# An approximate Bayesian geophysical inversion framework based on local-Gaussian likelihoods

Martin Jullum

Odd Kolbjørnsen

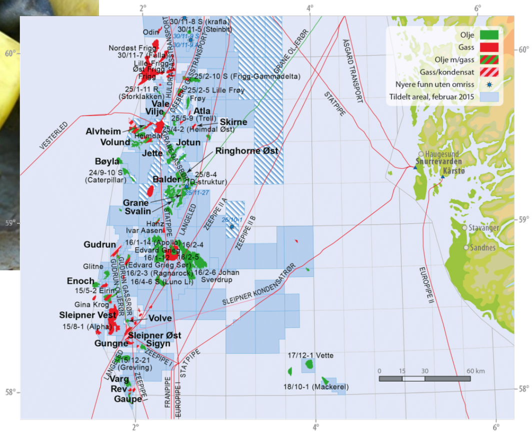
University of Oslo

martinju@math.uio.no

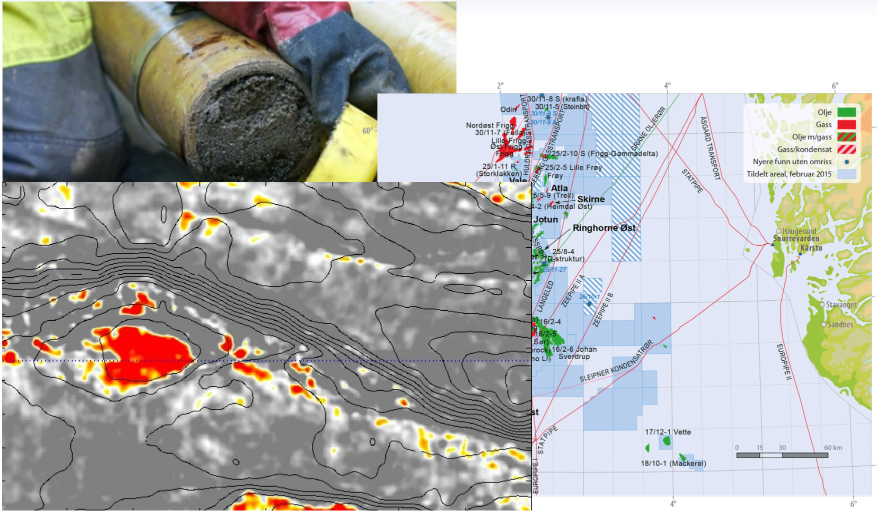
May 29, 2015

# Petroleum: Oil, gas etc.

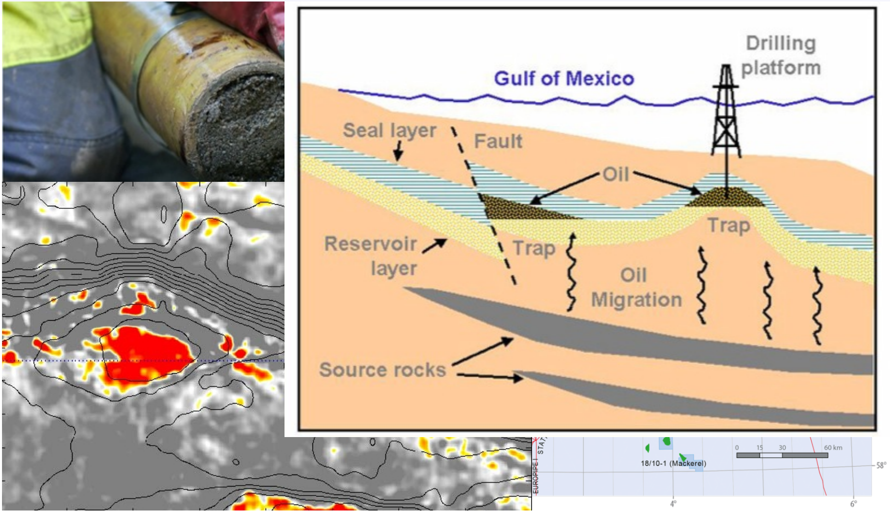
# Petroleum: Oil, gas etc.



# Petroleum: Oil, gas etc.

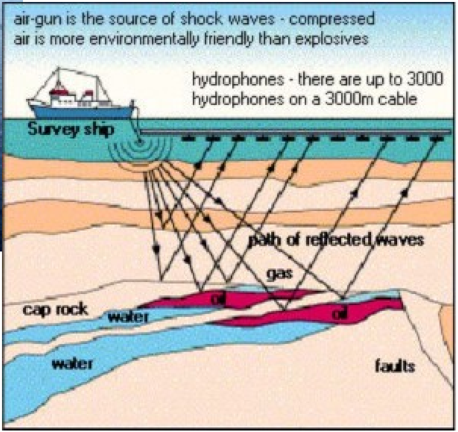


# Petroleum: Oil, gas etc.

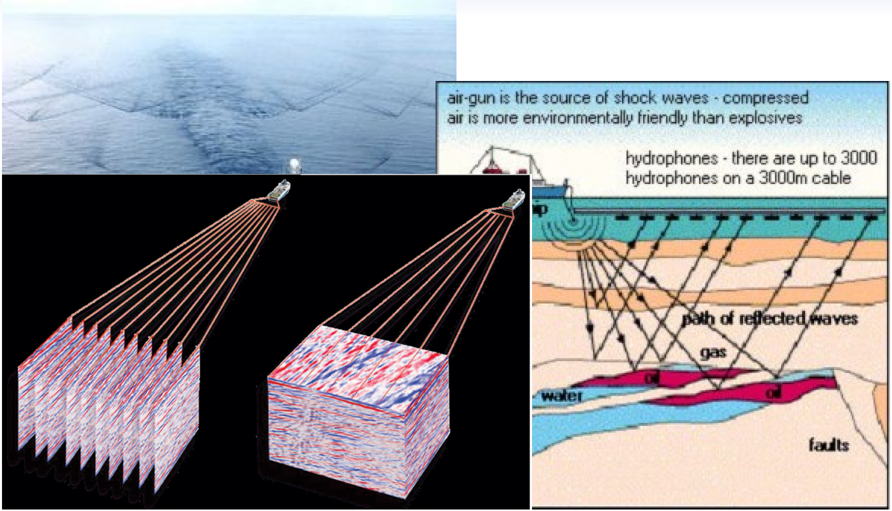


# Geophysical data: Seismic

# Geophysical data: Seismic



# Geophysical data: Seismic



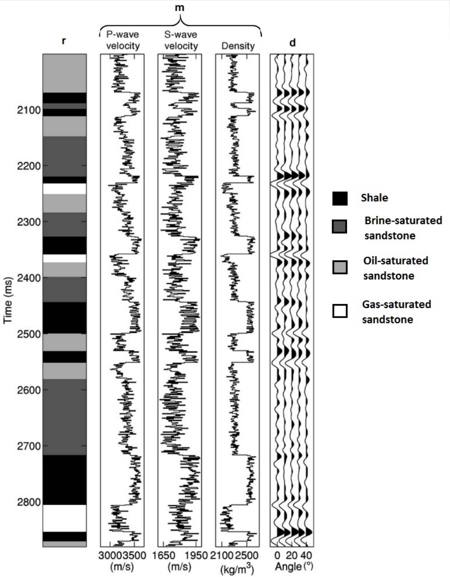
# Hierarchical model setup

Forward model

$$\mathbf{r} \rightarrow \mathbf{m} \rightarrow \mathbf{d}$$

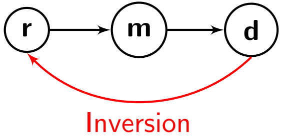
- Statistical approach
- Bayes is natural
  - Specify $p(\mathbf{r})$
  - Inversion $\Leftrightarrow$ consult $p(\mathbf{r}|\mathbf{d})$

- $\mathbf{r}$ = Rock properties/types
- $\mathbf{m}$ = Geophysical properties
- $\mathbf{d}$ = Geophysical data



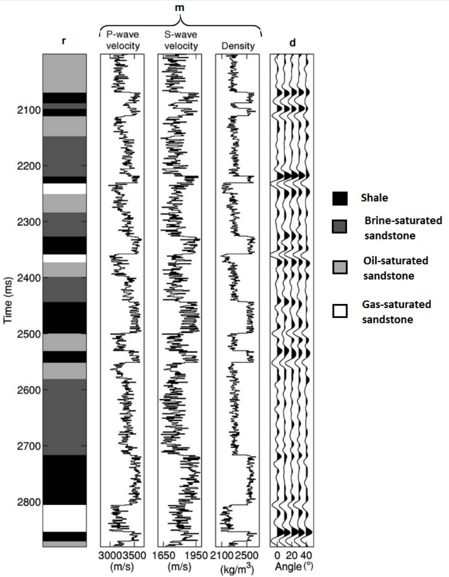
# Hierarchical model setup

Forward model

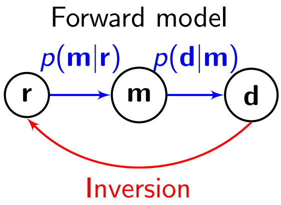



- Statistical approach
- Bayes is natural
  - Specify $p(\mathbf{r})$
  - Inversion $\Leftrightarrow$ consult $p(\mathbf{r}|\mathbf{d})$

- $\mathbf{r}$ = Rock properties/types
- $\mathbf{m}$ = Geophysical properties
- $\mathbf{d}$ = Geophysical data

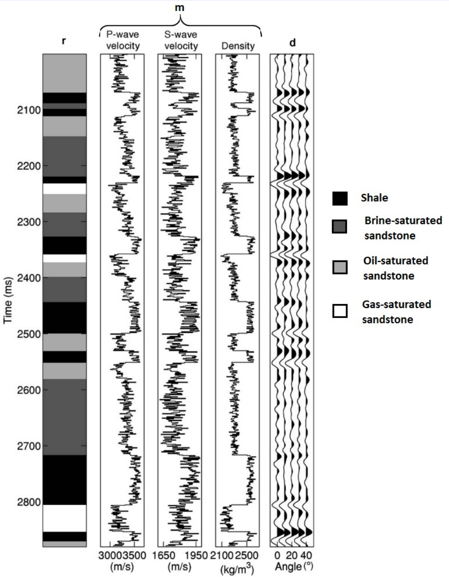


# Hierarchical model setup

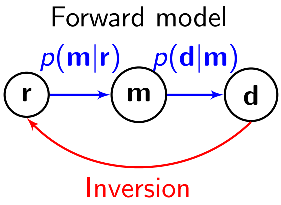


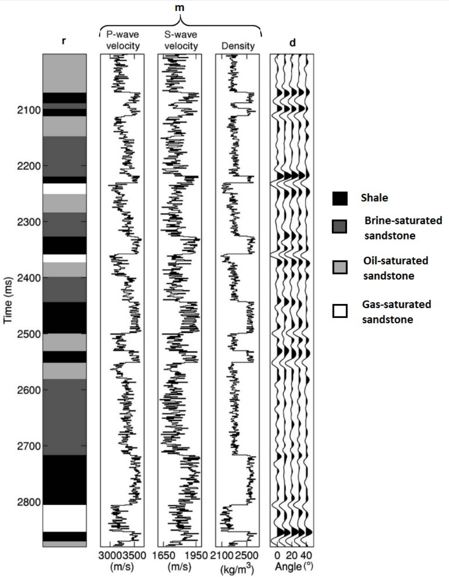

- Statistical approach
- Bayes is natural
  - Specify $p(\mathbf{r})$
  - Inversion $\Leftrightarrow$ consult $p(\mathbf{r}|\mathbf{d})$

- $\mathbf{r}$ = Rock properties/types
- $\mathbf{m}$ = Geophysical properties
- $\mathbf{d}$ = Geophysical data



# Hierarchical model setup



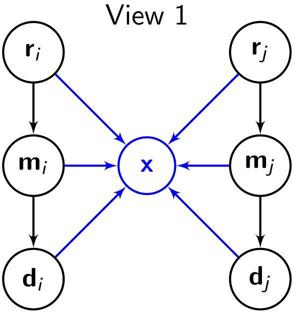

- Statistical approach
- Bayes is natural
  - Specify $p(\mathbf{r})$
  - Inversion $\Leftrightarrow$ consult $p(\mathbf{r}|\mathbf{d})$
- $\mathbf{r}$ = Rock properties/types
- $\mathbf{m}$ = Geophysical properties
- $\mathbf{d}$ = Geophysical data

# Typical problem

- Enormous amount of data
  - 5km $\times$ 5km $\times$ 2km (resolution 25m $\times$ 25m $\times$ 2m) $\Rightarrow 4 \cdot 10^6$ locations
- Wavelet smoothens the data, highly correlated data with complex dependency structures
- We are interested in the rock types/properties $r$ in ALL locations $t(i)$ at horizontal location $i = 1, \ldots, I$ and depth $t = 1, \ldots, T$.
  - Marginals $p(r_{t(i)}|\mathbf{d})$ typically 'sufficient'

# Typical working conditions

- Simplification: Horizontal dependencies are not modeled



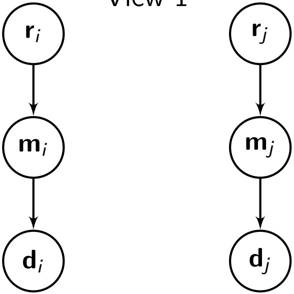

$\mathbf{x}$=State of the world

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma)$
- $p(\mathbf{m}_i|\mathbf{r}_i)$ and $p(\mathbf{r}_i)$ defined through sampling schemes

# Typical working conditions

- Simplification: Horizontal dependencies are not modeled

View 1



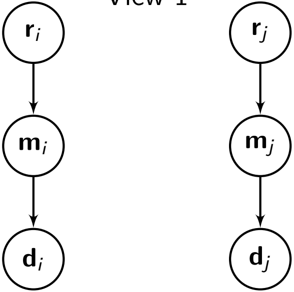

$\mathbf{x}$=State of the world

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma)$
- $p(\mathbf{m}_i|\mathbf{r}_i)$ and $p(\mathbf{r}_i)$ defined through sampling schemes

# Typical working conditions

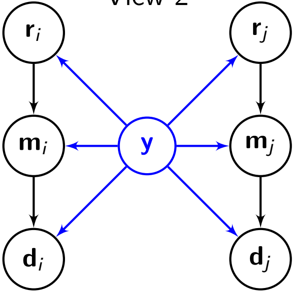
- Simplification: Horizontal dependencies are not modeled

View 1




$\mathbf{x}$=State of the world

View 2

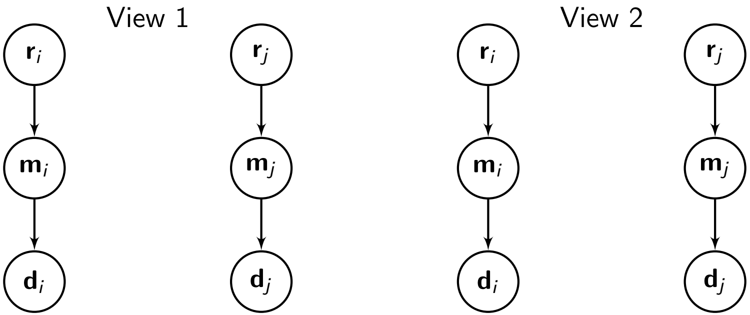



$\mathbf{y}$=Common feature

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma)$
- $p(\mathbf{m}_i|\mathbf{r}_i)$ and $p(\mathbf{r}_i)$ defined through sampling schemes

# Typical working conditions

- Simplification: Horizontal dependencies are not modeled




$\mathbf{x}$=State of the world

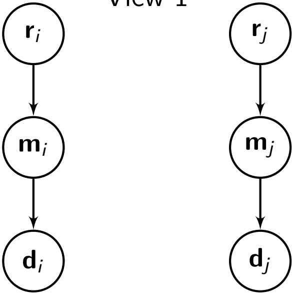
$\mathbf{y}$=Common feature

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma)$
- $p(\mathbf{m}_i|\mathbf{r}_i)$ and $p(\mathbf{r}_i)$ defined through sampling schemes

# Typical working conditions

- Simplification: Horizontal dependencies are not modeled

View 1




$\mathbf{x}$=State of the world

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma)$
- $p(\mathbf{m}_i|\mathbf{r}_i)$ and $p(\mathbf{r}_i)$ defined through sampling schemes

# Possible approaches

Full profile posterior:

- $p(\mathbf{r}_i|\mathbf{d}_i) \propto p(\mathbf{d}_i|\mathbf{r}_i)p(\mathbf{r}_i) = \int p(\mathbf{d}_i|\mathbf{m}_i)p(\mathbf{m}_i|\mathbf{r}_i)\,\mathrm{d}\mathbf{m}_i p(\mathbf{r}_i)$

Marginal posterior:

- $p(r_{t(i)}|\mathbf{d}_i) \propto p(\mathbf{d}_i|r_{t(i)})p(r_{t(i)}) = \int p(\mathbf{d}_i|\mathbf{r}_i)p(\mathbf{r}_i)\,\mathrm{d}\mathbf{r}_{t(-i)} = \int \left[\int p(\mathbf{d}_i|\mathbf{m}_i)p(\mathbf{m}_i|\mathbf{r}_i)\,\mathrm{d}\mathbf{m}_i\right] p(\mathbf{r}_i)\,\mathrm{d}\mathbf{r}_{t(-i)}$

- Exact computation?
- MCMC?
- Variational Bayes/Expectation Propagation?
- ABC?
- INLA?
- 'Everything Gaussian' Approximation?

# Possible approaches

Full profile posterior:

- $p(\mathbf{r}_i|\mathbf{d}_i) \propto p(\mathbf{d}_i|\mathbf{r}_i)p(\mathbf{r}_i) = \int p(\mathbf{d}_i|\mathbf{m}_i)p(\mathbf{m}_i|\mathbf{r}_i)\,\mathrm{d}\mathbf{m}_i p(\mathbf{r}_i)$

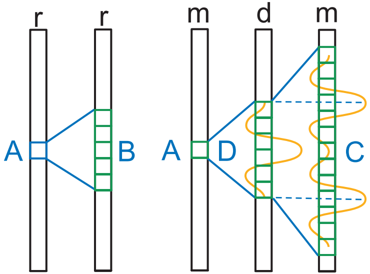
Marginal posterior:

- $p(r_{t(i)}|\mathbf{d}_i) \propto p(\mathbf{d}_i|r_{t(i)})p(r_{t(i)}) = \int p(\mathbf{d}_i|\mathbf{r}_i)p(\mathbf{r}_i)\,\mathrm{d}\mathbf{r}_{t(-i)} =$
  $\int \left[\int p(\mathbf{d}_i|\mathbf{m}_i)p(\mathbf{m}_i|\mathbf{r}_i)\,\mathrm{d}\mathbf{m}_i\right] p(\mathbf{r}_i)\,\mathrm{d}\mathbf{r}_{t(-i)}$

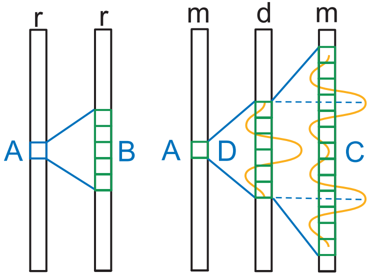
- Exact computation?
- MCMC?
- Variational Bayes/Expectation Propagation?
- ABC?
- INLA?
- ‘Everything Gaussian’ Approximation?

# Our solution: Local-Gaussian compound likelihoods I

- Let $A = t(i)$
- Define local subsets: $B = B(A), C = C(A), D = D(A)$
- $p(r_A|\mathbf{d}_i) \approx p(r_A|\mathbf{d}_D)$
- $p(r_A|\mathbf{d}_D) \propto$ $\int p(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)} =$ $\int \left[\int p(\mathbf{d}_D|\mathbf{m}_C)p(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C\right] p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)}$.



# Our solution: Local-Gaussian compound likelihoods I




- Let $A = t(i)$
- Define local subsets: $B = B(A), C = C(A), D = D(A)$
- $p(r_A|\mathbf{d}_i) \approx p(r_A|\mathbf{d}_D)$
- $p(r_A|\mathbf{d}_D) \propto$
  $\int p(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)} =$
  $\int \left[\int p(\mathbf{d}_D|\mathbf{m}_C)p(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C\right] p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)}.$

# Our solution: Local-Gaussian compound likelihoods II

Need to handle: $p(\mathbf{d}_D|\mathbf{r}_B) = \int p(\mathbf{d}_D|\mathbf{m}_C)p(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma) \Rightarrow p(\mathbf{d}_D|\mathbf{m}_i) \sim N(G_D m, \Sigma_{DD})$
  - $p(\mathbf{d}_D|\mathbf{m}_C) \approx p^*(\mathbf{d}_D|\mathbf{m}_C) \sim N(G_{DC}\mathbf{m}_C, \Sigma_{DD})$
- $p(\mathbf{m}_C|\mathbf{r}_B) \approx p^*(\mathbf{m}_C|\mathbf{r}_B) \sim N(\mu(\mathbf{r}_B), \Sigma(k)), k = k(\mathbf{r}_B)$
  - Sample lots of pairs $(\mathbf{m}_C, \mathbf{r}_B)$ from $p(\mathbf{m}_C, \mathbf{r}_B)$
  - Use flexible regression scheme (MARS, Projection pursuit etc.) and fit $\mu(\mathbf{r}_B)$ in $\mathbf{m}_C = \mu(\mathbf{r}_B) + \varepsilon$
  - Divide residuals $\varepsilon$ into groups $k = k(\mathbf{r}_B) \in \{1, \ldots, K\}$ and fit separate $\Sigma(k)$ with range spanning covariance estimation routine*
- Approximation: $p(\mathbf{d}_D|\mathbf{r}_B) \approx \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C = p^*(\mathbf{d}_D|\mathbf{r}_B) \sim N(G_{DC}\mu(\mathbf{r}_B), \Sigma_{DD} + G_{DC}\Sigma(k)G_{CD})$
- $p(r_A|\mathbf{d}_D) = p^*(r_A|\mathbf{d}_D) = \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)}$
  - Weighted Monte Carlo approach: Sample from $p(\mathbf{r}_B)$, weight corresponding $r_A$ by $p^*(\mathbf{d}_D|\mathbf{r}_B)$ and normalize
  - Properly aggregate weighted samples to approx. distribution quantities

# Our solution: Local-Gaussian compound likelihoods II

Need to handle: $p(\mathbf{d}_D|\mathbf{r}_B) = \int p(\mathbf{d}_D|\mathbf{m}_C)p(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma) \Rightarrow p(\mathbf{d}_D|\mathbf{m}_i) \sim N(G_D m, \Sigma_{DD})$
  - $p(\mathbf{d}_D|\mathbf{m}_C) \approx p^*(\mathbf{d}_D|\mathbf{m}_C) \sim N(G_{DC}\mathbf{m}_C, \Sigma_{DD})$
- $p(\mathbf{m}_C|\mathbf{r}_B) \approx p^*(\mathbf{m}_C|\mathbf{r}_B) \sim N(\mu(\mathbf{r}_B), \Sigma(k)), k = k(\mathbf{r}_B)$
  - Sample lots of pairs $(\mathbf{m}_C, \mathbf{r}_B)$ from $p(\mathbf{m}_C, \mathbf{r}_B)$
  - Use flexible regression scheme (MARS, Projection pursuit etc.) and fit $\mu(\mathbf{r}_B)$ in $\mathbf{m}_C = \mu(\mathbf{r}_B) + \varepsilon$
  - Divide residuals $\varepsilon$ into groups $k = k(\mathbf{r}_B) \in \{1, \ldots, K\}$ and fit separate $\Sigma(k)$ with range spanning covariance estimation routine*
- Approximation: $p(\mathbf{d}_D|\mathbf{r}_B) \approx \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C = p^*(\mathbf{d}_D|\mathbf{r}_B) \sim N(G_{DC}\mu(\mathbf{r}_B), \Sigma_{DD} + G_{DC}\Sigma(k)G_{CD})$
- $p(r_A|\mathbf{d}_D) = p^*(r_A|\mathbf{d}_D) = \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)}$
  - Weighted Monte Carlo approach: Sample from $p(\mathbf{r}_B)$, weight corresponding $r_A$ by $p^*(\mathbf{d}_D|\mathbf{r}_B)$ and normalize
  - Properly aggregate weighted samples to approx. distribution quantities

# Our solution: Local-Gaussian compound likelihoods II

Need to handle: $p(\mathbf{d}_D|\mathbf{r}_B) = \int p(\mathbf{d}_D|\mathbf{m}_C)p(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma) \Rightarrow p(\mathbf{d}_D|\mathbf{m}_i) \sim N(G_D m, \Sigma_{DD})$
  - $p(\mathbf{d}_D|\mathbf{m}_C) \approx p^*(\mathbf{d}_D|\mathbf{m}_C) \sim N(G_{DC}\mathbf{m}_C, \Sigma_{DD})$
- $p(\mathbf{m}_C|\mathbf{r}_B) \approx p^*(\mathbf{m}_C|\mathbf{r}_B) \sim N(\mu(\mathbf{r}_B), \Sigma(k)), k = k(\mathbf{r}_B)$
  - Sample lots of pairs $(\mathbf{m}_C, \mathbf{r}_B)$ from $p(\mathbf{m}_C, \mathbf{r}_B)$
  - Use flexible regression scheme (MARS, Projection pursuit etc.) and fit $\mu(\mathbf{r}_B)$ in $\mathbf{m}_C = \mu(\mathbf{r}_B) + \varepsilon$
  - Divide residuals $\varepsilon$ into groups $k = k(\mathbf{r}_B) \in \{1, \dots, K\}$ and fit separate $\Sigma(k)$ with range spanning covariance estimation routine*
- Approximation: $p(\mathbf{d}_D|\mathbf{r}_B) \approx \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C = p^*(\mathbf{d}_D|\mathbf{r}_B) \sim N(G_{DC}\mu(\mathbf{r}_B), \Sigma_{DD} + G_{DC}\Sigma(k)G_{CD})$
- $p(r_A|\mathbf{d}_D) = p^*(r_A|\mathbf{d}_D) = \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)}$
  - Weighted Monte Carlo approach: Sample from $p(\mathbf{r}_B)$, weight corresponding $r_A$ by $p^*(\mathbf{d}_D|\mathbf{r}_B)$ and normalize
  - Properly aggregate weighted samples to approx. distribution quantities

# Our solution: Local-Gaussian compound likelihoods II

Need to handle: $p(\mathbf{d}_D|\mathbf{r}_B) = \int p(\mathbf{d}_D|\mathbf{m}_C)p(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$

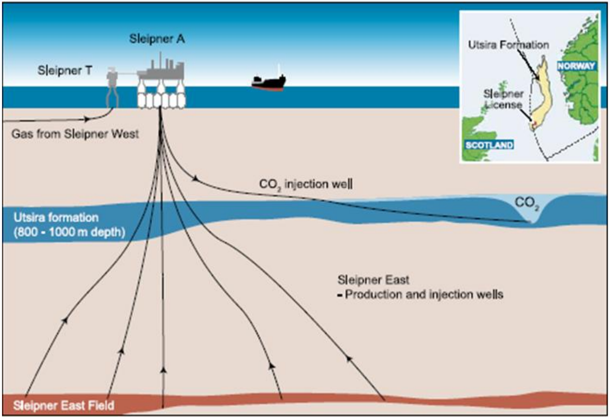
- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma) \Rightarrow p(\mathbf{d}_D|\mathbf{m}_i) \sim N(G_D m, \Sigma_{DD})$
  - $p(\mathbf{d}_D|\mathbf{m}_C) \approx p^*(\mathbf{d}_D|\mathbf{m}_C) \sim N(G_{DC}\mathbf{m}_C, \Sigma_{DD})$
- $p(\mathbf{m}_C|\mathbf{r}_B) \approx p^*(\mathbf{m}_C|\mathbf{r}_B) \sim N(\mu(\mathbf{r}_B), \Sigma(k)), k = k(\mathbf{r}_B)$
  - Sample lots of pairs $(\mathbf{m}_C, \mathbf{r}_B)$ from $p(\mathbf{m}_C, \mathbf{r}_B)$
  - Use flexible regression scheme (MARS, Projection pursuit etc.) and fit $\mu(\mathbf{r}_B)$ in $\mathbf{m}_C = \mu(\mathbf{r}_B) + \varepsilon$
  - Divide residuals $\varepsilon$ into groups $k = k(\mathbf{r}_B) \in \{1, \ldots, K\}$ and fit separate $\Sigma(k)$ with range spanning covariance estimation routine*
- Approximation: $p(\mathbf{d}_D|\mathbf{r}_B) \approx \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C = p^*(\mathbf{d}_D|\mathbf{r}_B) \sim N(G_{DC}\mu(\mathbf{r}_B), \Sigma_{DD} + G_{DC}\Sigma(k)G_{CD})$
- $p(r_A|\mathbf{d}_D) = p^*(r_A|\mathbf{d}_D) = \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)}$
  - Weighted Monte Carlo approach: Sample from $p(\mathbf{r}_B)$, weight corresponding $r_A$ by $p^*(\mathbf{d}_D|\mathbf{r}_B)$ and normalize
  - Properly aggregate weighted samples to approx. distribution quantities

# Our solution: Local-Gaussian compound likelihoods II

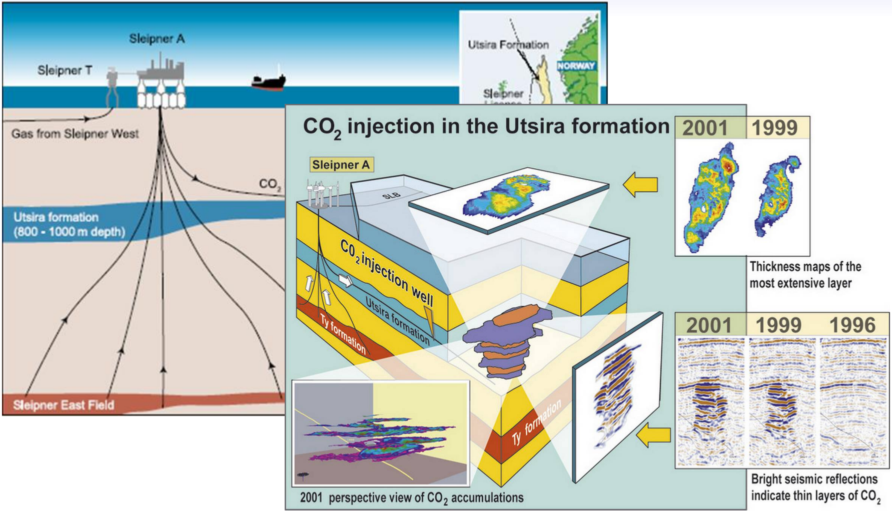
Need to handle: $p(\mathbf{d}_D|\mathbf{r}_B) = \int p(\mathbf{d}_D|\mathbf{m}_C)p(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C$

- $p(\mathbf{d}_i|\mathbf{m}_i) \sim N(Gm, \Sigma) \Rightarrow p(\mathbf{d}_D|\mathbf{m}_i) \sim N(G_D m, \Sigma_{DD})$
  - $p(\mathbf{d}_D|\mathbf{m}_C) \approx p^*(\mathbf{d}_D|\mathbf{m}_C) \sim N(G_{DC}\mathbf{m}_C, \Sigma_{DD})$
- $p(\mathbf{m}_C|\mathbf{r}_B) \approx p^*(\mathbf{m}_C|\mathbf{r}_B) \sim N(\mu(\mathbf{r}_B), \Sigma(k)), k = k(\mathbf{r}_B)$
  - Sample lots of pairs $(\mathbf{m}_C, \mathbf{r}_B)$ from $p(\mathbf{m}_C, \mathbf{r}_B)$
  - Use flexible regression scheme (MARS, Projection pursuit etc.) and fit $\mu(\mathbf{r}_B)$ in $\mathbf{m}_C = \mu(\mathbf{r}_B) + \varepsilon$
  - Divide residuals $\varepsilon$ into groups $k = k(\mathbf{r}_B) \in \{1, \dots, K\}$ and fit separate $\Sigma(k)$ with range spanning covariance estimation routine*
- Approximation: $p(\mathbf{d}_D|\mathbf{r}_B) \approx \int p^*(\mathbf{d}_D|\mathbf{m}_C)p^*(\mathbf{m}_C|\mathbf{r}_B)\,\mathrm{d}\mathbf{m}_C = p^*(\mathbf{d}_D|\mathbf{r}_B) \sim N(G_{DC}\mu(\mathbf{r}_B), \Sigma_{DD} + G_{DC}\Sigma(k)G_{CD})$
- $p(r_A|\mathbf{d}_D) = p^*(r_A|\mathbf{d}_D) = \int p^*(\mathbf{d}_D|\mathbf{r}_B)p(\mathbf{r}_B)\,\mathrm{d}\mathbf{r}_{B(-A)}$
  - Weighted Monte Carlo approach: Sample from $p(\mathbf{r}_B)$, weight corresponding $r_A$ by $p^*(\mathbf{d}_D|\mathbf{r}_B)$ and normalize
  - Properly aggregate weighted samples to approx. distribution quantities

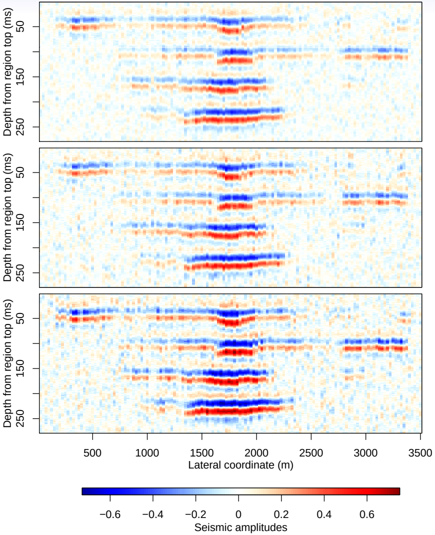
# Illustration: $CO_2$ monitoring $\Rightarrow$ map saturation

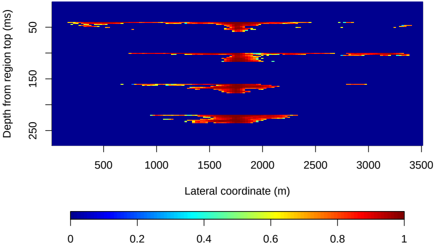


# Illustration: $CO_2$ monitoring $\Rightarrow$ map saturation

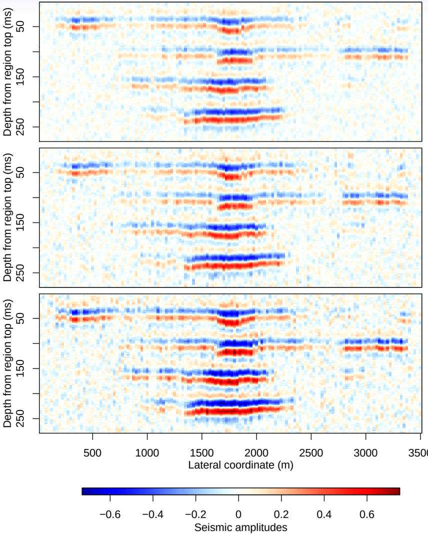


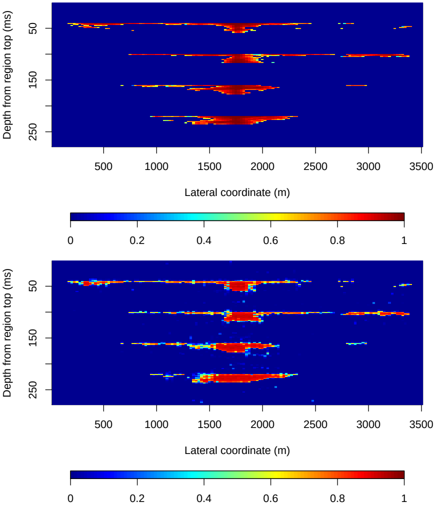
# Illustration: Synthetic case



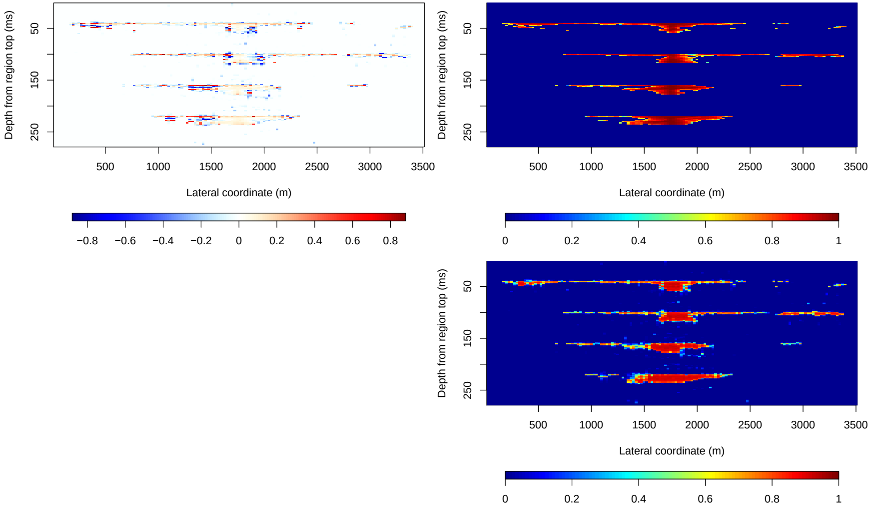



# Illustration: Synthetic case

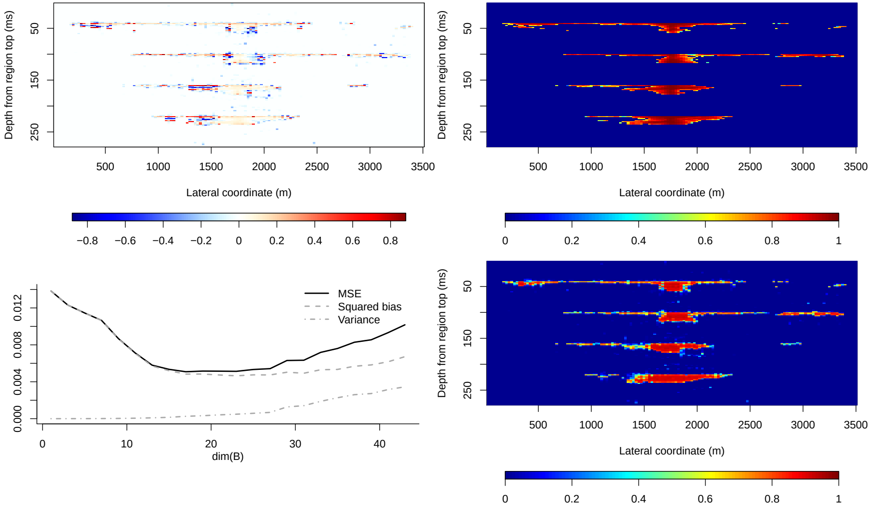


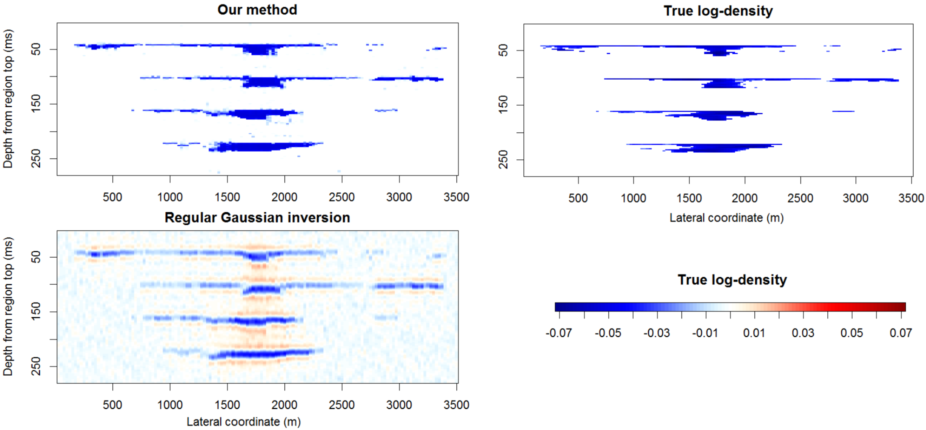
# Illustration: Synthetic case



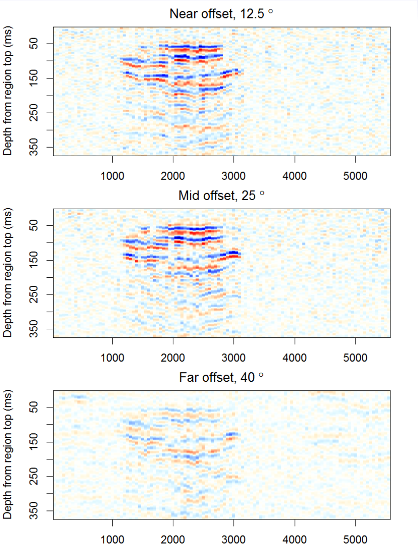
# Illustration: Synthetic case



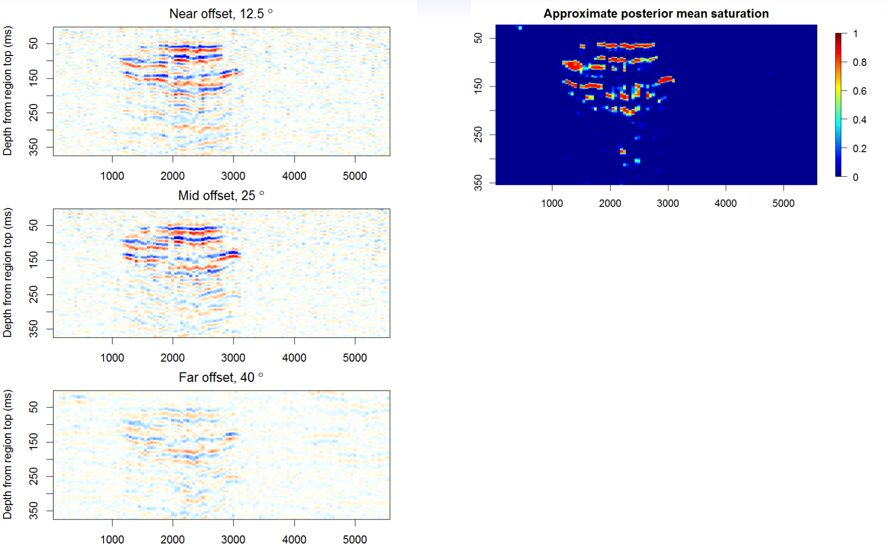
# Illustration: Synthetic case



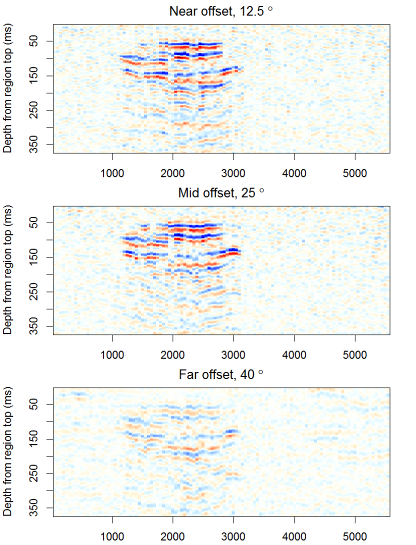
# Illustration: Real case

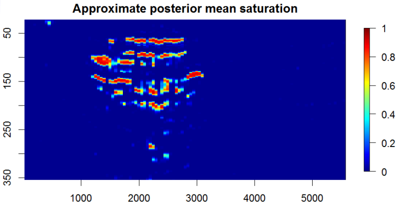


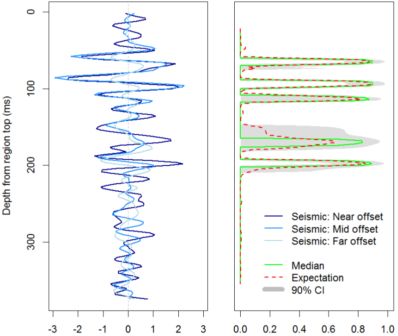
# Illustration: Real case



# Illustration: Real case



# Concluding remarks

Approximation ingredients:

- Compound local-Gaussian likelihood approximation
  - Linear Gaussian approx. directly from model knowledge + non-linear sampling based approx.
- Selecting/tuning local subset parameters (training on synthetic data)
- Weighted Monte Carlo sampling routine

- Some connections to INLA
- May approximate realistic models directly
- Well suited for parallellization (18 000 cells in 30' on 8 cored Windows laptop using plain R-programming)
- Application: Clearly improves upon common methodology

# Concluding remarks

Approximation ingredients:

- Compound local-Gaussian likelihood approximation
  - Linear Gaussian approx. directly from model knowledge + non-linear sampling based approx.
- Selecting/tuning local subset parameters (training on synthetic data)
- Weighted Monte Carlo sampling routine

- Some connections to INLA
- May approximate realistic models directly
- Well suited for parallellization (18 000 cells in 30' on 8 cored Windows laptop using plain R-programming)
- Application: Clearly improves upon common methodology