Partially conditional blocks approximations used in MCMC

Ingelin Steinsland & Håvard Rue

ingelins@math.ntnu.no

Norwegian University of Science and Technology

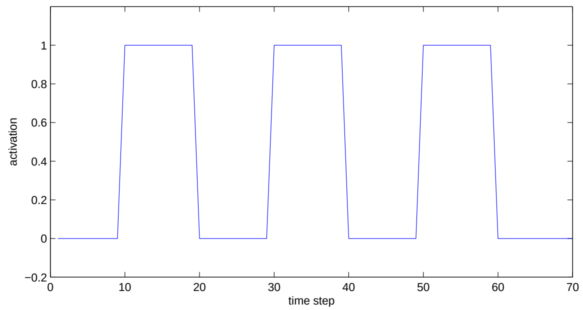
Outline

- Introduction functional Magnetic Resonance Imaging (fMRI) problem.
- Latent spatial Markov models.
- One-block Metropolis-Hastings algorithm.
- Partially conditional blocks approximations
- Results fMRI problem.
- Closing remarks.

fMRI

functional Magnetic Resonance Imaging -
Data from a visual stimulation experiment.

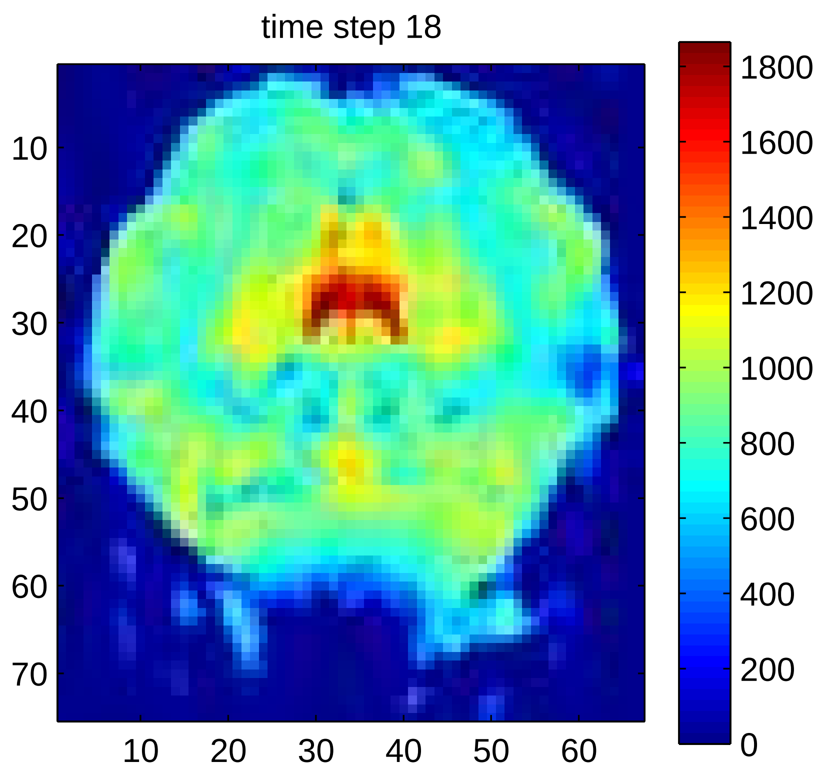
- Stimulus: 8 Hz flickering checkerboard.
- 4 periods (a 30 sec.) rest, 3 periods stimulus.

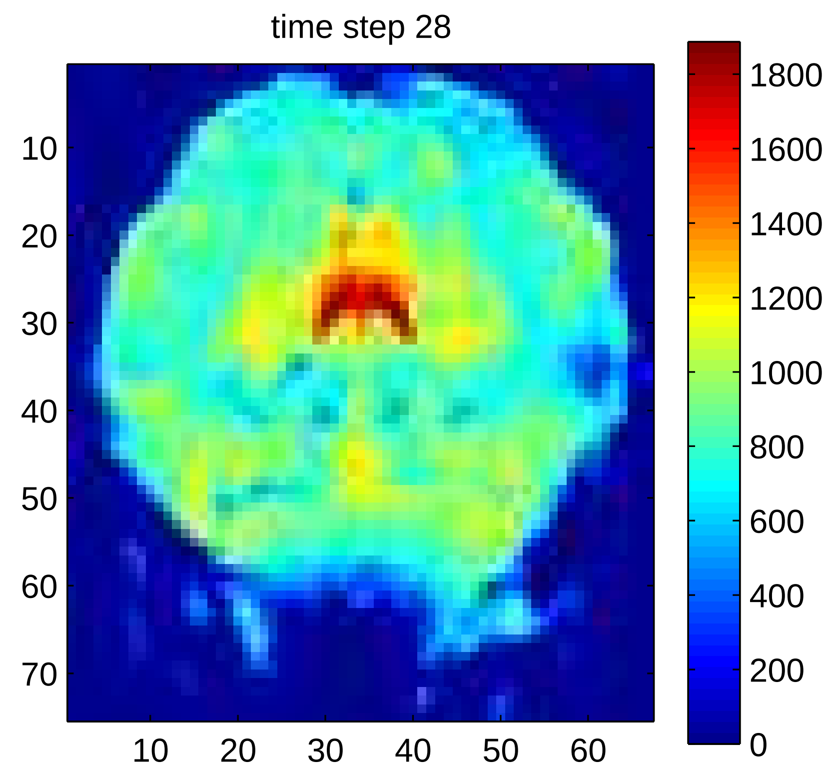


- Cross section of the brain observed every 3rd sec.
- Observe BOLD effects.

fMRI

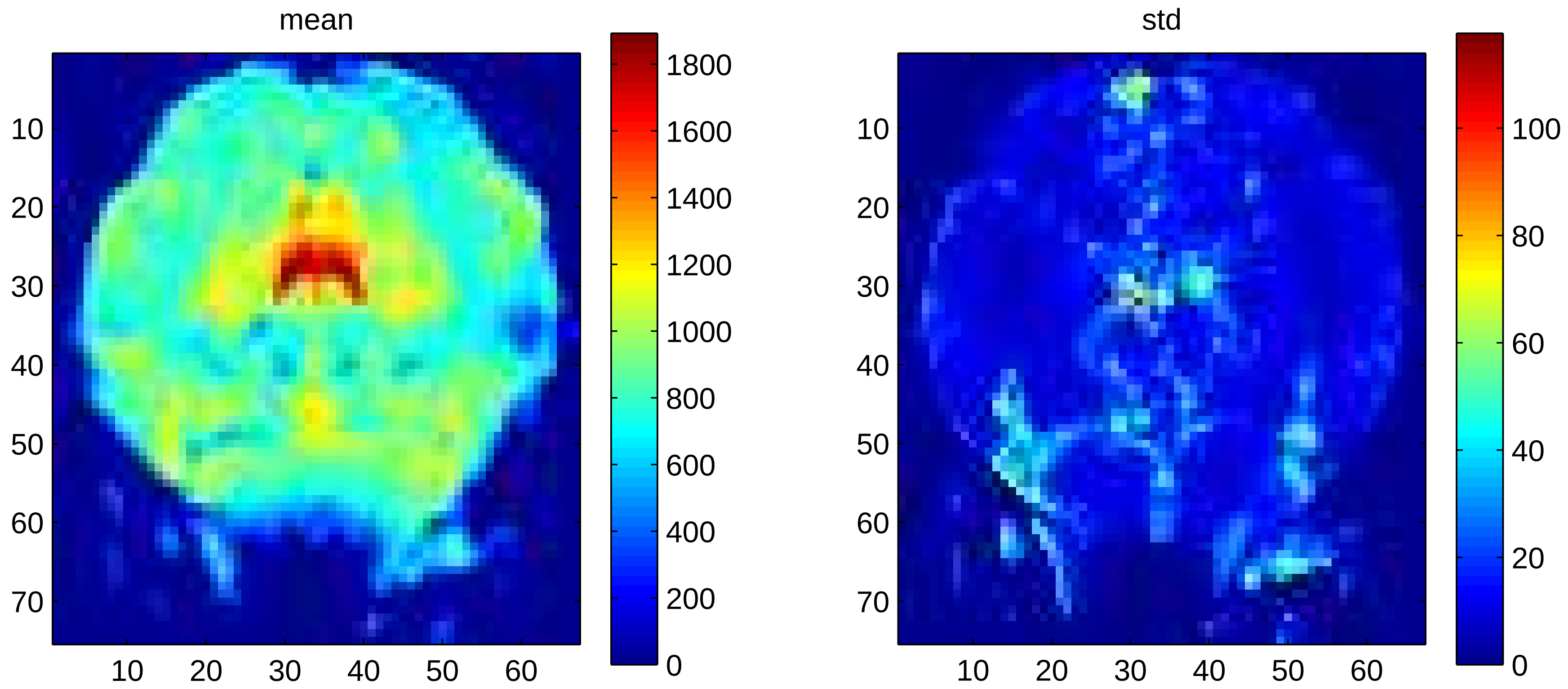
functional Magnetic Resonance Imaging -
Data from a visual stimulation experiment.





fMRI

functional Magnetic Resonance Imaging -
Data from a visual stimulation experiment.



Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$

Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67$, $t = 1, \ldots 70$
- a_i: Baseline image, pixel i, $i = 1, \ldots, 75 \times 67$

Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67, t = 1, \ldots 70$
- a_i: Baseline image, pixel i, $i = 1, \ldots, 75 \times 67$
- z_t: Transformed stimulus at time step t,
  $t = 1, \ldots 70$
- b_{it}: Activation effect of pixel i at time step t,
  $i = 1, \ldots, 75 \times 67, t = 1, \ldots 70$

Model fMRI

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- y_{it}: Data in pixel i at time step t
  $i = 1, \ldots, 75 \times 67, t = 1, \ldots 70$
- a_i: Baseline image, pixel i, $i = 1, \ldots, 75 \times 67$
- z_t: Transformed stimulus at time step t,
  $t = 1, \ldots 70$
- b_{it}: Activation effect of pixel i at time step t,
  $i = 1, \ldots, 75 \times 67, t = 1, \ldots 70$
- ϵ_{it}: Measurement error of pixel i at time step t
  $i = 1, \ldots, 75 \times 67, t = 1, \ldots 70$

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $\epsilon \sim N(0, \tau_{Data} I) \rightarrow \; y_{it} | a, b \sim N(a_i + z_t b_{it}, \tau_{Data})$

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $y_{it}|a, b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- z; use estimate from similar studies

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $y_{it} | a, b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- GMRF (Gaussian Markov random field) for a:
  $\pi(a) \propto \exp(-\frac{1}{2}\tau_A \sum_{i \overset{s}{\sim} j}(a_i - a_j)^2)$

Model specification

$$y_{it} = a_i + z_t b_{it} + \epsilon_{it}$$

- $y_{it}|a,b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- GMRF (Gaussian Markov random field) for a:
  $\pi(a) \propto \exp(-\frac{1}{2}\tau_A \sum_{i \overset{s}{\sim} j}(a_i - a_j)^2)$
- Time-space GMRF for b:

$$\begin{aligned}\pi(b) \quad \propto \quad & \exp(-\frac{1}{2}\tau_B \sum_{t=1}^{T} \sum_{i \overset{s}{\sim} j} (b_{it} - b_{jt})^2) \\ & \exp(-\frac{1}{2}\tau_T \sum_{i=1}^{N} \sum_{t \overset{t}{\sim} r} (b_{it} - b_{ir})^2)\end{aligned}$$

Model specification

- $y_{it}|a,b \sim N(a_i + z_t b_{it}, \tau_{Data})$
- GMRF for a: $\pi(a) \propto \exp(-\frac{1}{2}\tau_A \sum_{i \overset{s}{\sim} j}(a_i - a_j)^2)$
- Time-space GMRF for b:

$$\pi(b) \propto \exp(-\frac{1}{2}\tau_B \sum_{t=1}^{T} \sum_{i \overset{s}{\sim} j} (b_{it} - b_{jt})^2)$$

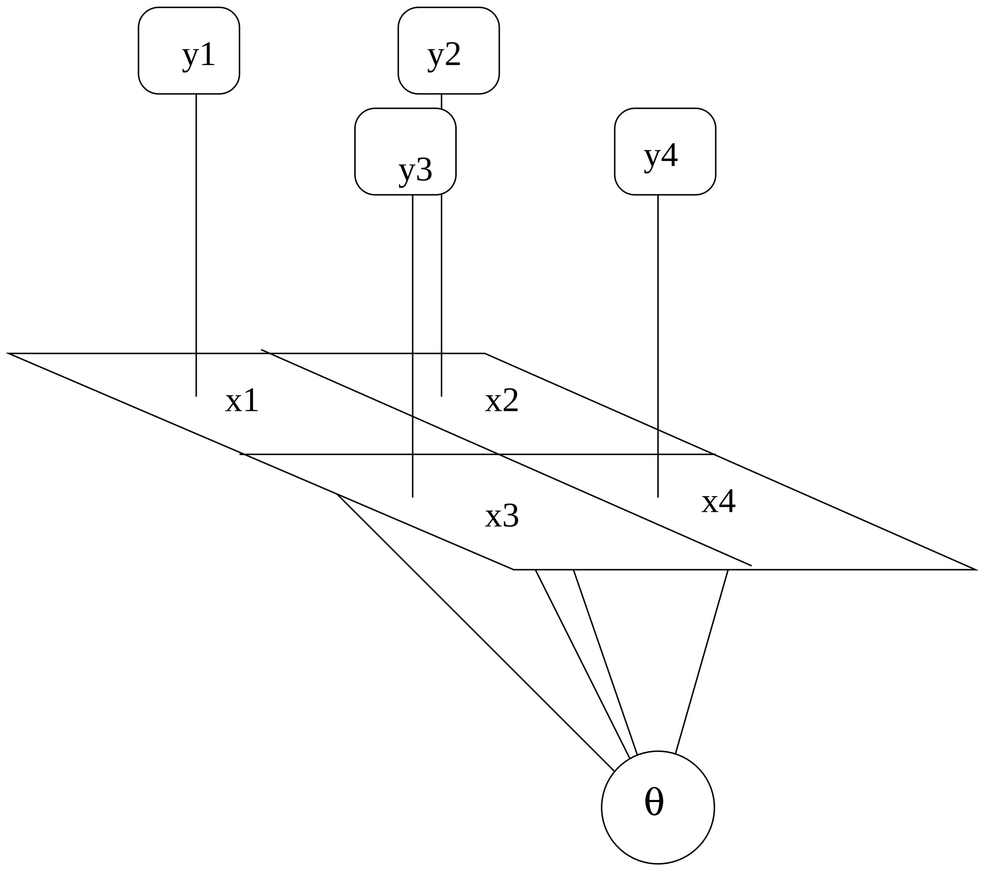
$$\exp(-\frac{1}{2}\tau_T \sum_{i=1}^{N} \sum_{t \overset{t}{\sim} r} (b_{it} - b_{ir})^2)$$

- Posterior, $x = (a, b)$ and $\theta = (\tau_A, \tau_B, \tau_T, \tau_{Data})$:

$$\pi(x, \theta|y) \propto \pi(y|x)\pi(x|\theta)\pi(\theta)$$

Latent GMRF models used

- Mutually independent likelihoods
- $\pi(x|\theta) \sim GMRF$



Gaussian Markov random field

A GMRF $x = (x_1, x_2, \ldots, x_n)$ is:

- Multivariate Gaussian distributed
- with a Markov property;
  - neighbourhood structure
  - x_i and x_j conditionally dependent only if they are neighbours

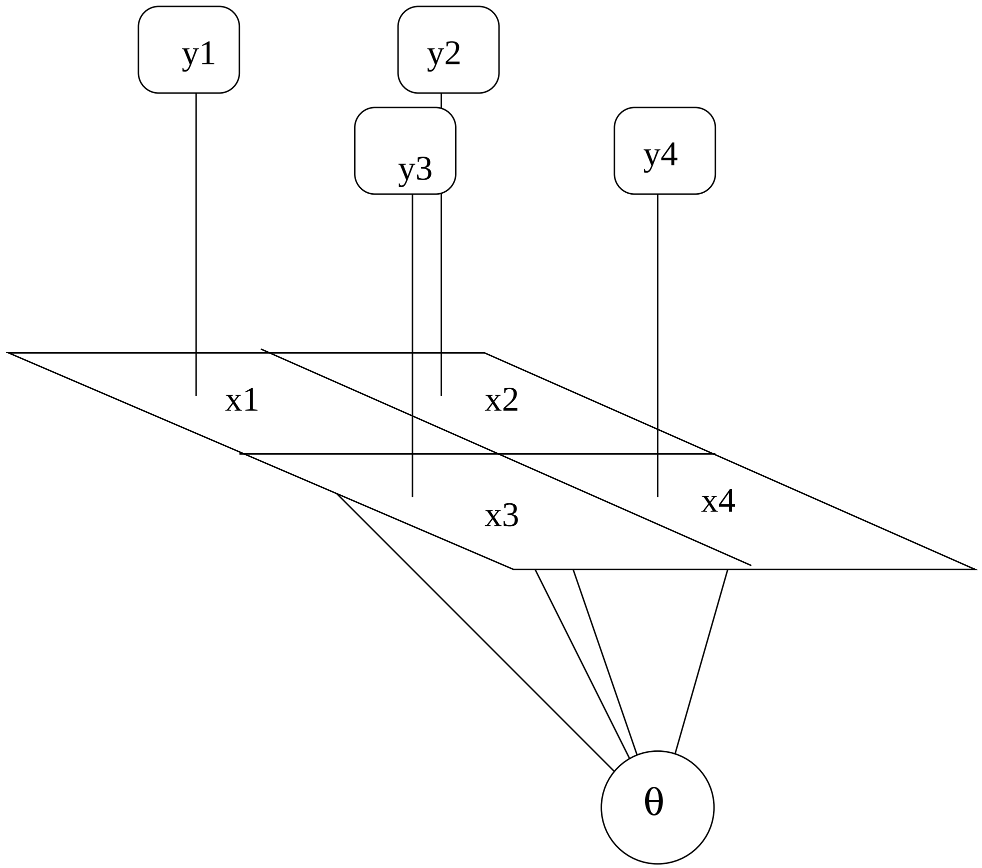
Gaussian Markov random field

A GMRF $x = (x_1, x_2, \ldots, x_n)$ is:

- Multivariate Gaussian distributed
- with a Markov property;
  - neighbourhood structure
  - x_i and x_j conditionally dependent only if they are neighbours
- Gives sparse precision matrix and computational benefi ts.
- All full conditional distribution are GMRFs, and easy to sample from and evaluate.

Metropolis-Hastings with one-block updating scheme

One-block updating scheme: x and θ are updated simultaneously.



Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim \; q(x|x^j, \theta^{new})$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim q(x|x^j, \theta^{new})$
  - Calculate α and accept / reject
  - if(accept)
    - $\theta^{j+1} = \theta^{new}$ and $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$ and $x^{j+1} = x^j$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim \; q(x|x^j, \theta^{new})$
  - Calculate α and accept / reject
  - if(accept)
    - $\theta^{j+1} = \theta^{new}$ and $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$ and $x^{j+1} = x^j$
- Return $(x^1, x^2, \ldots, x^{niter})$ and $(\theta^1, \theta^2, \ldots, \theta^{niter})$

Metropolis-Hastings with one-block updating scheme

- Given x^0 and θ^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample $x^{new} \sim$ $\boxed{q(x|x^j, \theta^{new})}$
  - Calculate α and accept / reject
  - if(accept)
    - $\theta^{j+1} = \theta^{new}$ and $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$ and $x^{j+1} = x^j$
- Return $(x^1, x^2, \ldots, x^{niter})$ and $(\theta^1, \theta^2, \ldots, \theta^{niter})$

Challenge: To make a good and cheap proposal for x.

Our setting

- Gold: Want to construct a $q(x|x^{old}, \theta^{new})$ that:
  - Produces nearly independent samples from approximately $\pi(x|y, \theta^{new})$.
  - Is computationally feasible to sample from.
- Constraint: Too expensive to sample from an n-dim. distribution.
- Here: Construct a proposal for x which we can sample from and evaluate working only with smaller blocks.

Proposal from blocks

- Can use one full scan of a Gibbs sampler as $q(x|x^{old}, \theta^{new})$.
- Block Gibbs sampler used to get better convergence.

Traditional blocking

- x: 100×100 GMRF, $E(x) = 0$ but $x^0 = 3$
- 5×5 neighbourhood
- A GMRF approximation to correlation function:

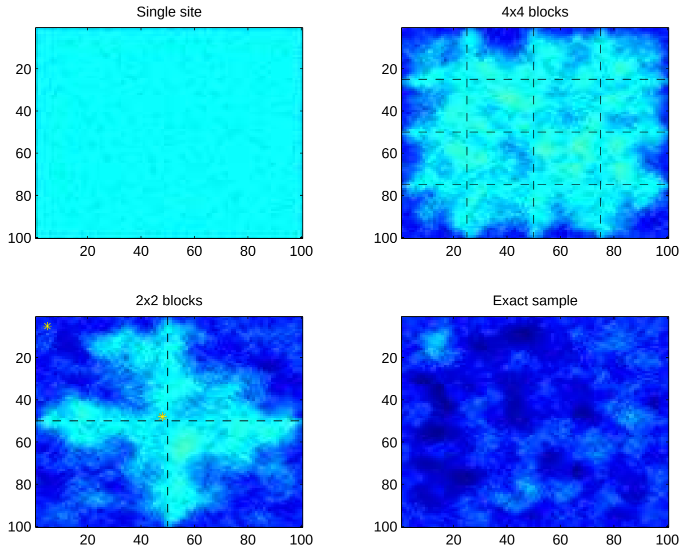
$$\rho(x_i, x_j) = \exp(\frac{-3d(x_i, x_j)}{r})$$

Traditional blocking

- x: 100×100 GMRF, $E(x) = 0$ but $x^0 = 3$
- 5×5 neighbourhood
- A GMRF approximation to correlation function:

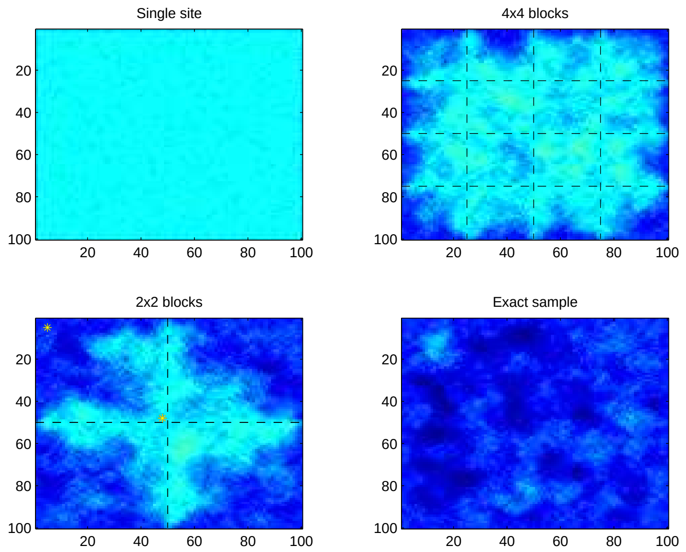
$$\rho(x_i, x_j) = \exp(\frac{-3d(x_i, x_j)}{r})$$

First iteration ($r = 40$)

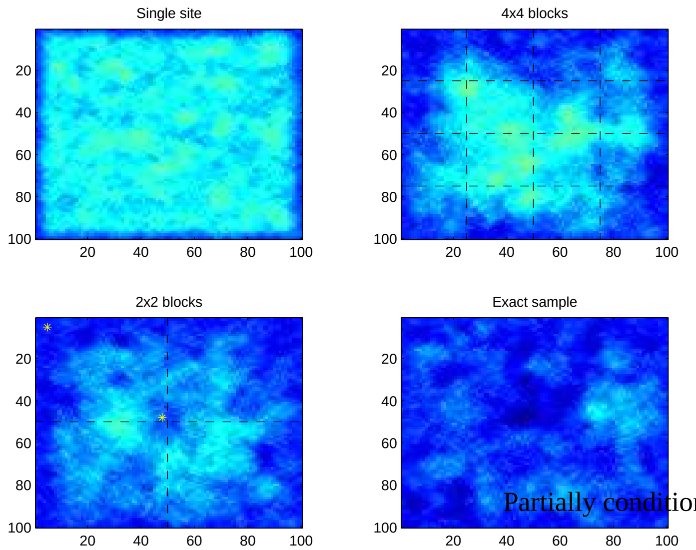


Traditional blocking

First iteration ($r = 40$)

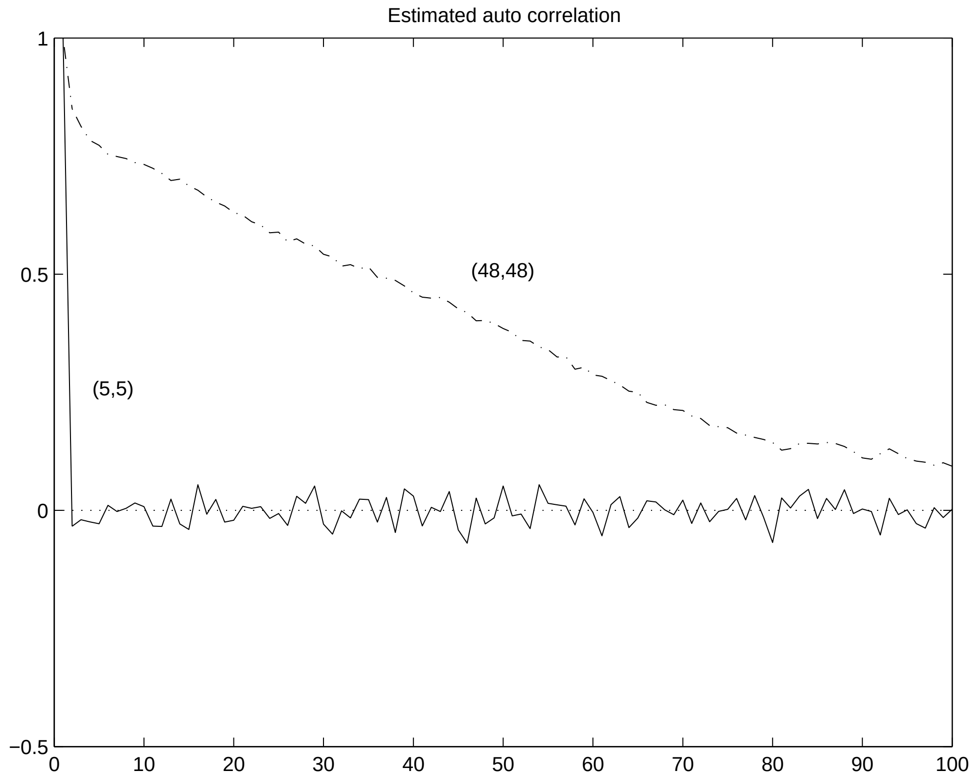


200th iteration ($r = 40$)



Traditional blocking

Estimated autocorrelation:

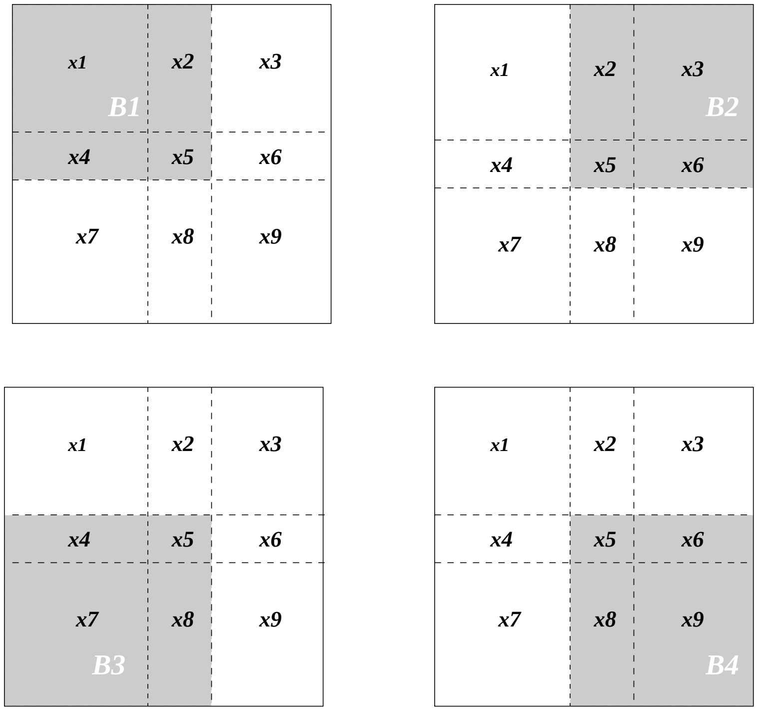


Overlapping block Gibbs sampler

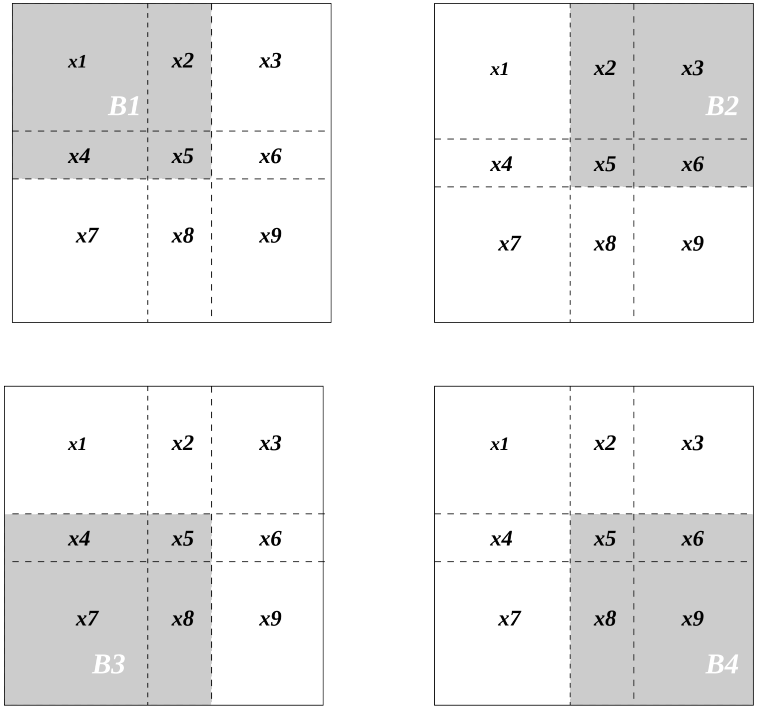
Idea: Let the blocks overlap.

Overlapping block Gibbs sampler

Idea: Let the blocks overlap.

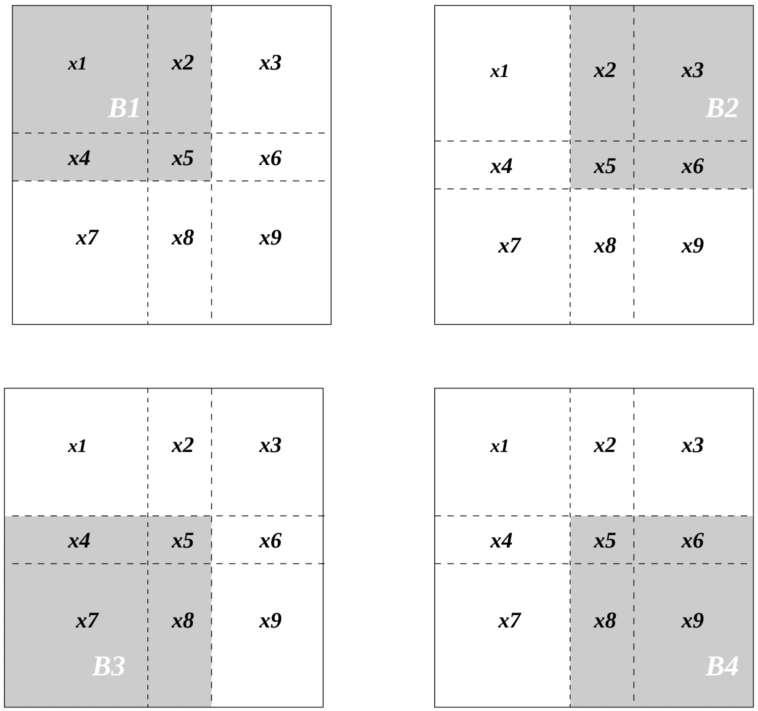


Overlapping block Gibbs sampler



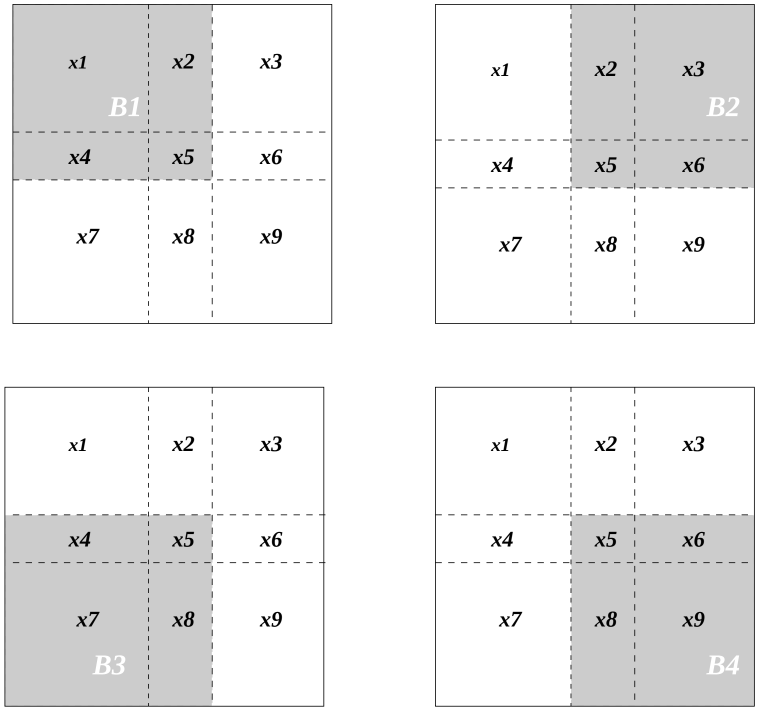
- Given x^i
  - Sample $(x_1^{i+1}, x_2^{B1}, x_4^{B1}, x_5^{B1}) \sim \pi(x_{B_1} | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$

Overlapping block Gibbs sampler



- Given x^i
  - Sample $(x_1^{i+1}, x_2^{B1}, x_4^{B1}, x_5^{B1}) \sim \pi(x_{B_1} | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_2^{i+1}, x_3^{i+1}, x_5^{B2}, x_6^{B2}) \sim \pi(x_{B_2} | x_1^{i+1}, x_4^{B1}, x_7^i, x_8^i, x_9^i)$

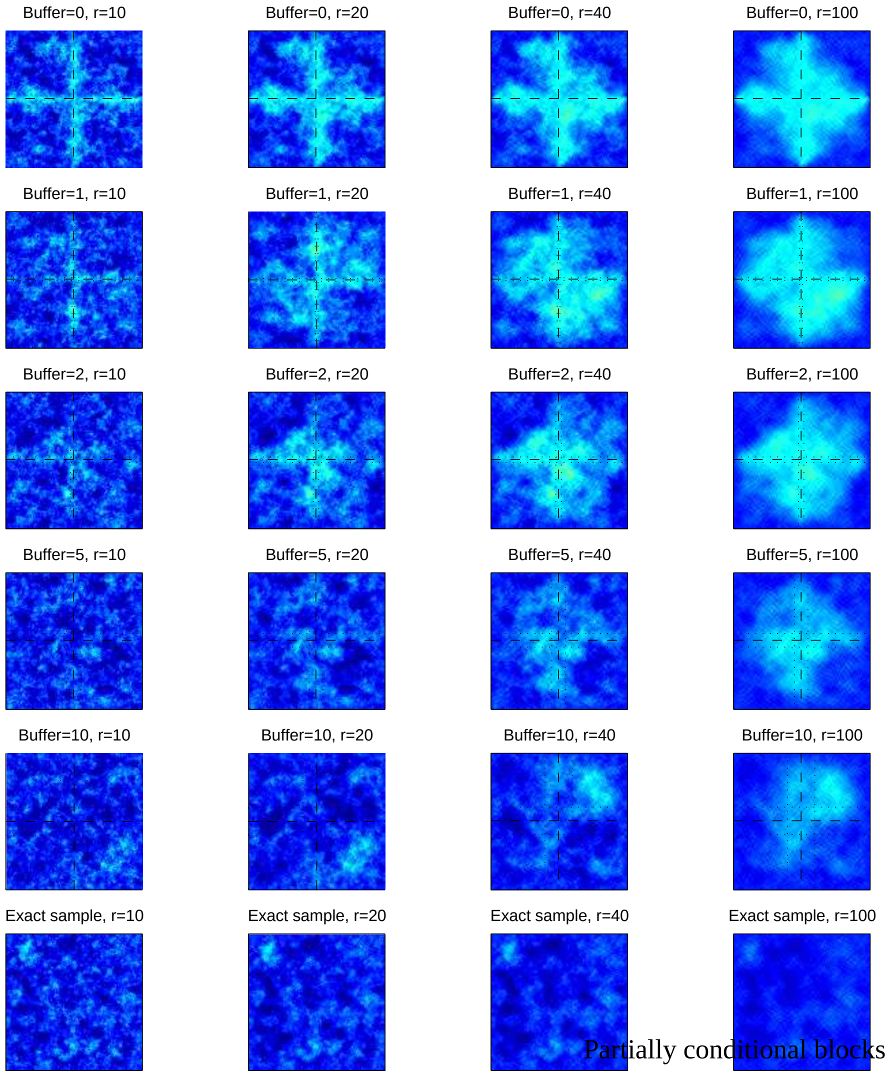
Overlapping block Gibbs sampler



- Given x^i
  - Sample $(x_1^{i+1}, x_2^{B1}, x_4^{B1}, x_5^{B1}) \sim \pi(x_{B_1} | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_2^{i+1}, x_3^{i+1}, x_5^{B2}, x_6^{B2}) \sim \pi(x_{B_2} | x_1^{i+1}, x_4^{B1}, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_4^{i+1}, x_5^{B3}, x_7^{i+1}, x_8^{B3}) \sim \pi(x_{B_3} | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_6^{B2}, x_9^i)$
  - Sample $(x_5^{i+1}, x_6^{i+1}, x_8^{i+1}, x_9^{i+1}) \sim \pi(x_{B_4} | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_4^{i+1}, x_7^{i+1})$
- Return x^{i+1}

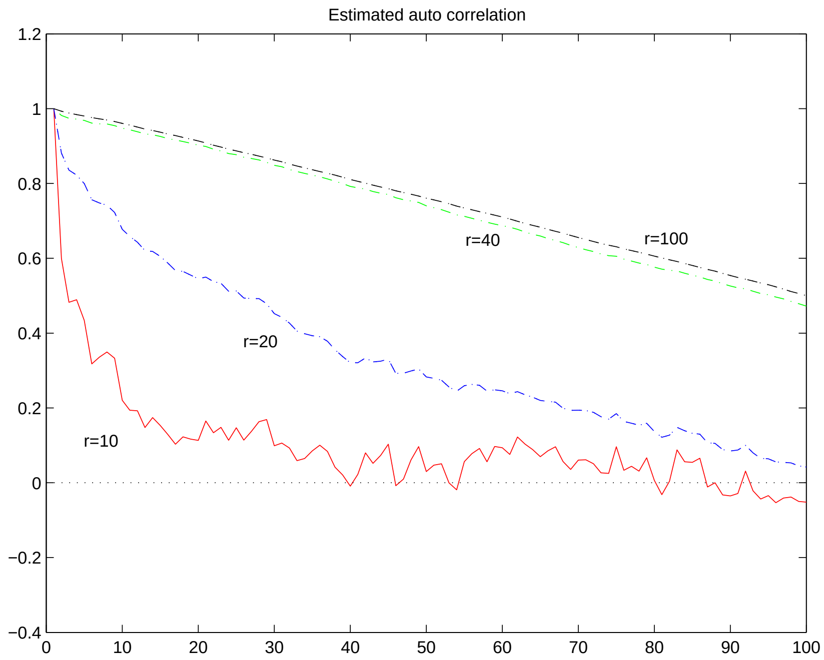
Does it work?

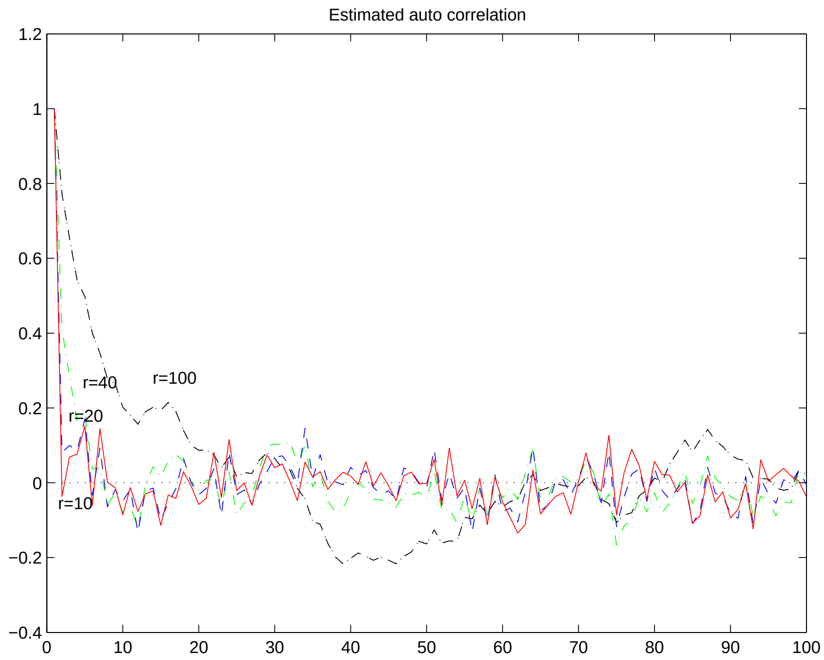
As previous example with $r = \{10, 20, 40, 100\}$ and buffer $= \{0, 1, 2, 5, 10\}$



Does it work?

Estimated auto-correlation function at pixel $(48, 48)$.
Left: Block Gibbs without buffers
Right: With buffer five



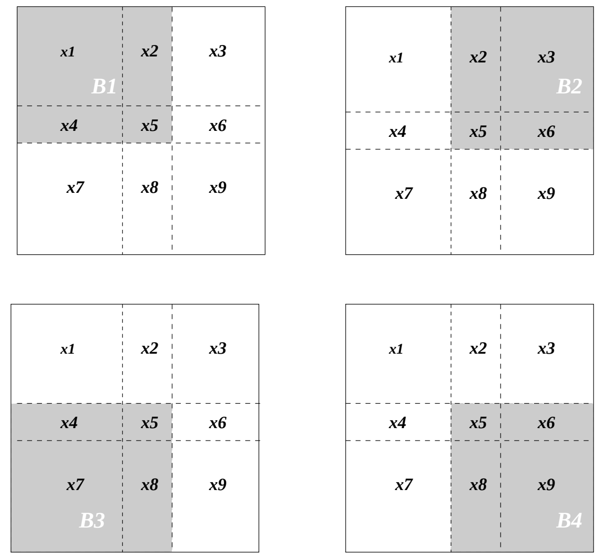


Transition probability

Hard to calculate the transition probability:

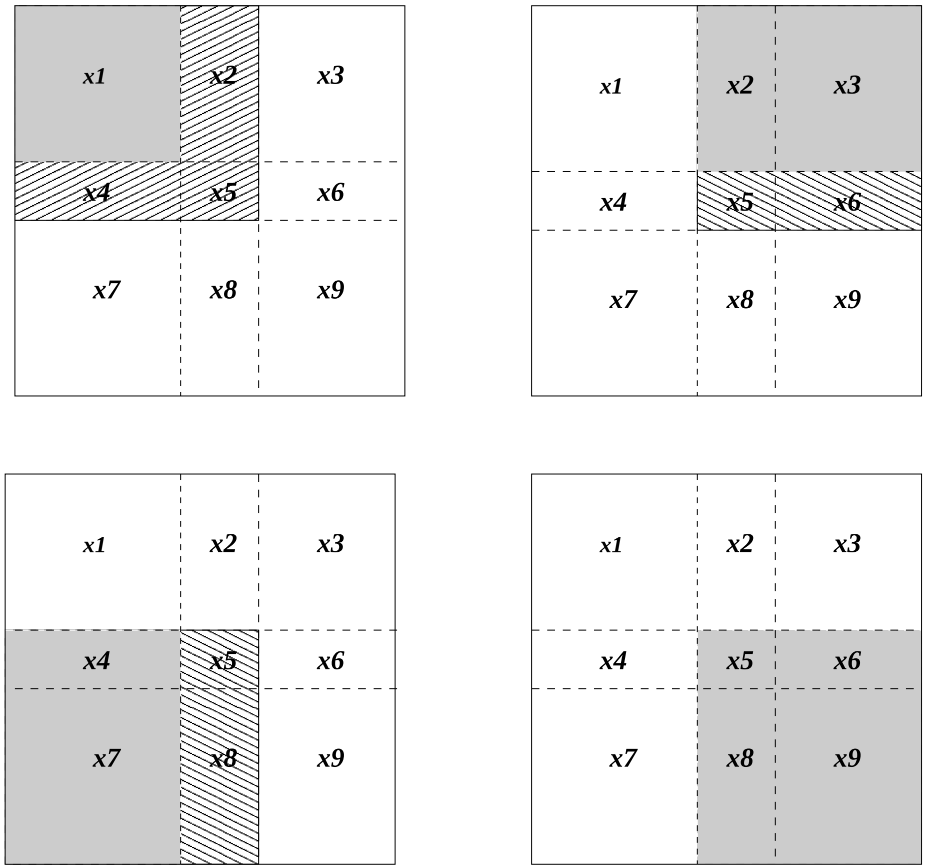
Transition probability

Hard to calculate the transition probability:

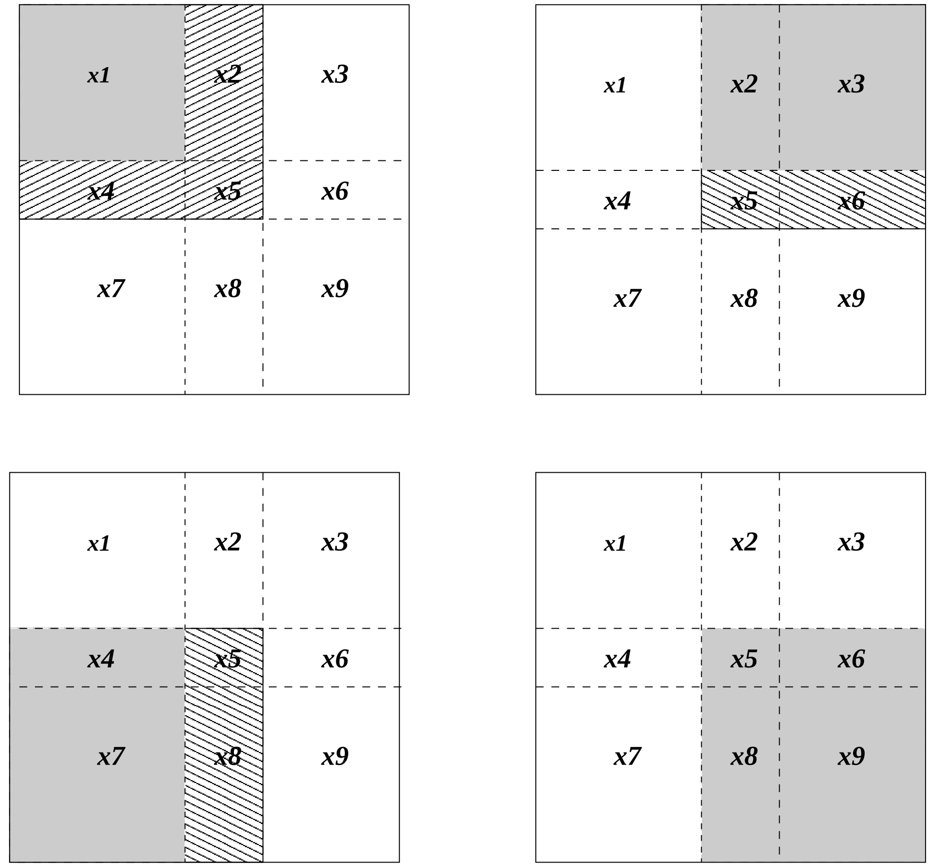


$$
\begin{aligned}
q(x|x') \quad = \quad & \int [\pi_1(x'_1, x_2^{B1}, x_4^{B1}, x_5^{B1}|x_3, x_6, x_7, x_8, x_9) \\
& \pi_2(x'_2, x'_3, x_5^{B2}, x_6^{B2}|x'_1, x_4^{B1}, x_7, x_8, x_9) \\
& \pi_3(x'_4, x_5^{B3}, x'_7, x_8^{B3}|x'_1, x'_2, x'_3, x_6^{B2}, x_9) \\
& \pi_4(x'_5, x'_6, x'_8, x'_9|x'_1, x'_2, x'_3, x'_4, x'_7)] dx_2^{B1} dx_4^{B1} dx_5^{B1} dx_5^{B2} dx_5^{B3} dx_6^{B2} dx_8^{B3}
\end{aligned}
$$

Partially conditional blocks approximation

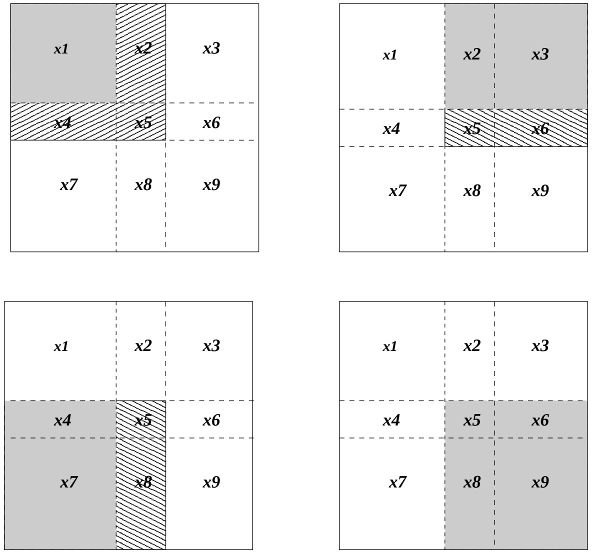


Partially conditional blocks approximation



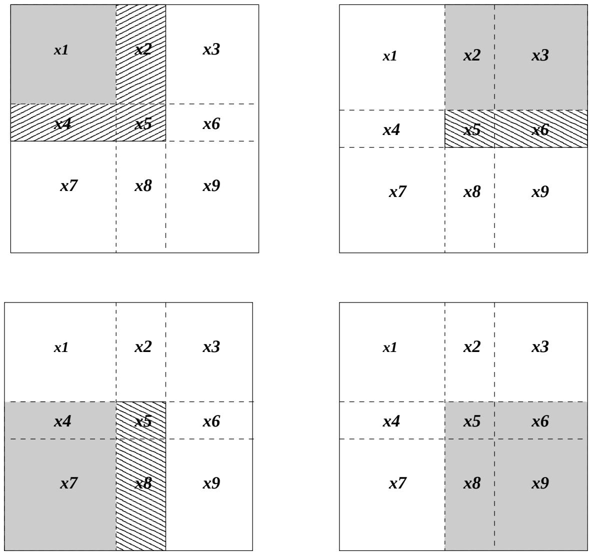
- Given x^0
- for $i = 0 : (niter - 1)$
  - Sample $(x_1^{i+1}) \sim \pi(x_1 | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$

Partially conditional blocks approximation



- Given x^0
- for $i = 0 : (niter - 1)$
  - Sample $(x_1^{i+1}) \sim \pi(x_1 | x_3^i, x_6^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_2^{i+1}, x_3^{i+1}) \sim \pi(x_2, x_3 | x_1^{i+1}, x_4^i, x_7^i, x_8^i, x_9^i)$
  - Sample $(x_4^{i+1}, x_7^{i+1}) \sim \pi(x_4, x_7 | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_6^i, x_9^i)$
  - Sample $(x_5^{i+1}, x_6^{i+1}, x_8^{i+1}, x_9^{i+1}) \sim \pi(x_5, x_6, x_8, x_9 | x_1^{i+1}, x_2^{i+1}, x_3^{i+1}, x_4^{i+1}, x_7^{i+1})$
- Return $((x_1^1, x_2^1, \ldots, x_9^1), (x_1^2, x_2^2, \ldots, x_9^2), \ldots, (x_1^{niter}, x_2^{niter}, \ldots, x_9^{niter}))$

Partially conditional blocks approximation



Transition probability:

$$
\begin{aligned}
q(x|x^{'}) \;=\; & \pi(x_1'|x_3, x_6, x_7, x_8, x_9) \\
& \pi(x_2', x_3'|x_1', x_4, x_7, x_8, x_9) \\
& \pi(x_4', x_7'|x_1', x_2', x_3', x_6, x_9) \\
& \pi(x_5', x_6', x_8', x_9'|x_1', x_2', x_3', x_4', x_7')
\end{aligned}
$$

Partially conditional blocks approximation

Transition probability:

$$
\begin{aligned}
q(x|x') \;=\; & \pi(x'_1|x_3, x_6, x_7, x_8, x_9) \\
& \pi(x'_2, x'_3|x'_1, x_4, x_7, x_8, x_9) \\
& \pi(x'_4, x'_7|x'_1, x'_2, x'_3, x_6, x_9) \\
& \pi(x'_5, x'_6, x'_8, x'_9|x'_1, x'_2, x'_3, x'_4, x'_7)
\end{aligned}
$$

Can use that:

$$\pi(a|c) = \frac{\pi(a, b|c)}{\pi(b|a, c)}$$

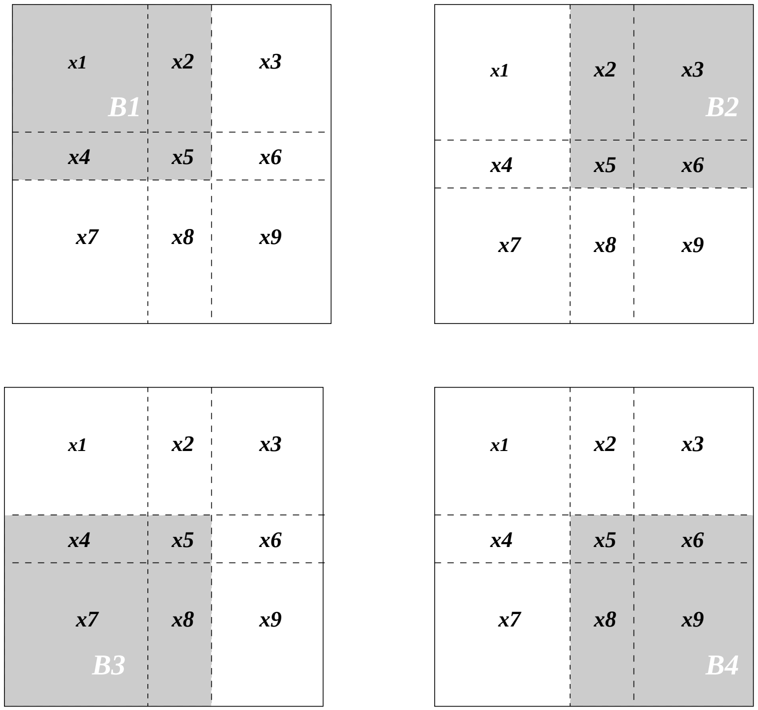
for any b

Opposite reversal

- A M-H proposal constructed by Gibbs steps doesn't give acceptance probability 1.

Opposite reversal

- A M-H proposal constructed by Gibbs steps doesn't give acceptance probability 1.
- Sample first a direction $i = \{0, 1\}$
  - if $i == 0$ use $q_0 : B_1 \rightarrow B_2 \rightarrow B_3 \rightarrow B_4$
  - if $i == 1$ use $q_1 : B_4 \rightarrow B_3 \rightarrow B_2 \rightarrow B_1$



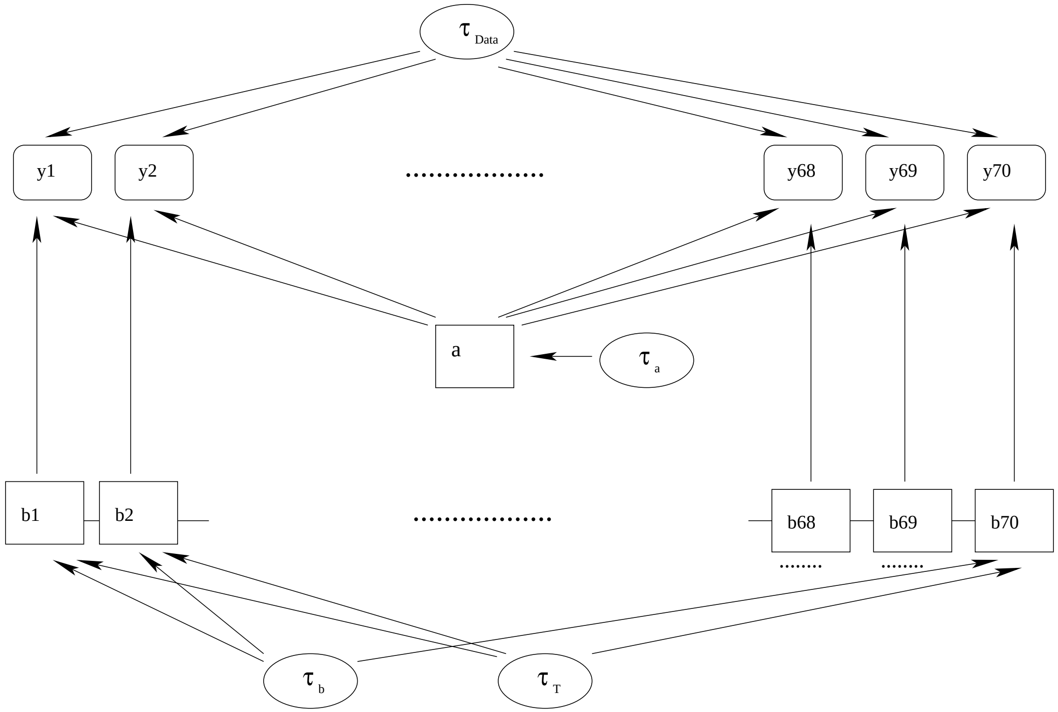
Opposite reversal

- A M-H proposal constructed by Gibbs steps doesn't give acceptance probability 1.
- Sample first a direction $i = \{0, 1\}$
  - if $i == 0$ use $q_0 : B_1 \rightarrow B_2 \rightarrow B_3 \rightarrow B_4$
  - if $i == 1$ use $q_1 : B_4 \rightarrow B_3 \rightarrow B_2 \rightarrow B_1$
- Use acceptance probability (Tjelmeland & Hegstad, 2002)

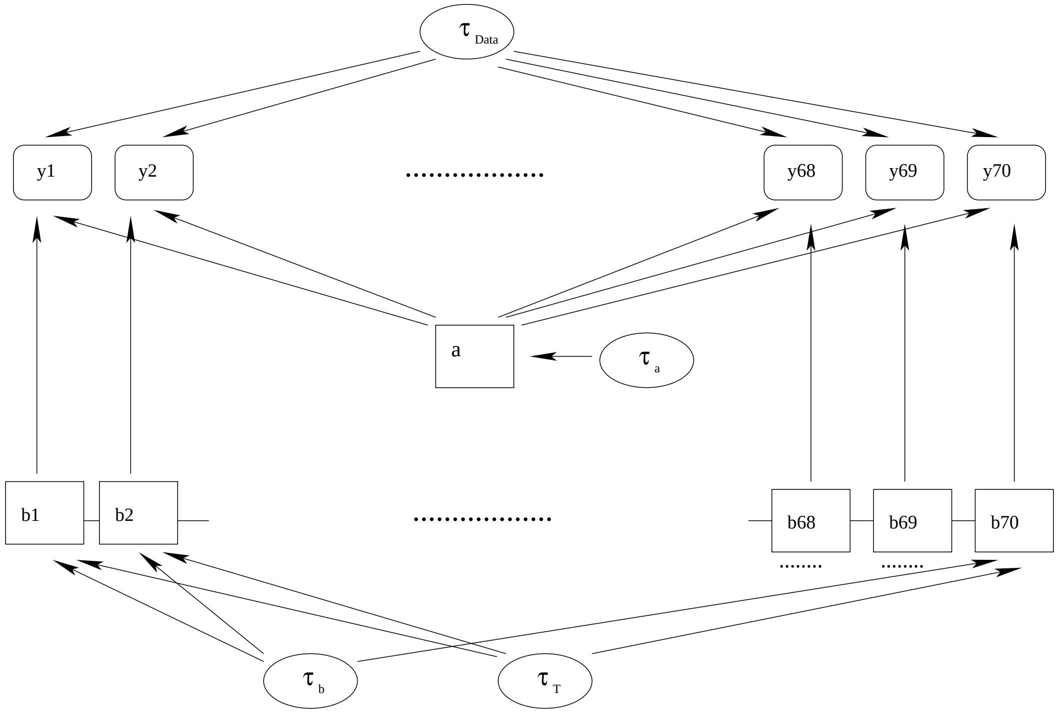
$$\alpha_{i,1-i}(y|x) = \min\left\{1, \frac{\pi(x')q_{1-i}(x|x')}{\pi(x)q_i(x'|x)}\right\}$$

- This gives $\alpha = 1$ for overlapping block Gibbs proposal, but generally not for a partial conditioning sampler

DAG fMRI



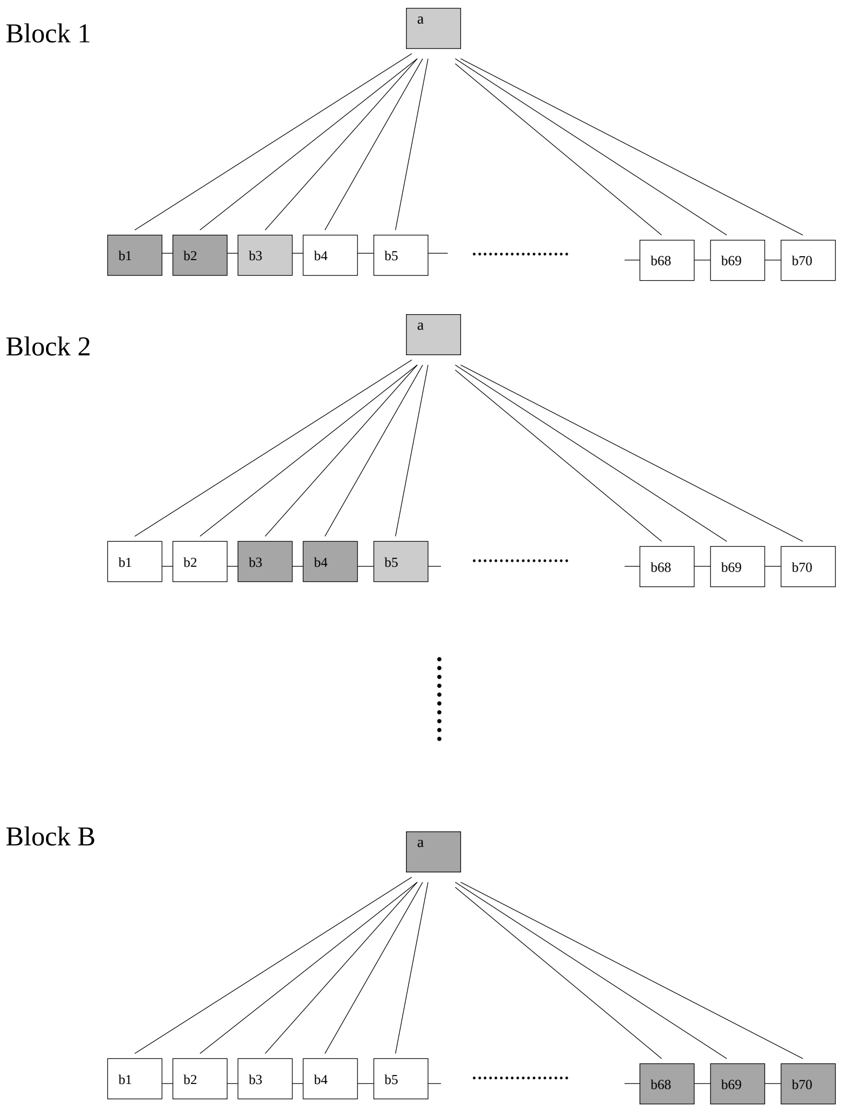
DAG fMRI



- Dimension (a, b) full problem 356 775.
- Dimension (a, b) reduced problem 111 825.

Solution fMRI

Sampling scheme:



Algorithm

- Given θ^0 and x^0
- for $j = 0 : (niter - 1)$ $niter = 20000$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$ Independent random walk, τ_{Data} estimated beforehand

Algorithm

- Given θ^0 and x^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample i: $P(i = 0) = P(i = 1) = 0.5$.
  - Sample from overlapping block Gibbs proposal $x^{new} \sim q_i(x|x^{old}, \theta^{new})$
    - Each block: a and five b_t.
    - Overlap: a and two b_t.

Algorithm

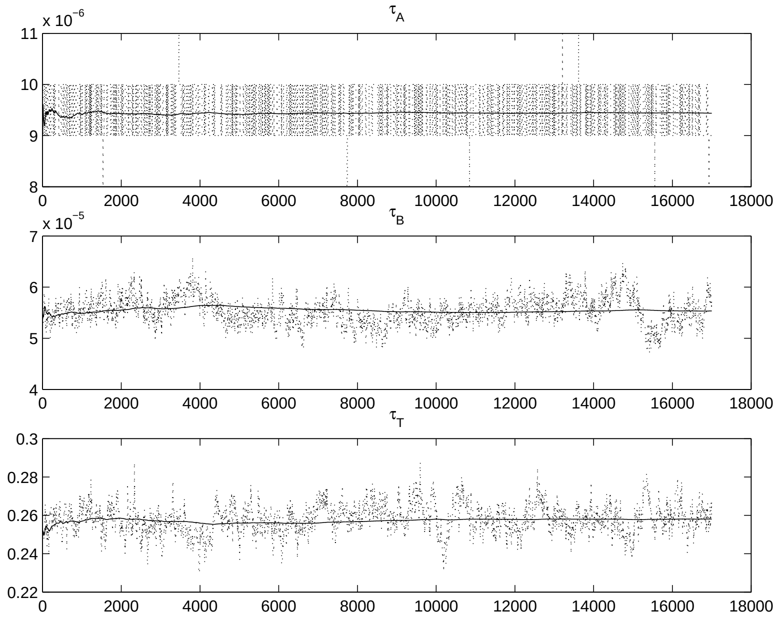
- Given θ^0 and x^0
- for $j = 0 : (niter - 1)$
  - Sample $\theta^{new} \sim q(\theta|\theta^j)$
  - Sample i: $P(i = 0) = P(i = 1) = 0.5$.
  - Sample from overlapping block Gibbs proposal $x^{new} \sim q_i(x|x^{old}, \theta^{new})$
  - Calculate acceptance probability

$$\alpha = \min(1, \frac{\pi(y|x^{new})\pi(x^{new}|\theta^{new})\pi(\theta^{new})q(\theta^j|\theta^{new})q_i(x^j|x^{new}, \theta^j)}{\pi(y|x^j)\pi(x^j|\theta^j)\pi(\theta^j)q(\theta^{new}|\theta^j)q_{1-i}(x^{new}|x^j, \theta^{new})})$$

  - Sample $u \sim \text{Unif}(0, 1)$
  - if($u < \alpha$)
    - $\theta^{j+1} = \theta^{new}$
    - $x^{j+1} = x^{new}$
  - else
    - $\theta^{j+1} = \theta^j$
    - $x^{j+1} = x^j$
- Return $((\theta^1, x^1), (\theta^2, x^2), \dots, (\theta^n, x^n))$.

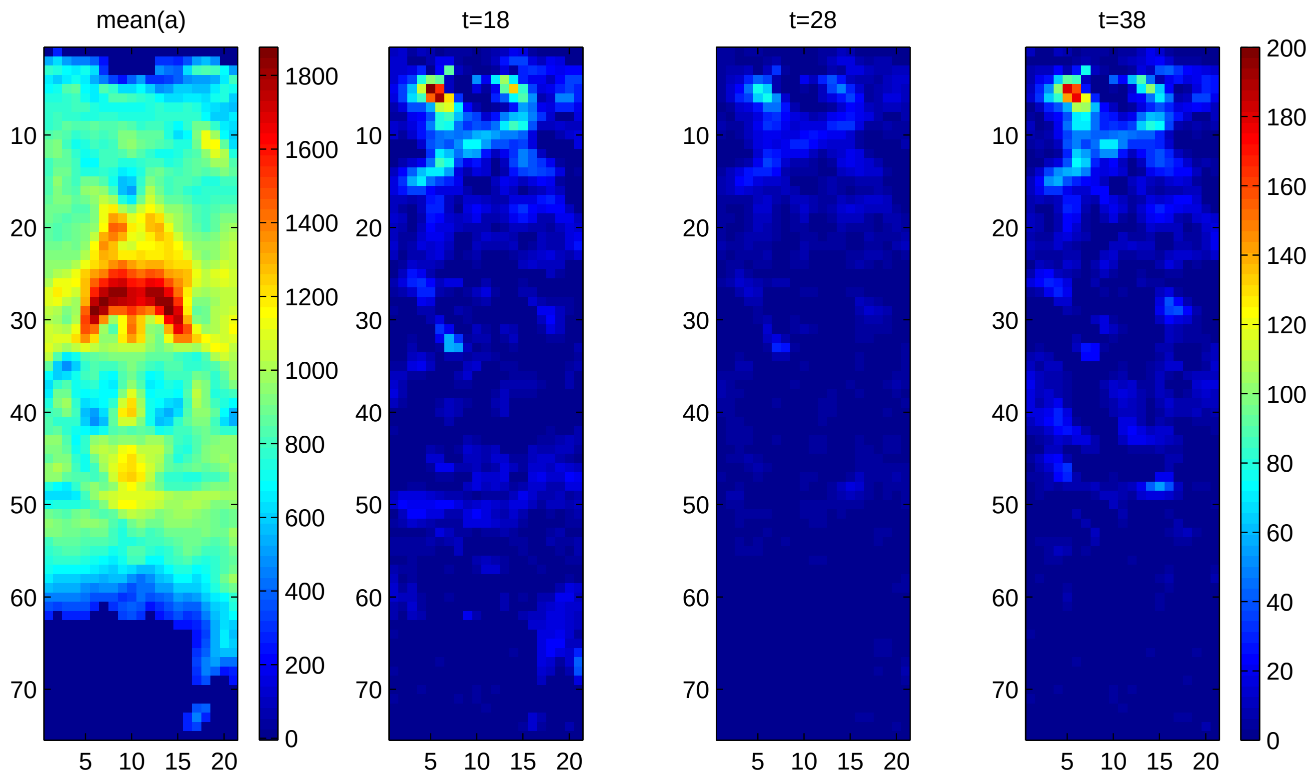
Results fMRI

Trace plots hyper-parameters:



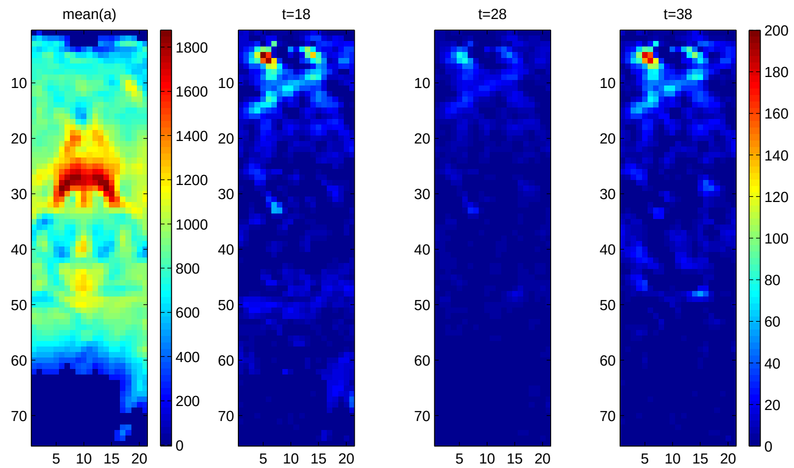
Results fMRI

Estimated mean a and b_{18}, b_{28} and b_{38}:

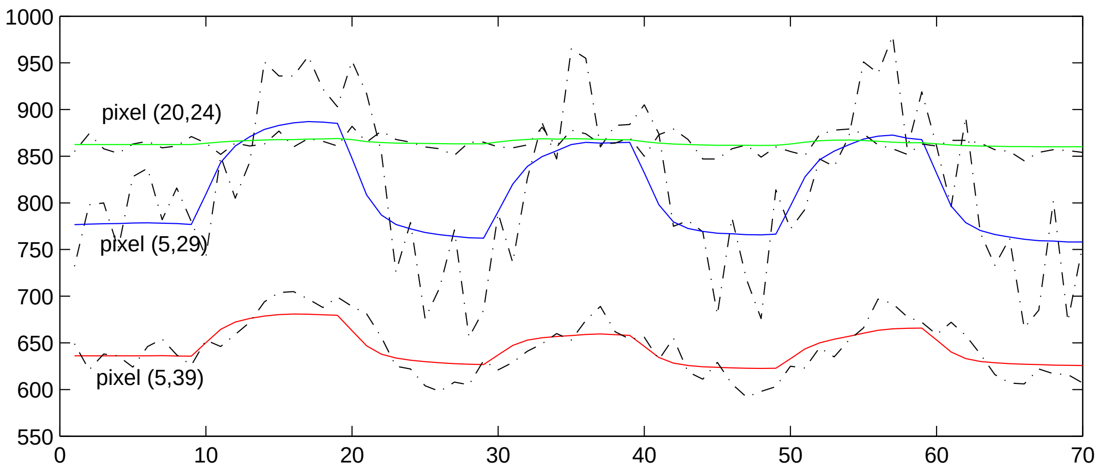


Results fMRI

Estimated mean a and b_{18}, b_{28} and b_{38}:



Estimate for some pixels in time:



Partially conditional approximated blocks approximations

- Can sample each block from an approximation to $\pi(x_B|x_{-B}, \theta^{new}, y)$.
- Enable us to make inference from hidden GMRF models with non-Gaussian likelihood.
- Have used this for a time-space disease-mapping example with GMRF latent field and Poisson likelihood.

How to choose block and buffer sizes

- Blocks: What is OK from a computational point of view.
- Buffers: Depends on the problem:
  - Larger spatial dependents $\Rightarrow$ larger buffers
  - I.e. often depends on the dataset and the current value of θ^{new}.

Summary

Background:

- Latent spatial Markov models describe a large class of problems.
- One-block updating schemes important for mixing of Metropolis-Hastings samplers.

Challenge: Proposal for x, $q(x|x^{old}, \theta^{new})$

- Make an approximation from partially conditional blocks.
- Use knowledge from the dependence structure to set up blocks and buffers.
- Computational benefi ts because only smaller blocks are involved.

This presentation...

can be found on

`www.math.ntnu.no/~ingelins`