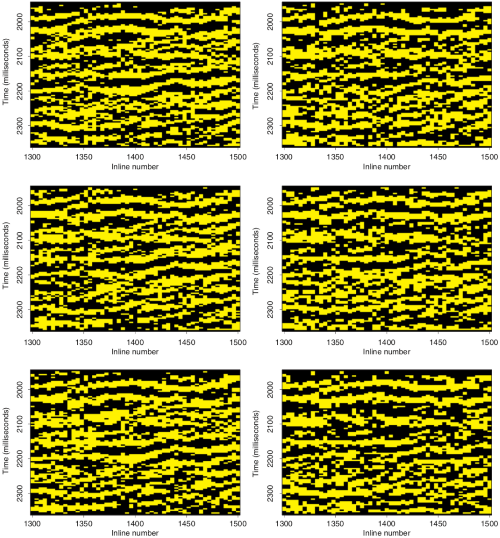
- ⋆ Proposal distributions
  - Gibbs proposal for $\kappa_i = (\kappa_{ij}, j = 1, \ldots, nx)$



# Lithology/fluid class prediction: Results

⋆ Monte Carlo samples from posterior

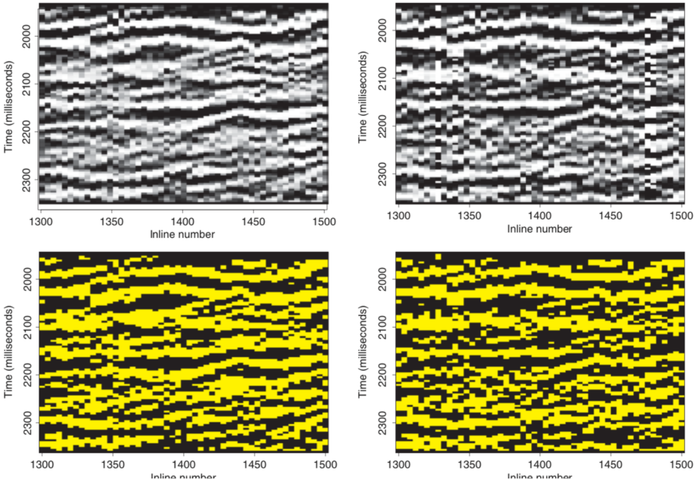
Markov mesh prior    Simpler prior



# Lithology/fluid class prediction: Results

⋆ Marginal probabilities and most probable value

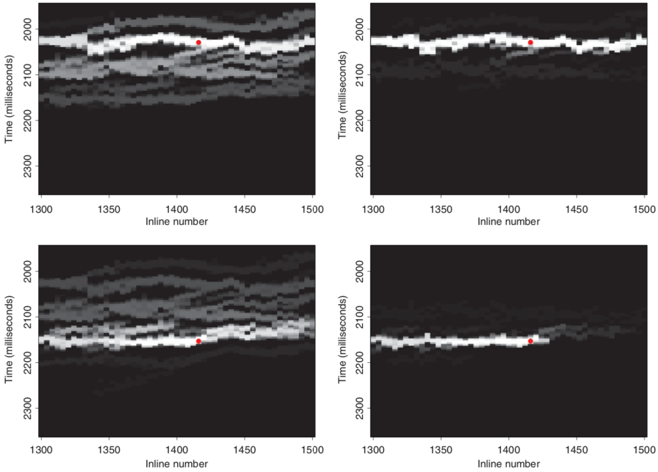
Markov mesh prior | Simpler prior



# Lithology/fluid class prediction: Results

⋆ Probabilities for connectivity

Markov mesh prior | Simpler prior



# Markov chain Monte Carlo summary

- Metropolis–Hastings is extremely flexible
  - can sample from almost any distribution
- Metropolis–Hastings can be reasonable fast or extremely slow
- If you can sample without using Metropolis–Hastings:
  - do not use Metropolis–Hastings
- If you cannot sample without using Metropolis–Hastings:
  - use Metropolis–Hastings

# Markov chain Monte Carlo summary

- Metropolis–Hastings is extremely flexible
  - can sample from almost any distribution
- Metropolis–Hastings can be reasonable fast or extremely slow
- If you can sample without using Metropolis–Hastings:
  - do not use Metropolis–Hastings
- If you cannot sample without using Metropolis–Hastings:
  - use Metropolis–Hastings

- Next Friday: Alternatives to MCMC for specific classes of models