Hidden Markov models

Jo Eidsvik

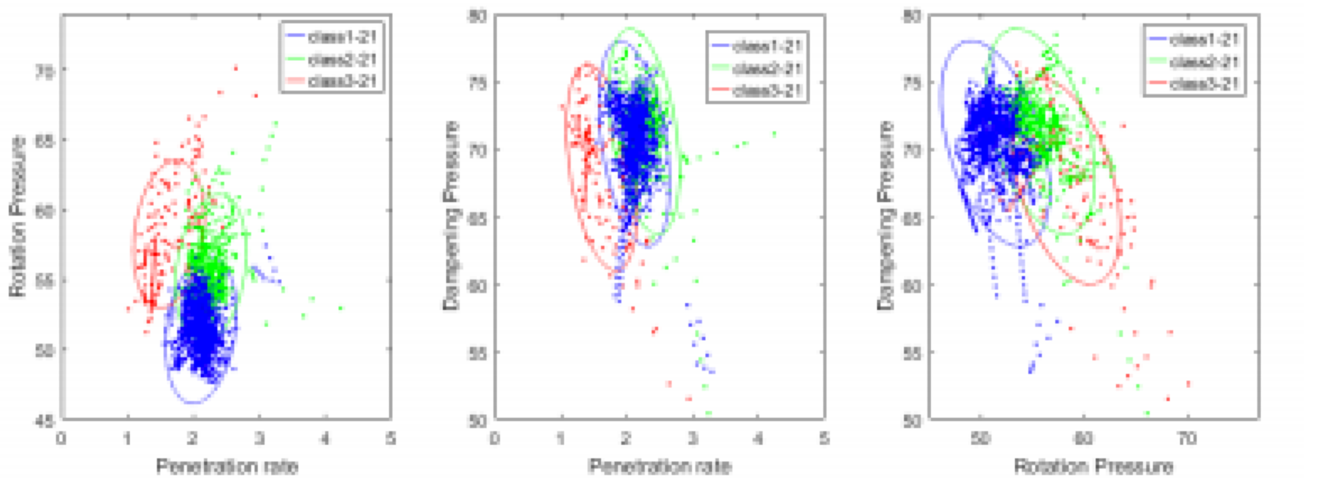
Contents

- Markov models (discrete time and sample space)
- Forward-Backward algorithm
- Example : classification of borehole data
- Related topics: Higher order, Markov random fields, graphs.

Applications

- Rock type classification

Rig hardness data – what is underlying rock type?

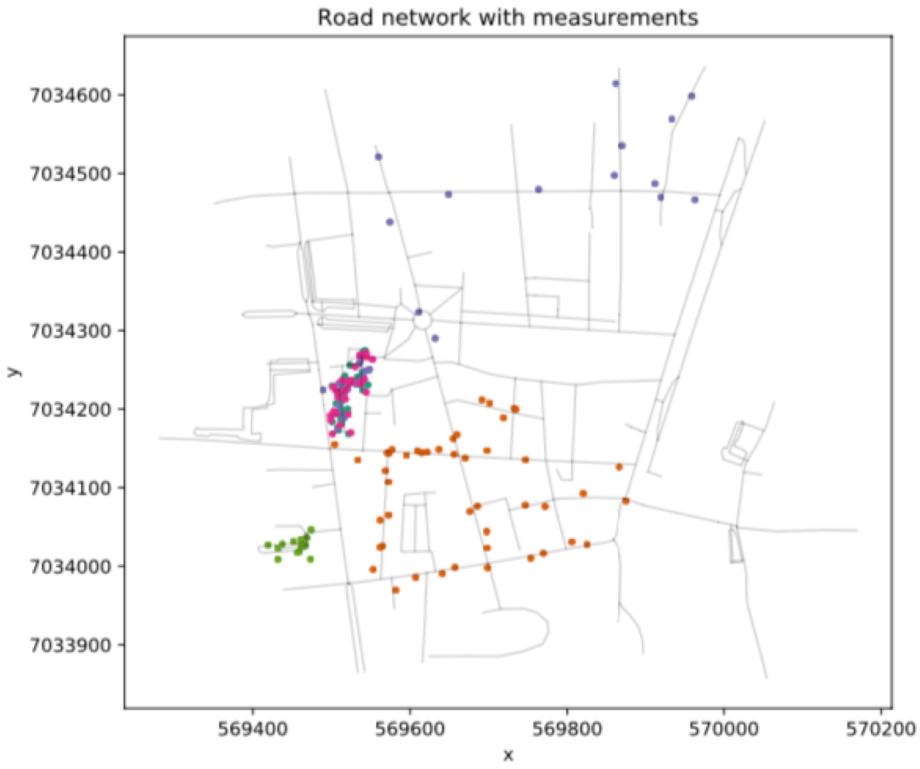


Veena S Vezhapparambu

Applications

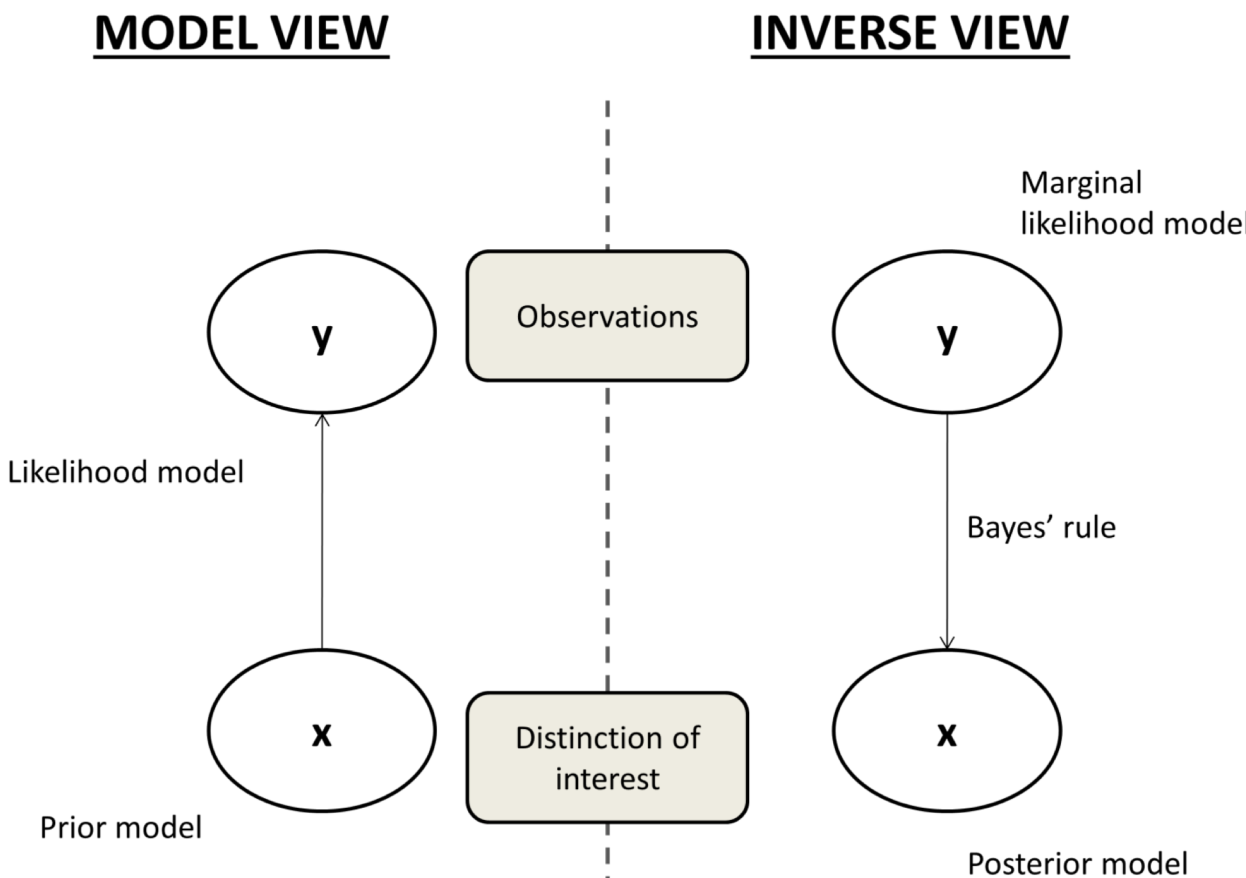
- Map matching

GPS measurements – what is 'true' road segment?



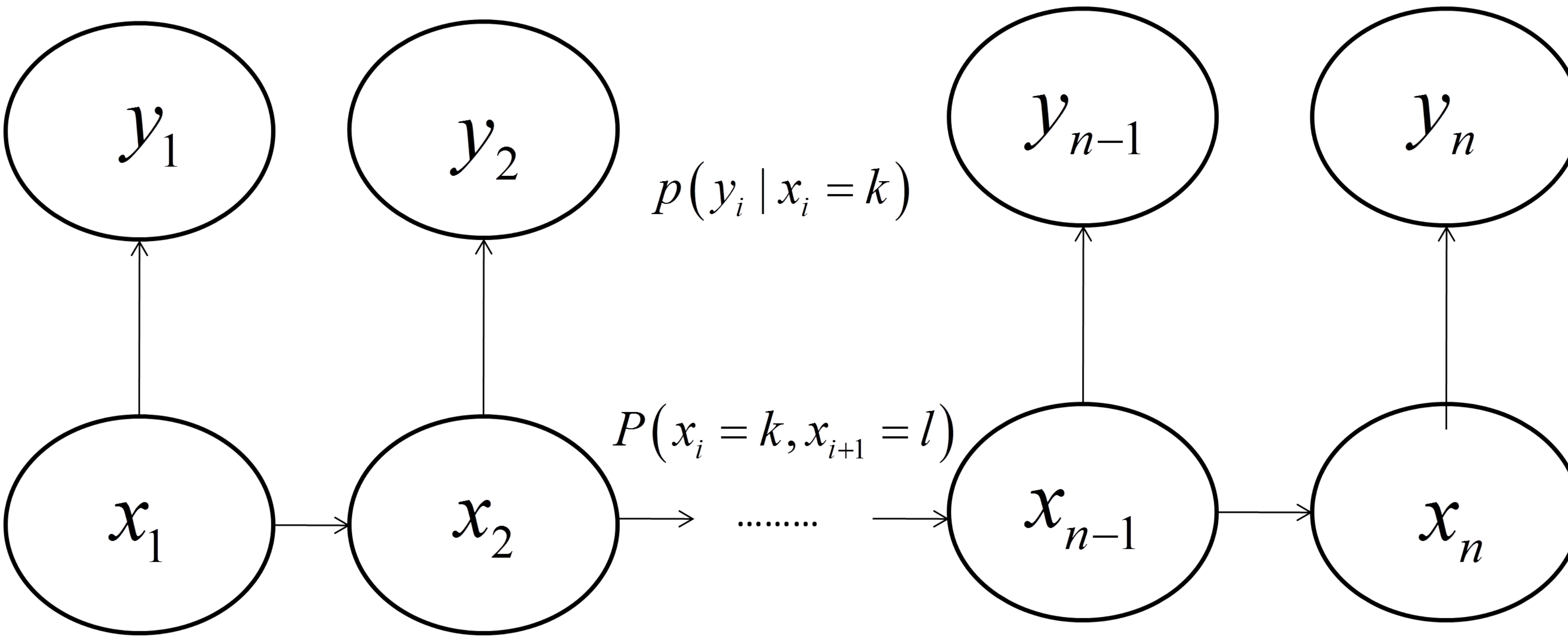
Øyvind Klåpbakken

Bayesian inversion



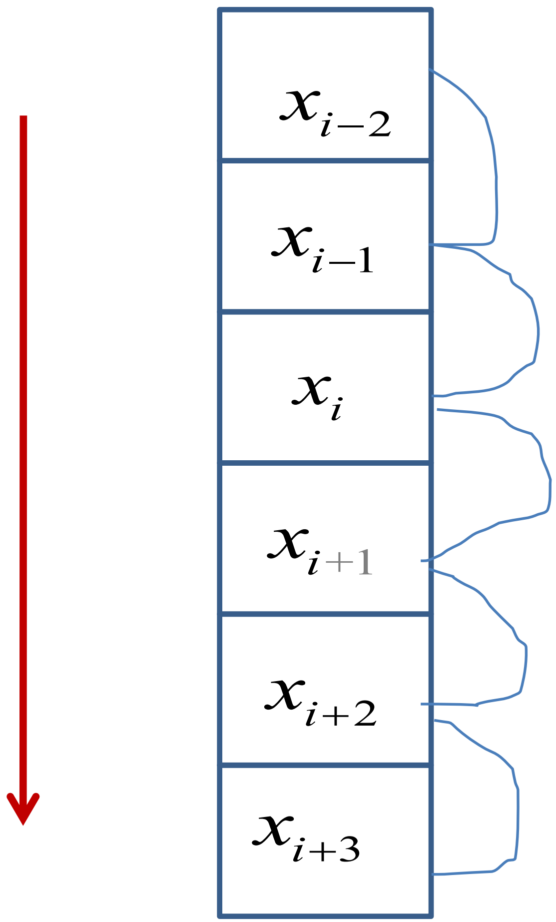
Hidden Markov model (HMM)

Imperfect observations.
Conditionally independent



Latent distinction of interest. Markov chain

Markov chain



$$p(x_i \mid x_{i-1},\ldots,x_1) = p(x_i \mid x_{i-1}),$$

$$p\left(x_1\right).$$

$$x_i \in \left\{1,\ldots,d\right\}, i = 1,\ldots,n$$

$$\sum_{k=1}^{d} p(x_i = k \mid x_{i-1}) = 1$$

All dependence from the past is inherited through the previous state.

Markov transition matrix

- *d=2*

	1	**2**
1	*p*	*1-p*
2	*1-q*	*q*

$$p(x_i = 1 \mid x_{i-1} = 1) = p,$$
$$p(x_i = 2 \mid x_{i-1} = 2) = q$$

Markov chains

- Stationary, p=q, p=0.9 (left), p=0.5 (middle), p=0.1 (right).

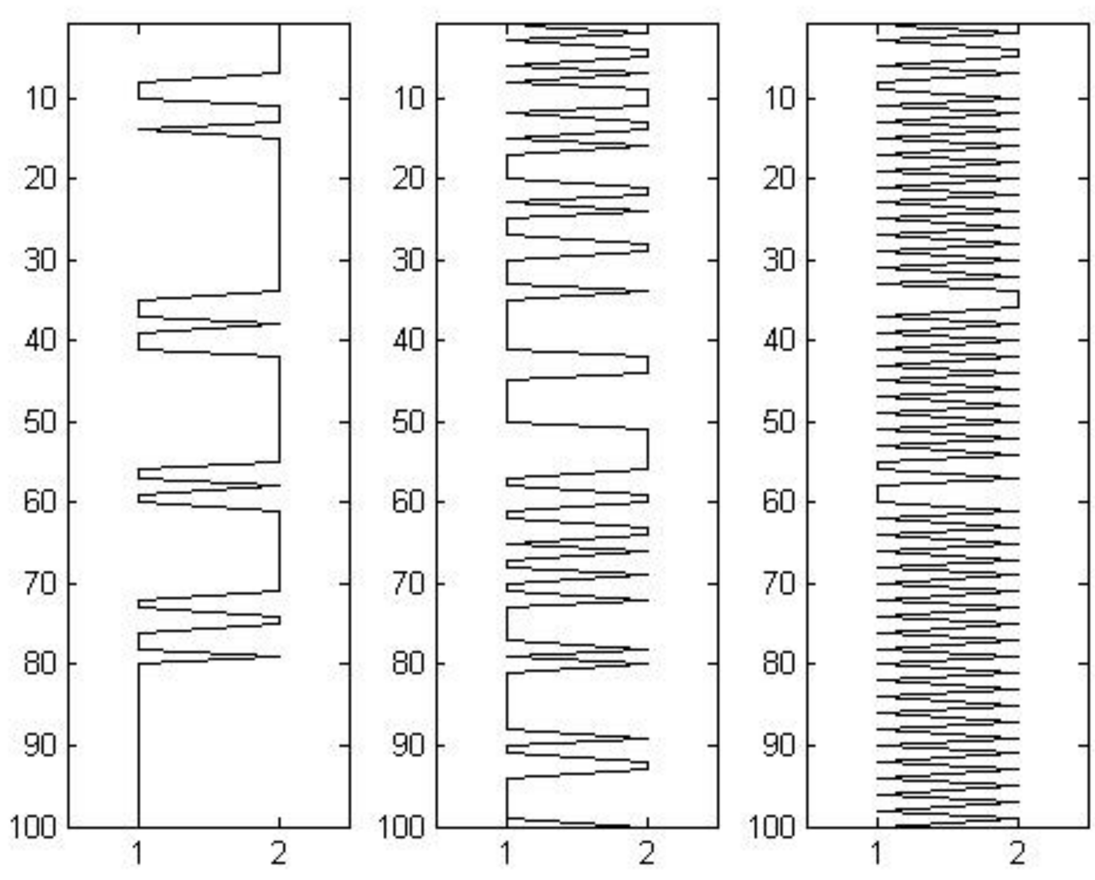
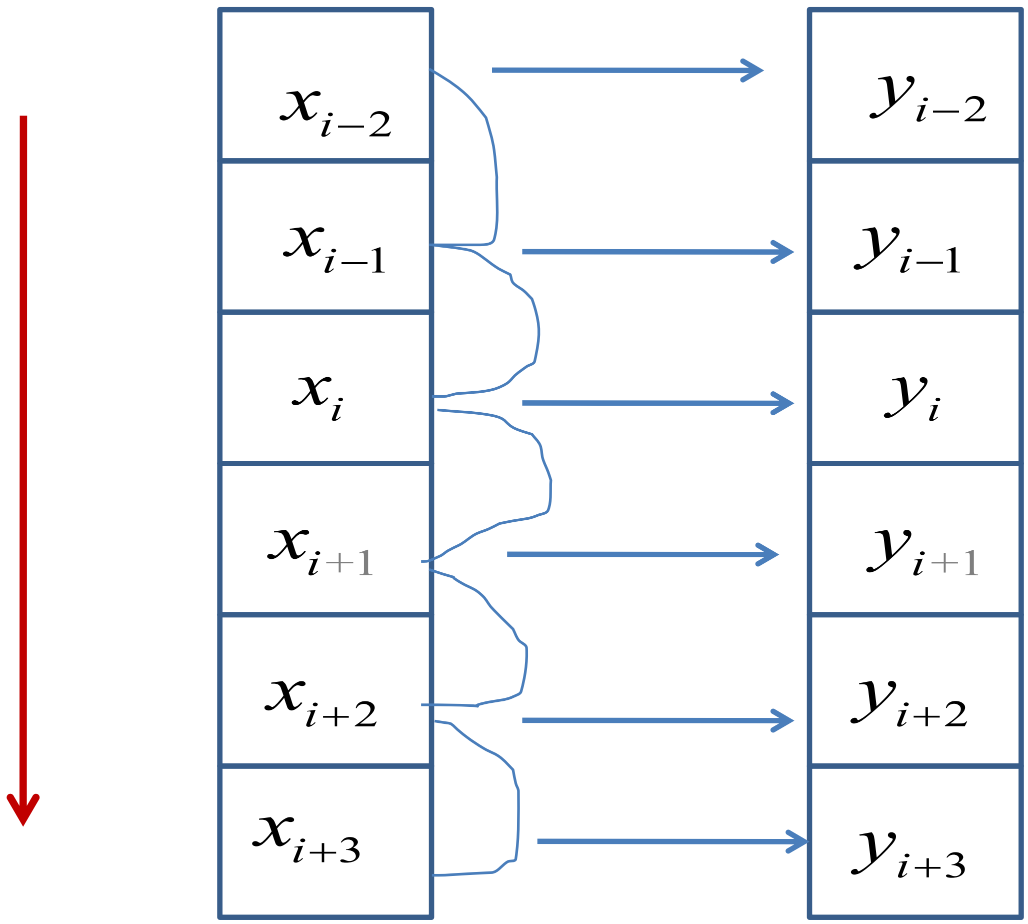


Illustration of Hidden Markov model



«Local data», conditional independence

Hidden Markov models

$$p(x_i \mid x_{i-1},\ldots,x_1) = p(x_i \mid x_{i-1}),$$

$$p\left(x_1\right)$$

$$x_i \in \left\{1,\ldots,d\right\}, i = 1,\ldots,n,$$

$$p(y_i \mid x_i)$$

Data can be discrete or continuous.
Distinction of interest is discrete and Markovian.

HMM: Inference and prediction

Main goal is likelihood evaluation, marginal posterior probabilities or posterior realizations

$$p(y_1,\ldots,y_n) = p_{\boldsymbol{\theta}}(y_1,\ldots,y_n)$$

$$p(x_i = k \mid y_1,\ldots,y_n), k = \{1,..,d\}, i = \{1,..,n\},$$

$$\boldsymbol{x} = \left(x_1,\ldots,x_n\right) \sim p(\boldsymbol{x} \mid y_1,\ldots,y_n)$$

Forward-Backward algorithm

Forward recursion:

$$p(x_i \mid y_1, \ldots, y_{i-1}) = \sum_{k=1}^{d} p(x_i \mid x_{i-1} = k)\, p(x_{i-1} = k \mid y_1, \ldots, y_{i-1})$$

$$p(x_i = k \mid y_1, \ldots, y_i) = \frac{p(y_i \mid x_i = k)\, p(x_i = k \mid y_1, \ldots, y_{i-1})}{p(y_i \mid y_1, \ldots, y_{i-1})}$$

Forward-Backward algorithm

Sequential likelihood:

$$p(y_i \mid y_1,\ldots,y_{i-1}) = \sum_{k=1}^{d} p(y_i \mid x_i = k)\, p(x_i = k \mid y_1,\ldots,y_{i-1})$$

$$p(y_1,\ldots,y_i) = p(y_i \mid y_1,\ldots,y_{i-1})\, p(y_1,\ldots,y_{i-1})$$

$$p(y_1,\ldots,y_n) = p_{\boldsymbol{\theta}}(y_1,\ldots,y_n)$$

Last step of forward run.

Forward-Backward algorithm

Backward recursion:

1. By end of forward:

$$p(x_n = k \mid y_1, \ldots, y_n)$$

2. Stepping back:

$$p(x_i = k \mid y_1, \ldots, y_n), i = n-1, \ldots, 1$$

Forward-Backward algorithm

Backward recursion:

$$p\left(x_i = k \mid y_1,\ldots,y_n\right) = \sum_{l=1}^{d} p\left(x_i = k \mid y_1,\ldots,y_{i+1},x_{i+1} = l\right) p\left(x_{i+1} = l \mid y_1,\ldots,y_n\right)$$

$$p\left(x_i = k \mid y_1,\ldots,y_{i+1},x_{i+1}\right) = \frac{p\left(x_i = k, x_{i+1}, y_{i+1} \mid y_1,\ldots,y_i\right)}{p\left(y_{i+1} \mid x_{i+1}, y_1,\ldots,y_i\right) p\left(x_{i+1} \mid y_1,\ldots,y_i\right)}$$

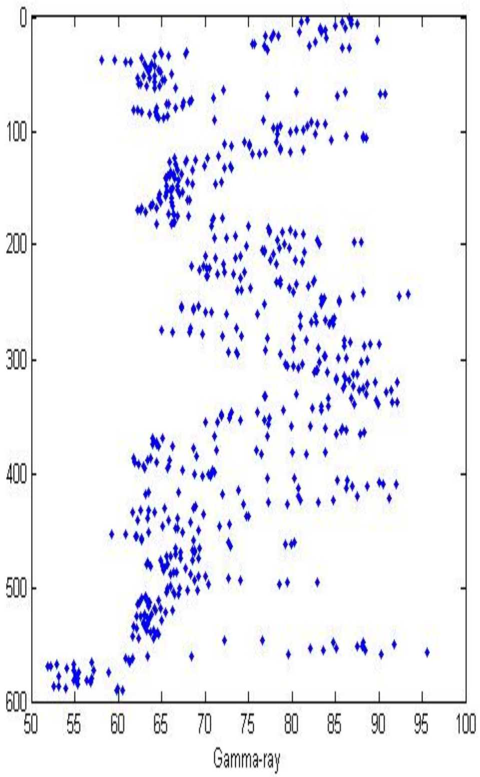
$$= \frac{p\left(x_{i+1} \mid x_i = k\right) p\left(x_i = k \mid y_1,\ldots,y_i\right)}{p\left(x_{i+1} \mid y_1,\ldots,y_i\right)}.$$

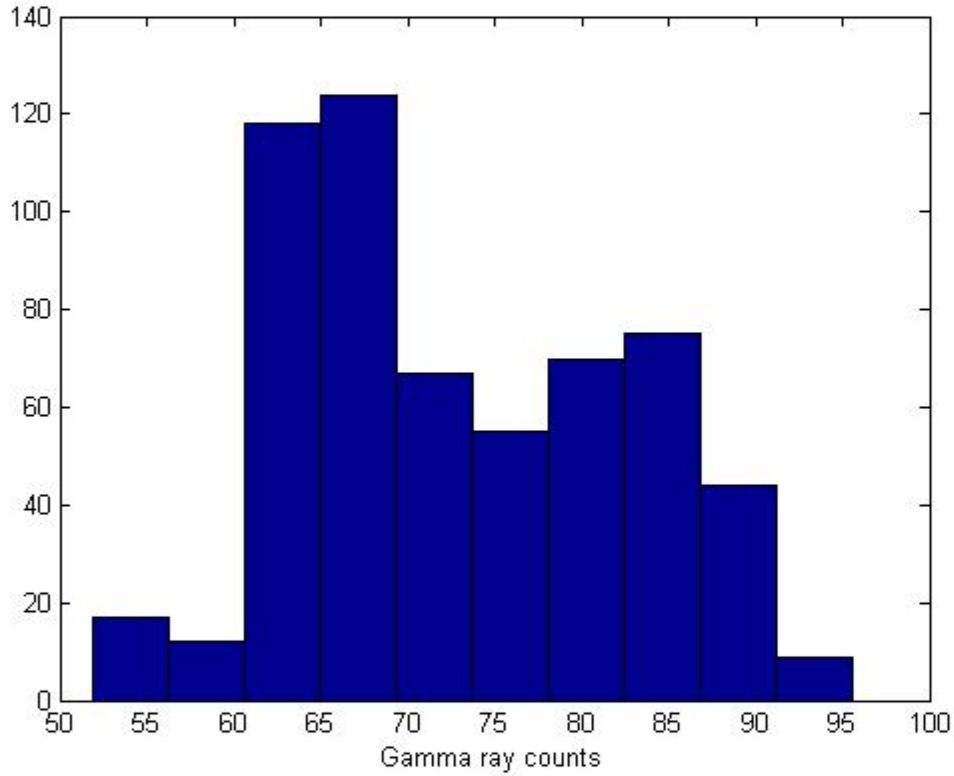
Forward-Backward algorithm

- Useful in applied statistics:
  - Likelihood modeled from measurement accuracy, using discipline-specific knowledge.
  - Prior modeled from a priori understanding.
- Fast
  - Order $O(nd^2)$ calculations.

Example: Hidden Markov models

- Gamma ray information in borehole.





Hidden Markov models

- Latent rock type is sand or shale (at every depth). Goal is to classify rock type, given gamma-ray counts and Markovian dependence in the rock types.

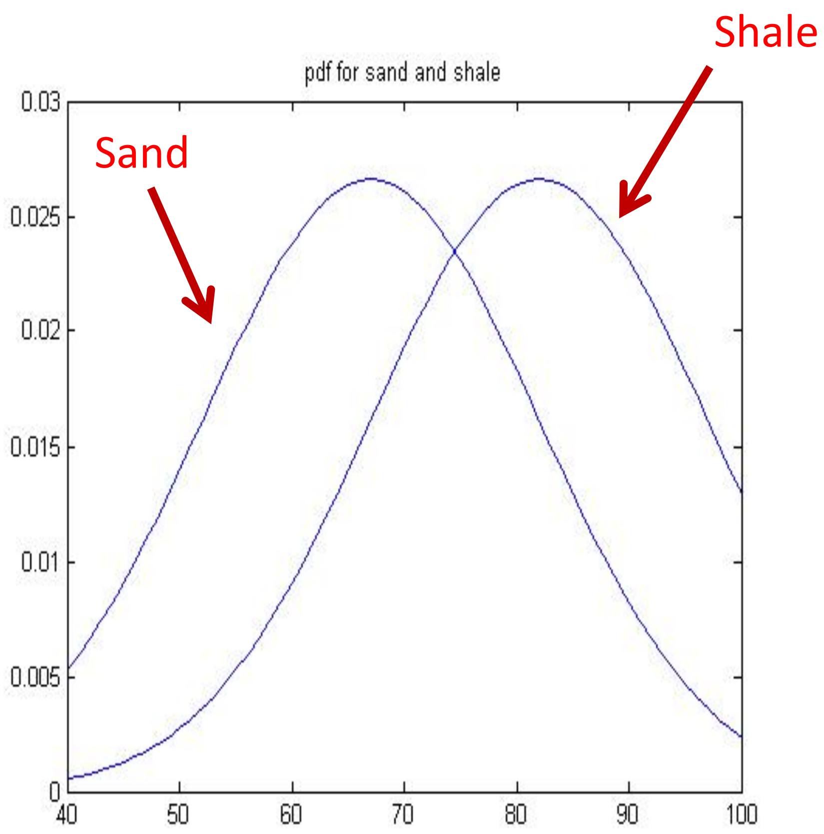
$$p(x_i \mid x_{i-1}, \ldots, x_1) = p(x_i = k \mid x_{i-1} = k) = 0.9,$$

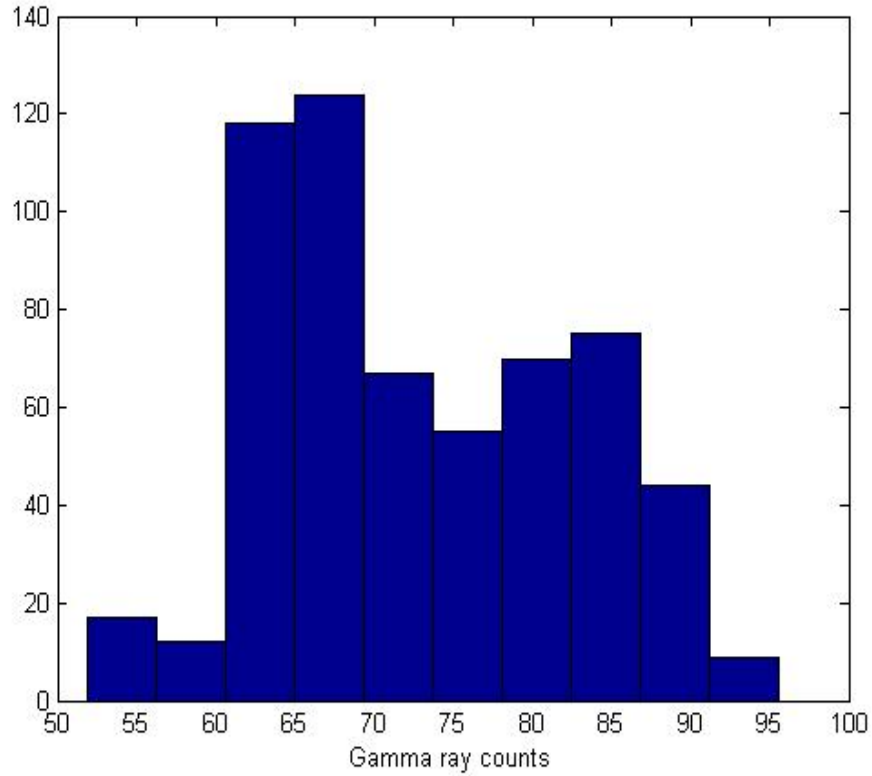
$$x_i \in \{1, 2\},$$

$$p(y_i \mid x_i = k) = N\left(\mu_k, \sigma_k^2\right)$$

Likelihood of gamma-ray data

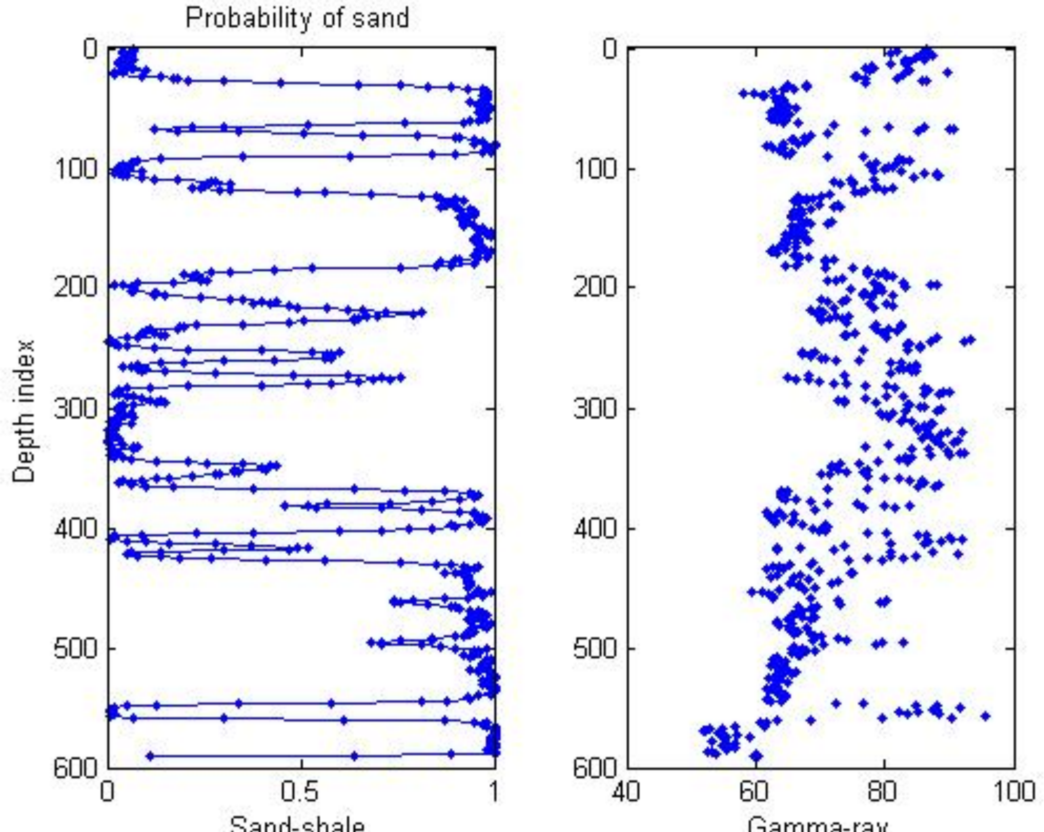
- Sand (smaller counts) and shale (higher counts).





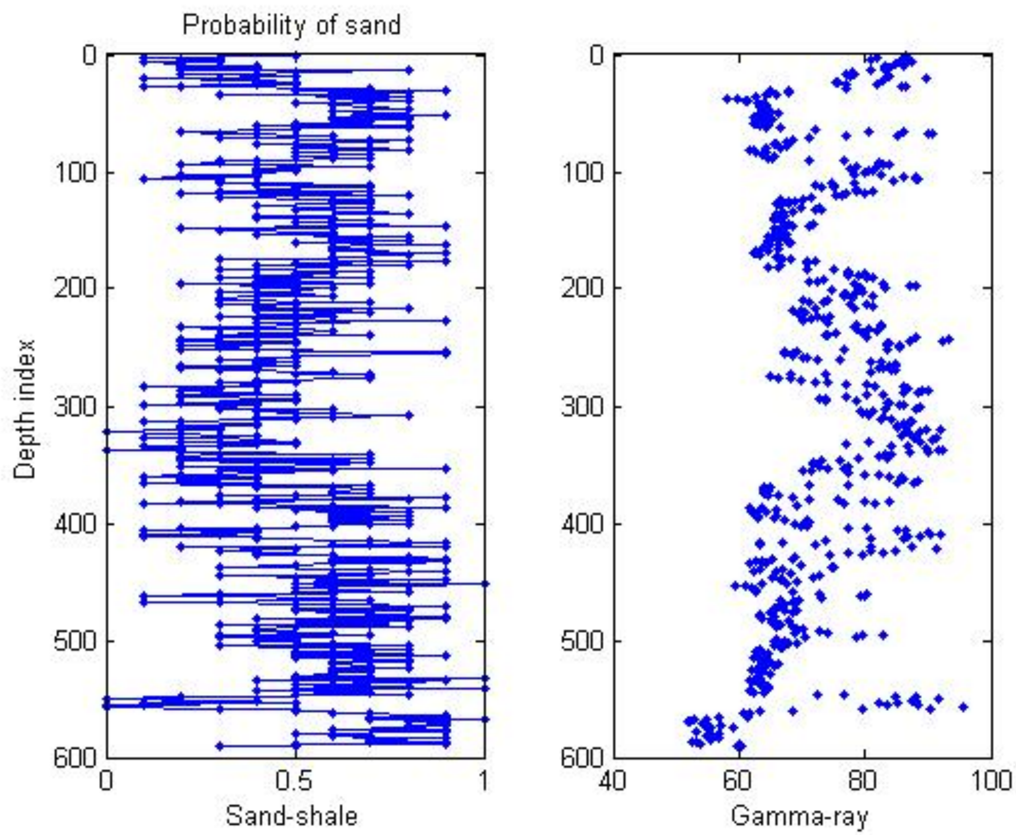
Result 1: Hidden Markov models

- Probability of sand (at every depth) using gamma ray information, and Markovian prior.

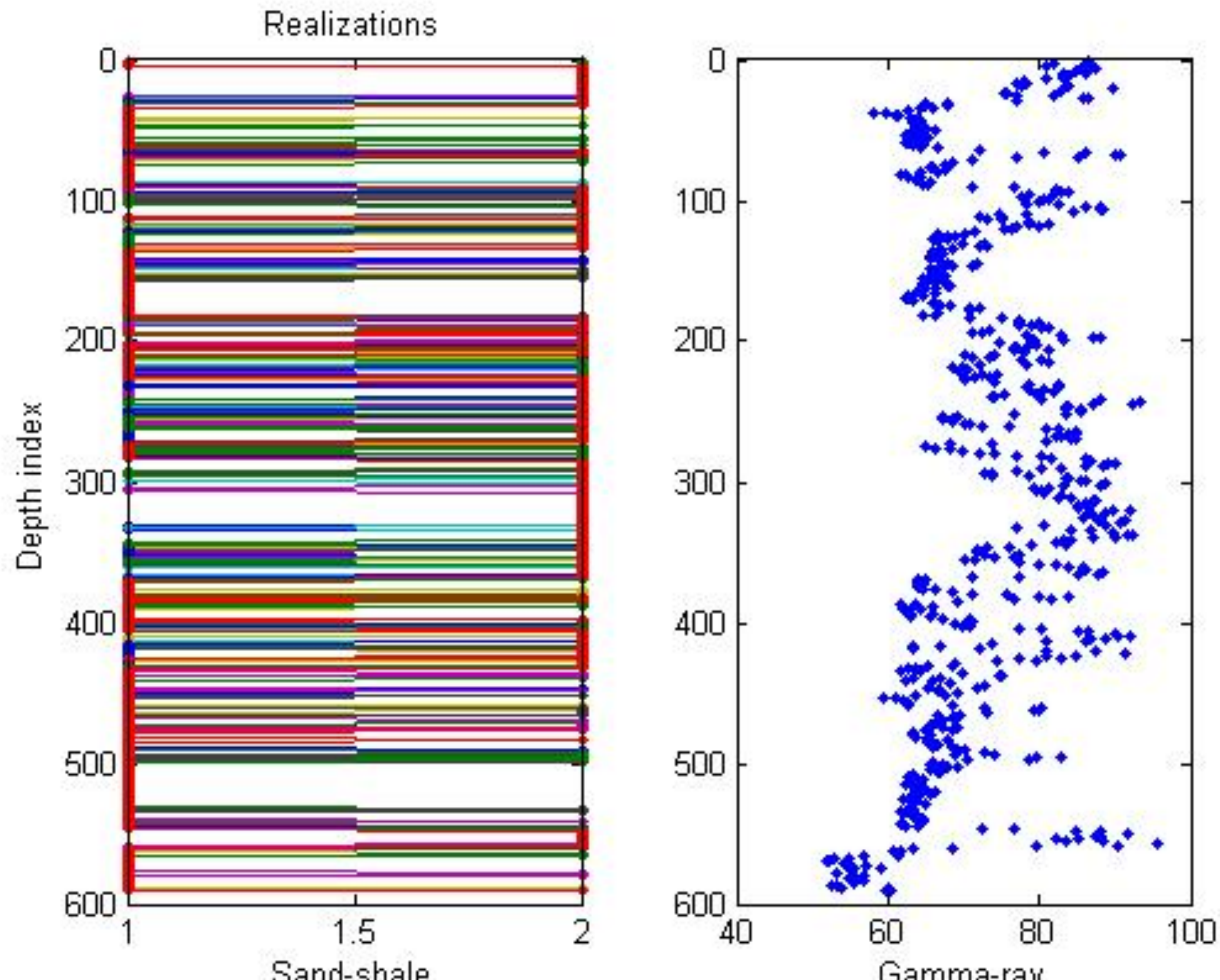


Result 2: Hidden Markov models

- Probability of sand (depth by depth) using only likelihood, and *no dependence* in prior.

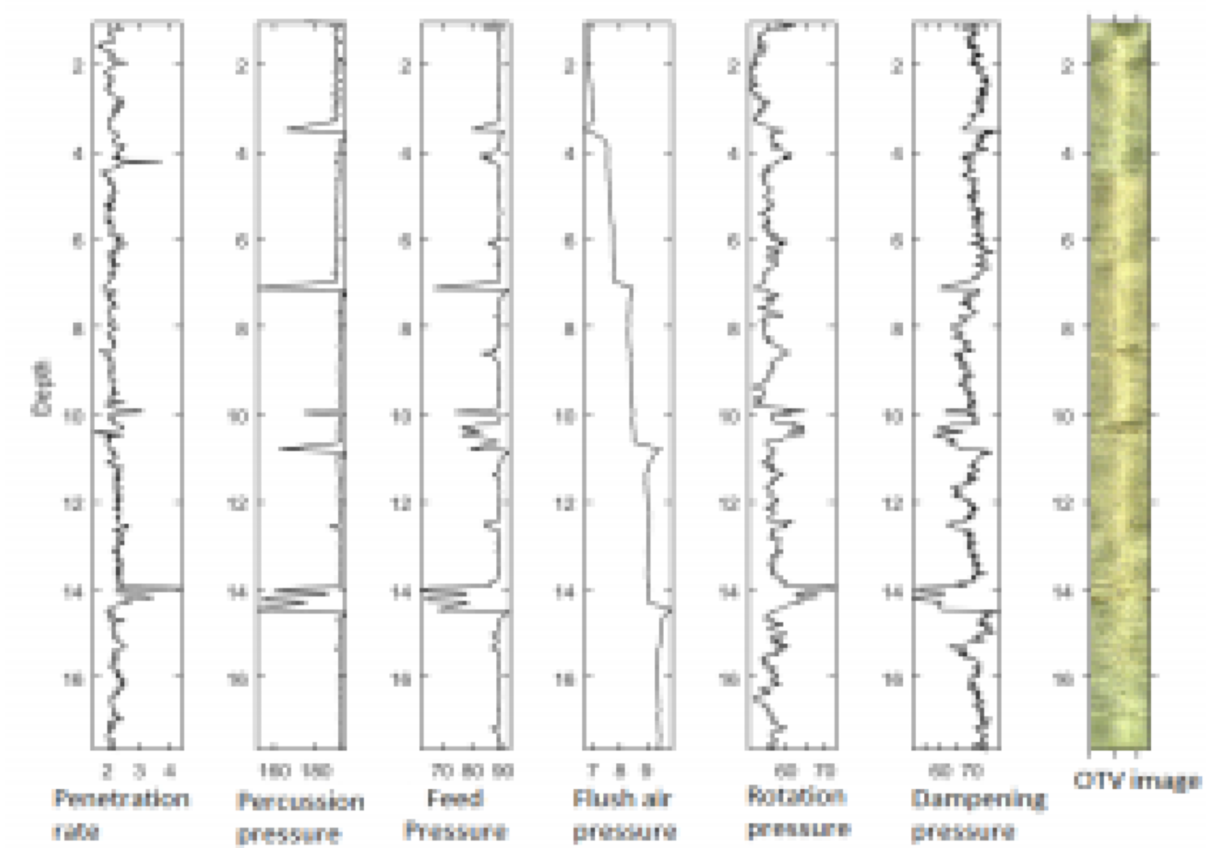


Samples of sand/shale, given gamma ray information



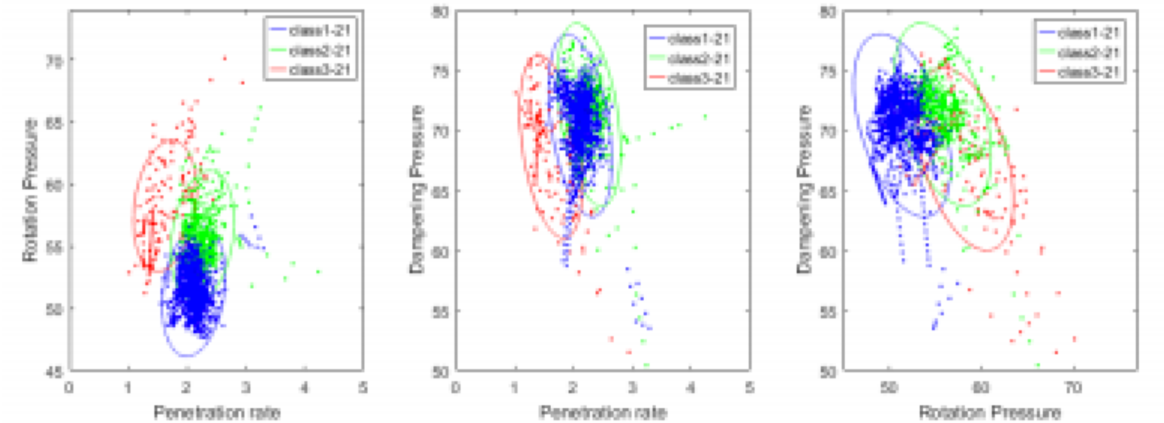
Applications

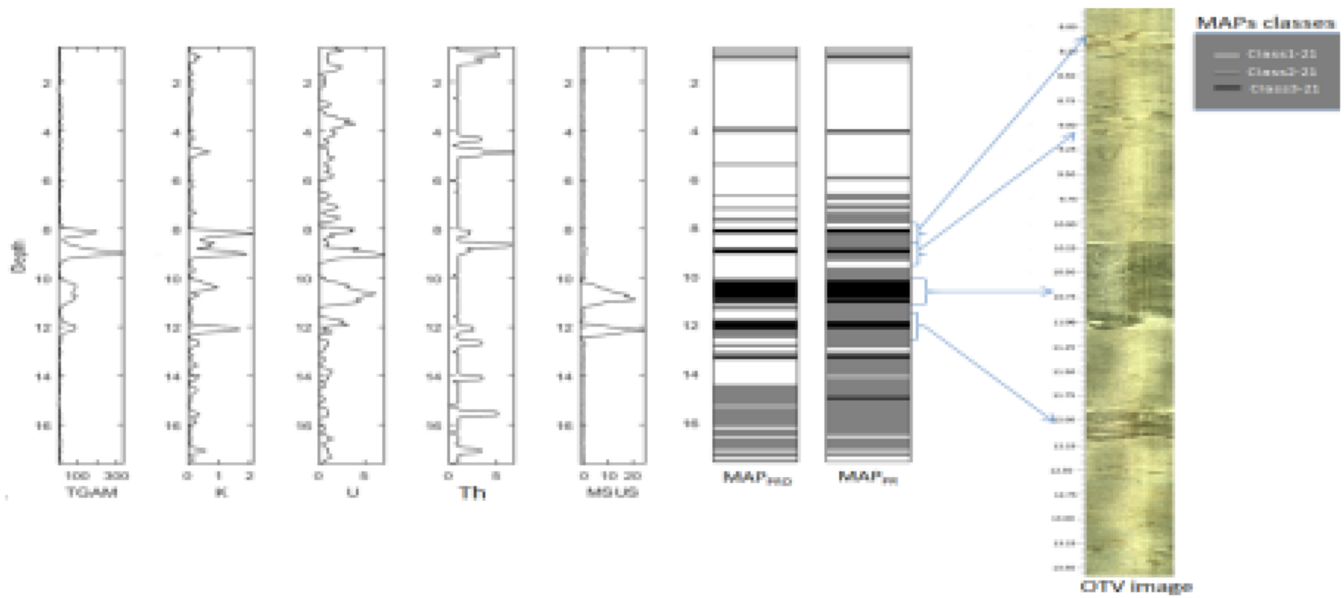
- Rock type classification from hardness data



Veena S Vezhapparambu

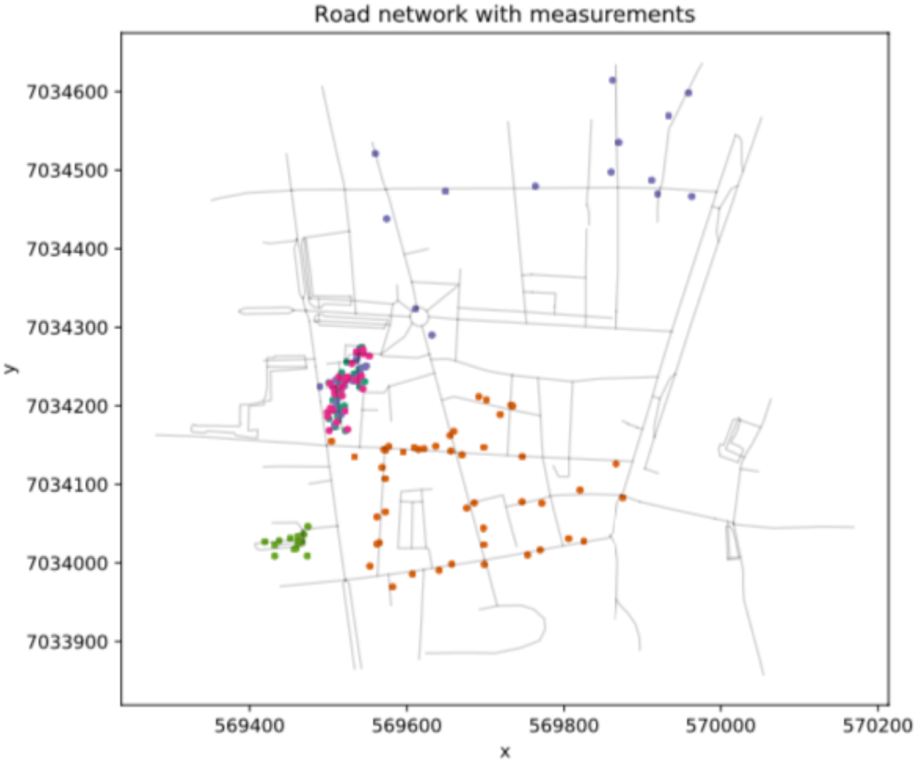
Applications





Applications

- Map matching

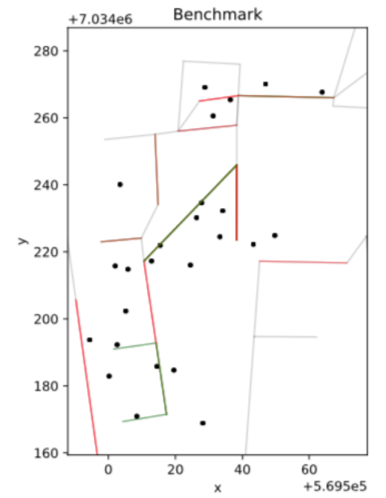


Øyvind Klåpbakken

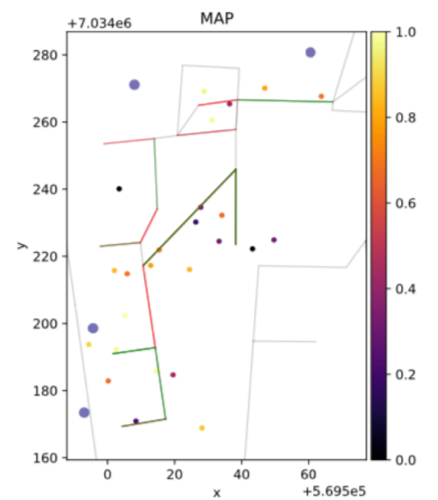
Applications

- Map matching

Borrow information to improve prediction. Better than simply using nearest road segment



(a) BM estimate for measurement with 25% missing GPS - measurements and presence of signal sources.

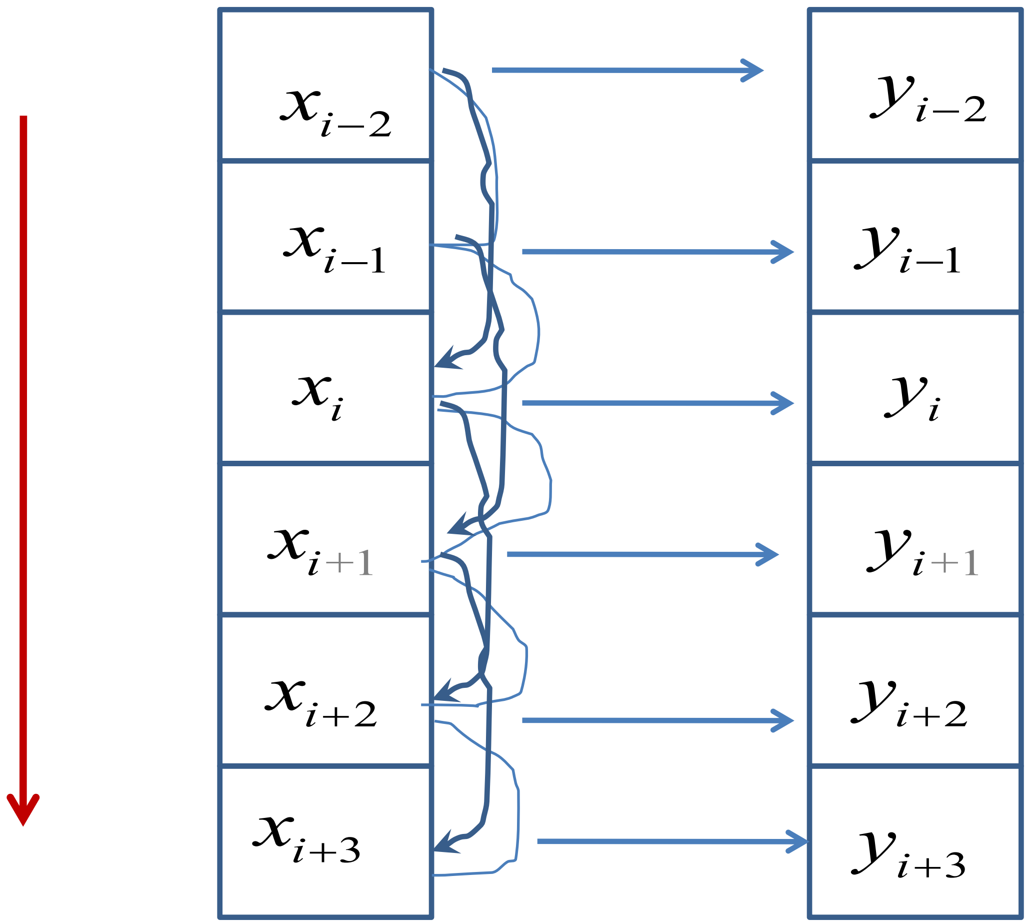


(b) MAP-estimate with $\sigma = 1$ and $\gamma = 1/100$ for measurement with 25% missing GPS - measurements and presence of signal sources.

Related topics:

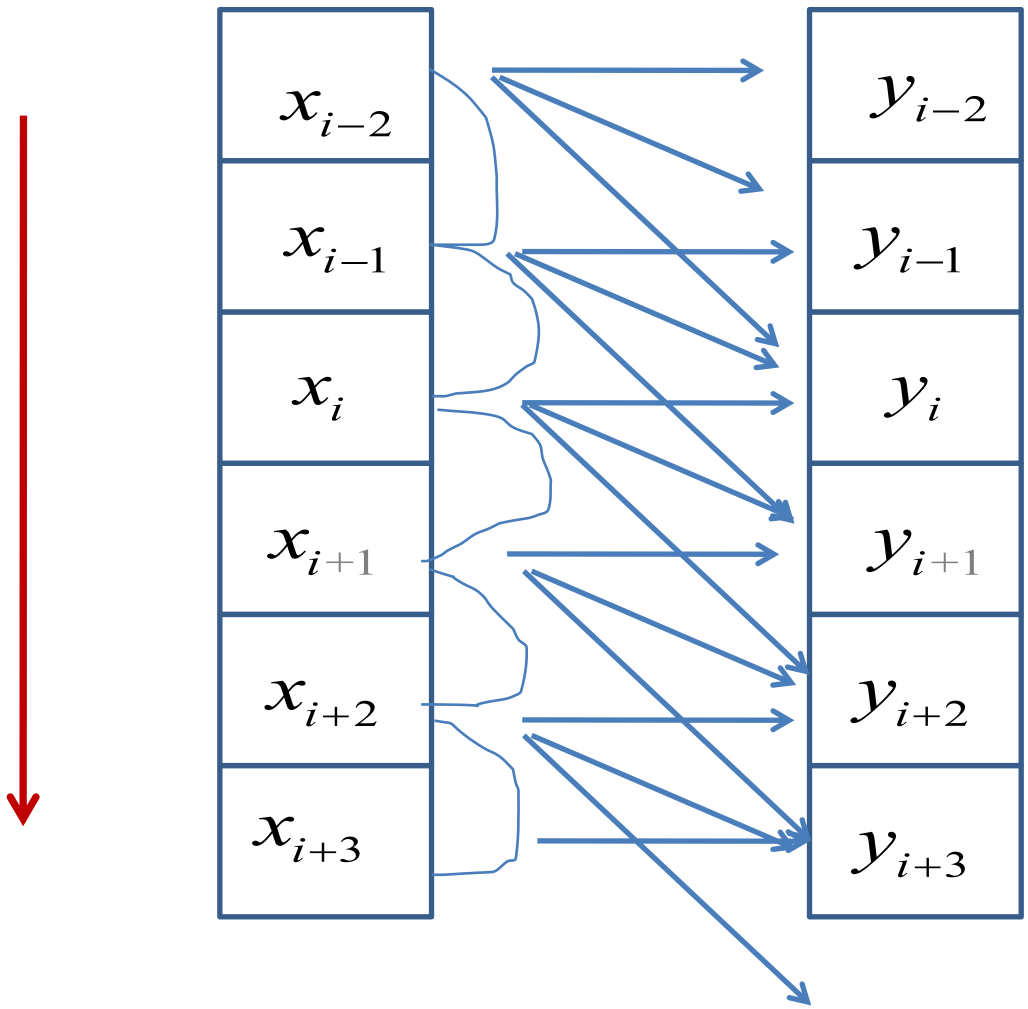
- Higher order Markov chains, coupling in the likelihood
- Markov random fields
- Graphical models
- Sequential updating

Higher-order models



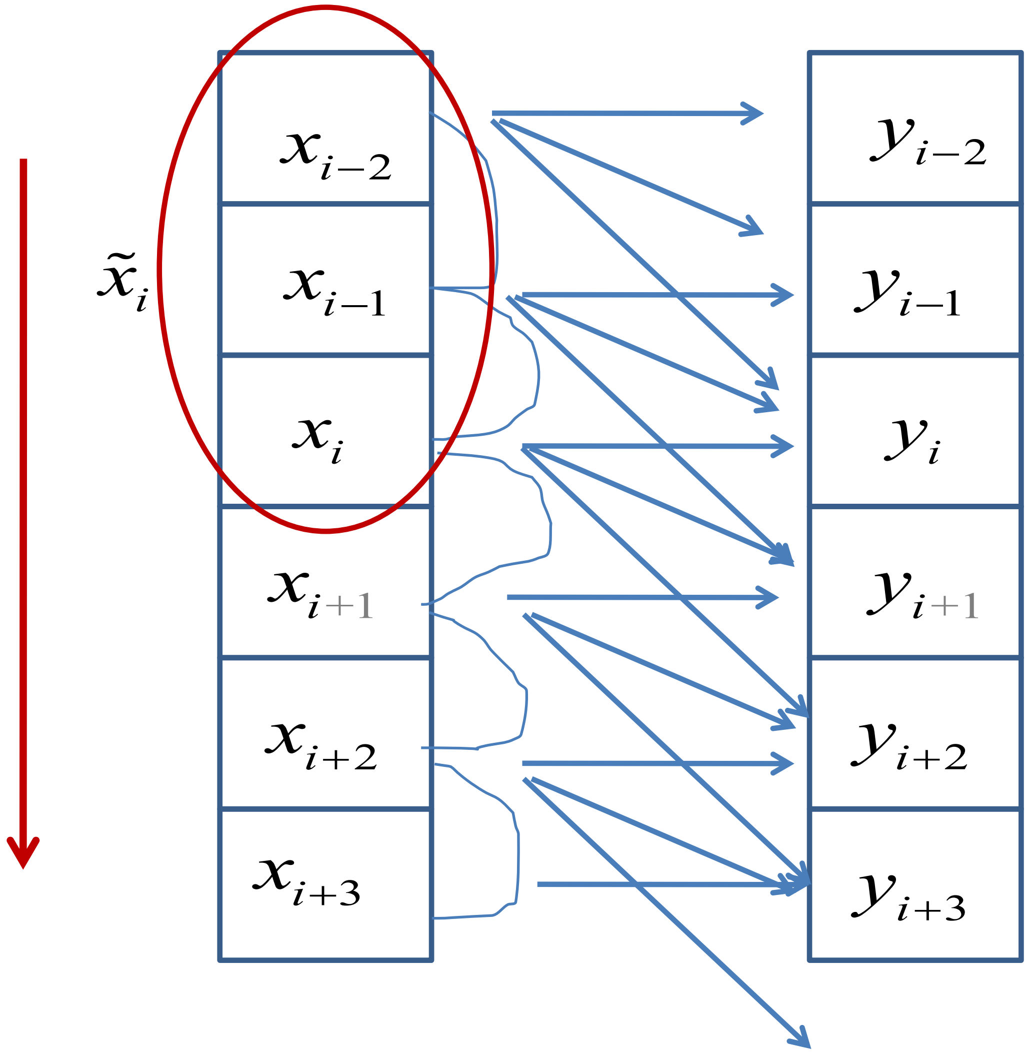
Higher-order dependence in latent model.

Higher-order models



«Convolved data», no longer conditional independence

Higher-order models



Introduce augmented variable.

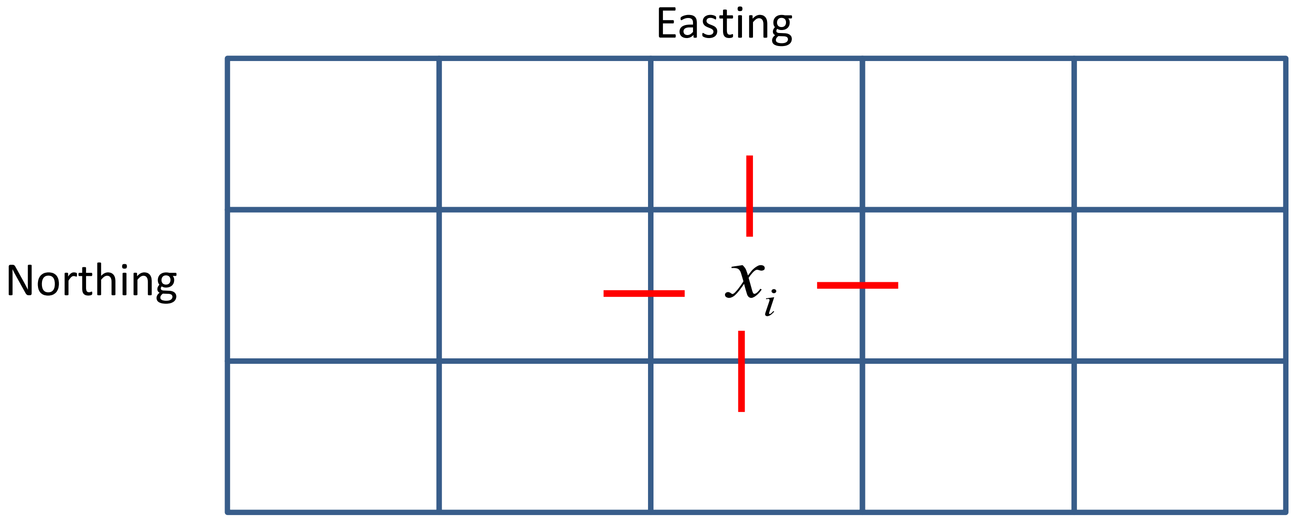
Usual Markov model for the augmented state space.

Challenge:

$O\left(nd^2\right)$

Markov random fields

First order Markov random field: 4 neighbors.



Markov random fields

Full conditional distribution depends only on neighbors.

$$p(x_i \mid x_{-i}) = p(x_i \mid x_j ; j \in N_i),$$

$$x_i \in \{1, \ldots, d\}, i = 1, \ldots, n = n_1 n_2$$

n_1 North grid dimension

n_2 East grid dimension

Markov random fields

Equivalent formulation using joint distribution (Hammersley-Clifford theorem):

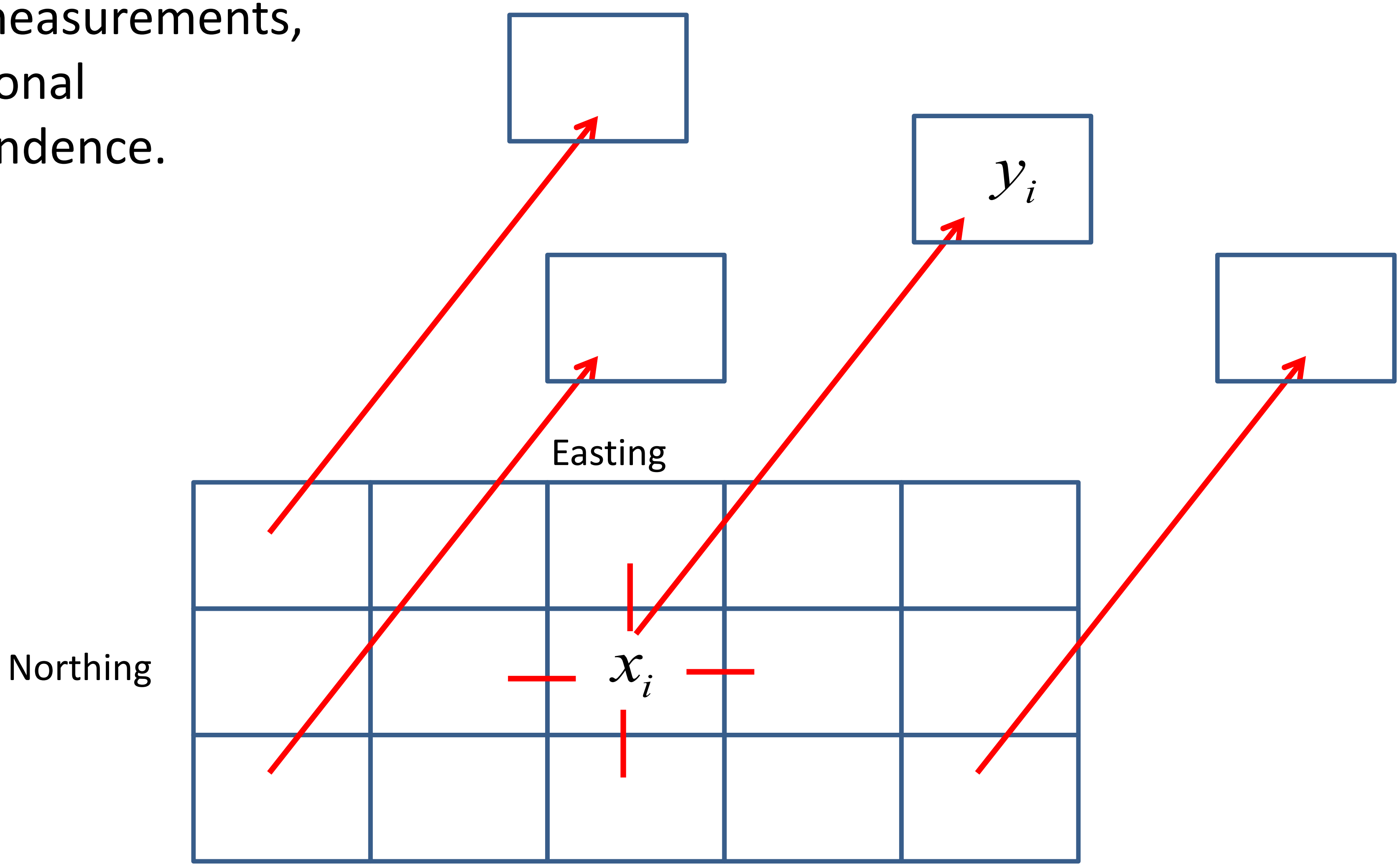
$$p(x_1,\ldots,x_n) = c\exp\left(\beta\sum_{i\sim j} I\left[x_i = x_j\right]\right),$$

i, j neighbors on the grid

$$c = \sum_{x_1=1}^{d}\ldots\sum_{x_n=1}^{d}\exp\left(\beta\sum_{i\sim j} I\left[x_i = x_j\right]\right)$$

Hidden Markov random fields

Local measurements, conditional independence.



Hidden Markov random fields

$$p(x_i = k \mid x_{-i}) = p(x_i = k \mid x_j; j \in N_i)$$

$$p\left(y_i \mid x_i\right), i = 1, \ldots, n$$

Data can be continuous or discrete. Latent variable is discrete.

Hidden Markov random fields

Main goal: posterior marginal probabilities or realizations

$$p(x_i \mid y_1,\ldots,y_n), k=\{1,..,d\}, i=\{1,..,n\},$$
$$x=\left(x_1,\ldots,x_n\right)\sim p(\boldsymbol{x} \mid y_1,\ldots,y_n)$$

For small grids, the forward-backward algorithm still works.
For large grids we use approximations or we use MCMC.

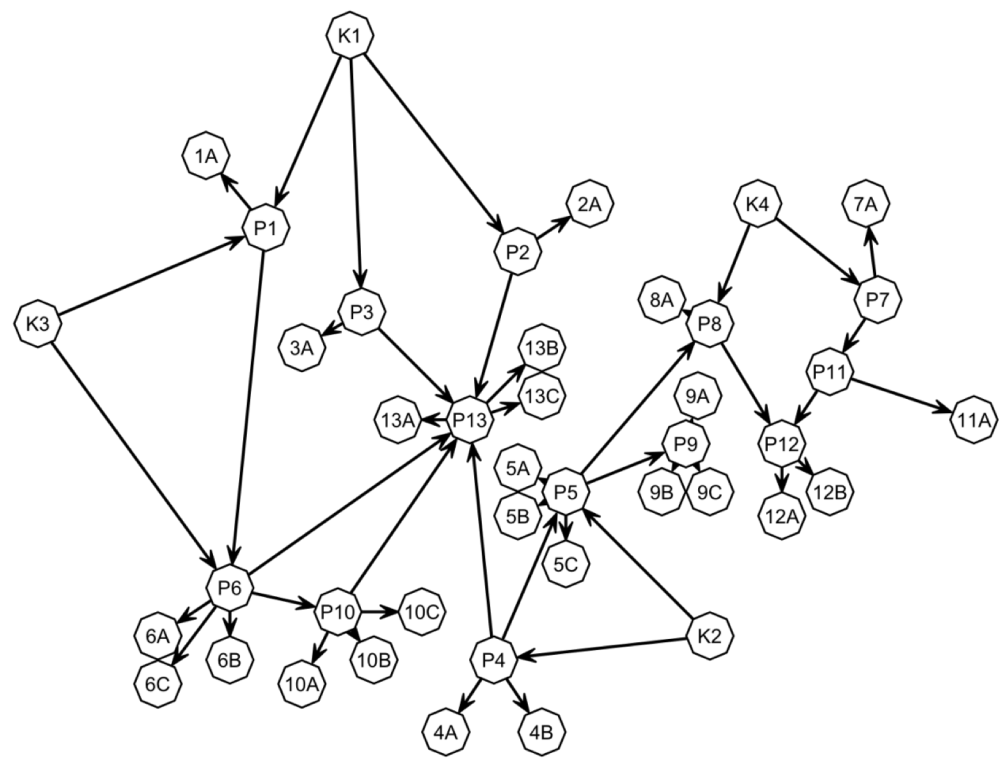
Conditional modeling (graphs)

$$p(x_1, x_2) = p(x_1)p(x_2 \mid x_1),$$

$$p(x_1, x_2) = p(x_2)p(x_1 \mid x_2)$$

$$p\left(x_2\right) = \sum_{x_1} p(x_1, x_2)$$

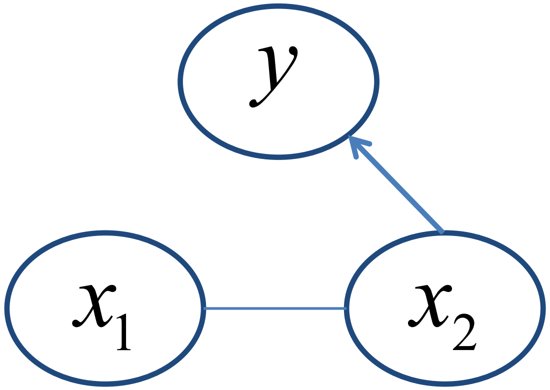
$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p(x_2)}$$



Smart ordering of variables

$$p(x_1, x_2 \mid y) = p(x_2 \mid y)\, p(x_1 \mid x_2, y),$$

$$p(x_1, x_2 \mid y) = p(x_2 \mid y)\, p(x_1 \mid x_2)$$



Project HMM

- Simulate a Markov chain and data, with known parameters.
- Run forward algorithm and fit maximum likelihood estimates from data.
- Run forward-backward algorithm and do marginal predictions and joint sampling, given parameter estimates.
- Do predictions using the data alone, without any dependence in the latent class variables.