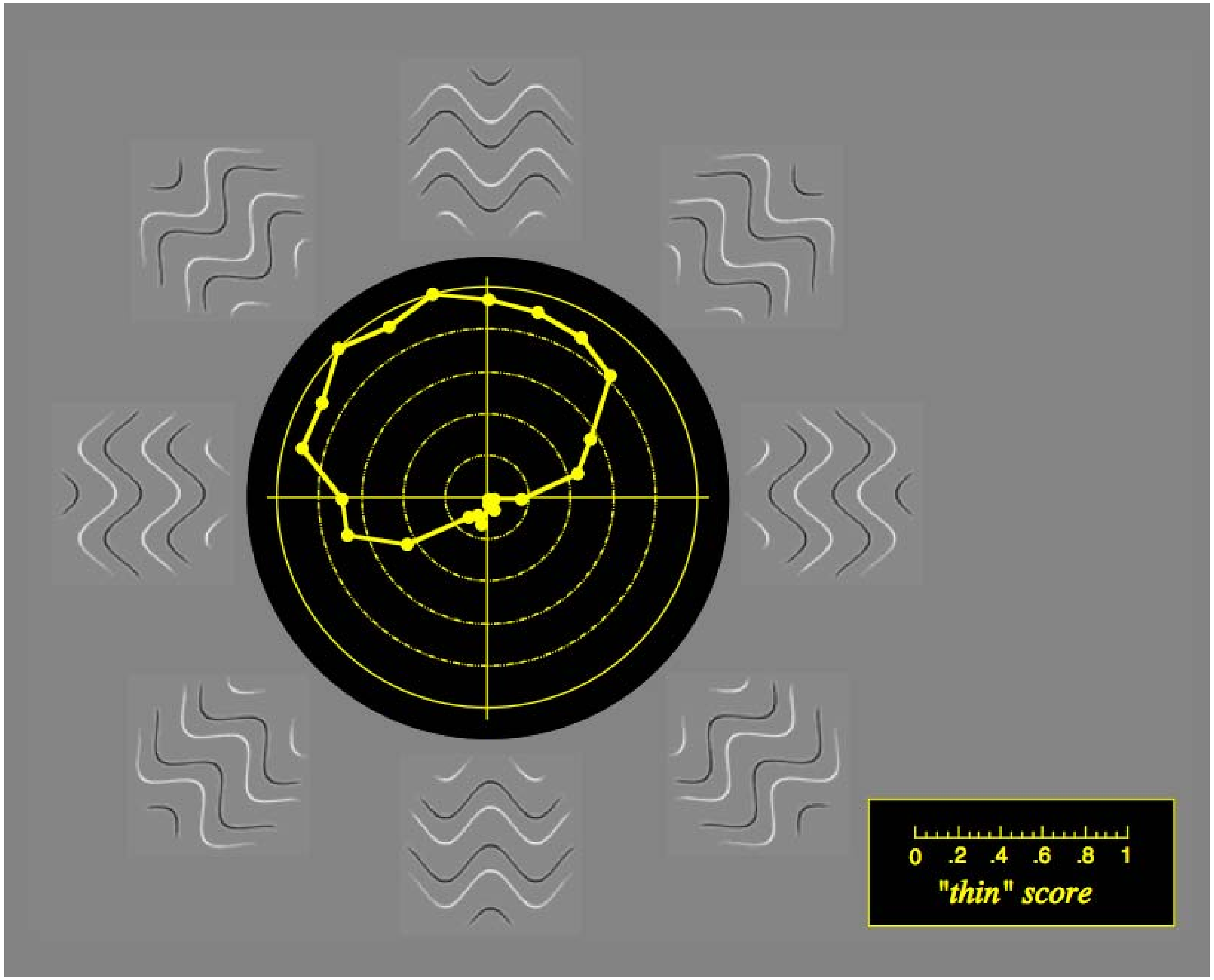
Bayesian inference

J. Daunizeau

Wellcome Trust Centre for Neuroimaging, London, UK
Institute of Empirical Research in Economics, Zurich, Switzerland



Prior knowledge on the illumination position
Mamassian & Goutcher 2001.

Overview of the talk

1 Probabilistic modelling and representation of uncertainty

1.1 Bayesian paradigm

1.2 Hierarchical models

1.3 Frequentist versus Bayesian inference

2 Numerical Bayesian inference methods

2.1 Sampling methods

2.2 Variational methods (EM, VB)

3 SPM applications

3.1 aMRI segmentation

3.2 fMRI time series analysis with spatial priors

3.3 Dynamic causal modelling

3.4 EEG source reconstruction

Overview of the talk

1 Probabilistic modelling and representation of uncertainty

1.1 Bayesian paradigm

1.2 Hierarchical models

1.3 Frequentist versus Bayesian inference

2 Numerical Bayesian inference methods

2.1 Sampling methods

2.2 Variational methods (EM, VB)

3 SPM applications

3.1 aMRI segmentation

3.2 fMRI time series analysis with spatial priors

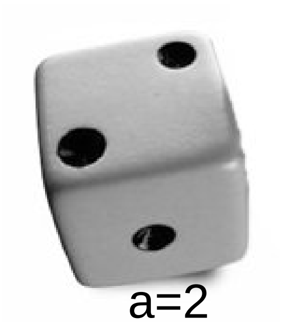
3.3 Dynamic causal modelling

3.4 EEG source reconstruction

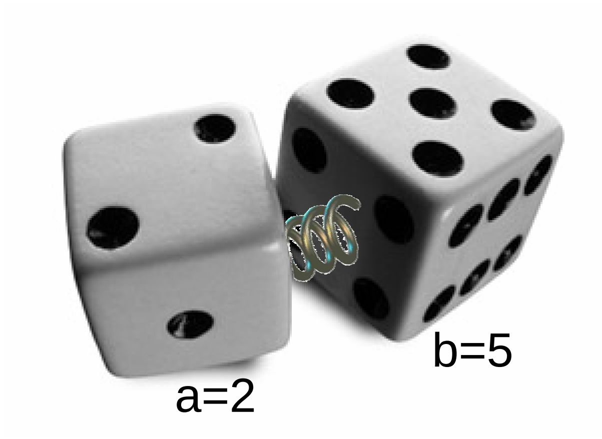
Bayesian paradigm

theory of probability

Degree of plausibility *desiderata*:
- should be represented using real numbers (D1)
- should conform with intuition (D2)
- should be consistent (D3)



• normalization: $$\sum_a P(a) = 1$$

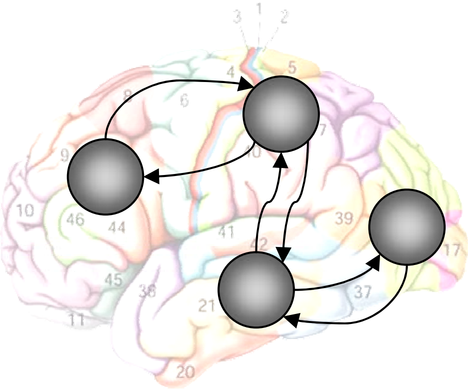


• marginalization: $$P(b) = \sum_a P(a,b)$$

• conditioning : (*Bayes rule*) $$P(a,b) = P(a|b)\,P(b) = P(b|a)\,P(a)$$

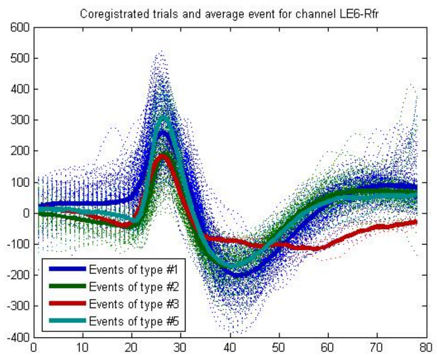
Bayesian paradigm

Likelihood and priors



generative model m

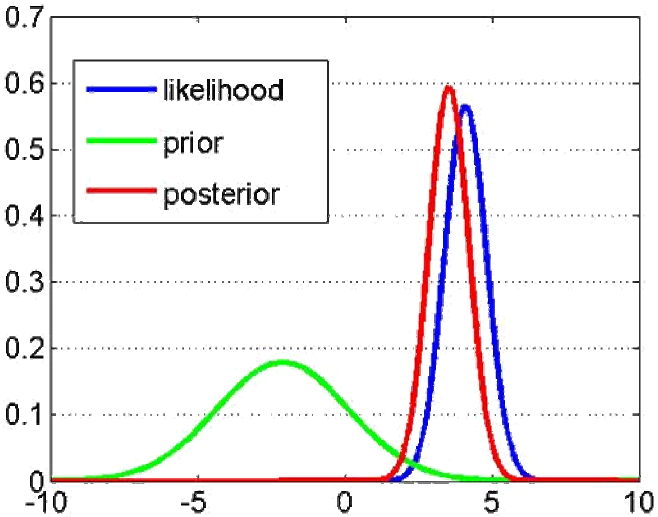
y



Likelihood: $p(y|\theta,m)$

Prior: $p(\theta|m)$

Bayes rule: $$p(\theta|y,m)=\frac{p(y|\theta,m)\,p(\theta|m)}{p(y|m)}$$



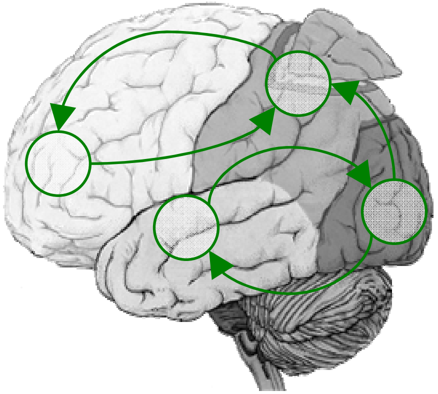
Bayesian paradigm

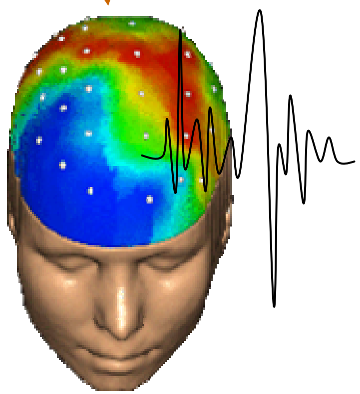
forward and inverse problems

forward problem

$$p\left(y\middle|\vartheta,m\right)$$

likelihood





posterior distribution

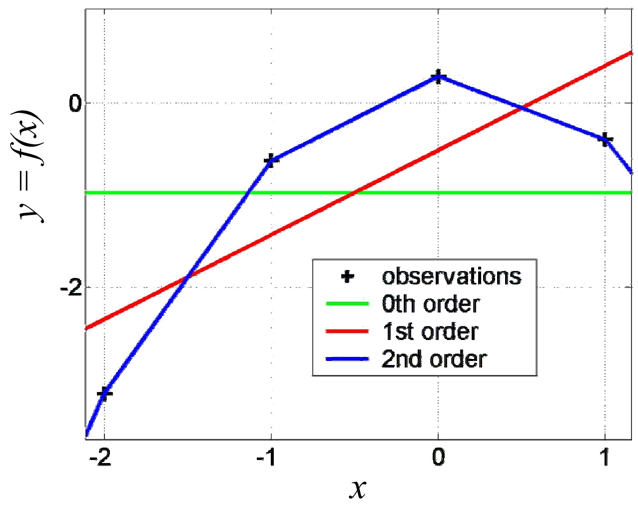
$$p\left(\vartheta\middle|y,m\right)$$

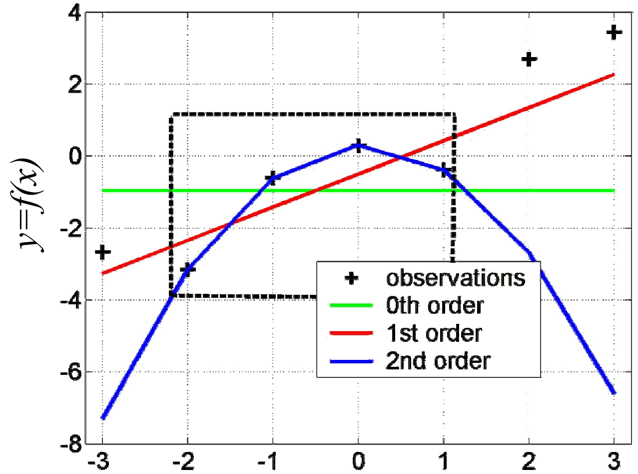
inverse problem

Bayesian paradigm

Model comparison

Principle of parsimony :
« plurality should not be assumed without necessity »

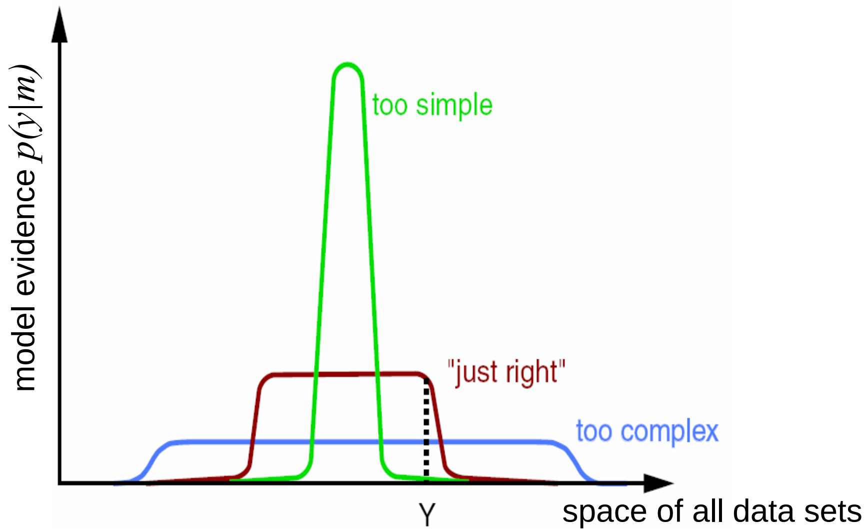




Model evidence:

$$p(y|m) = \int p(y|\theta, m) p(\theta|m) d\theta$$

"Occam's razor" :

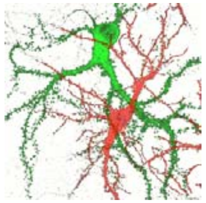


Hierarchical models

principle

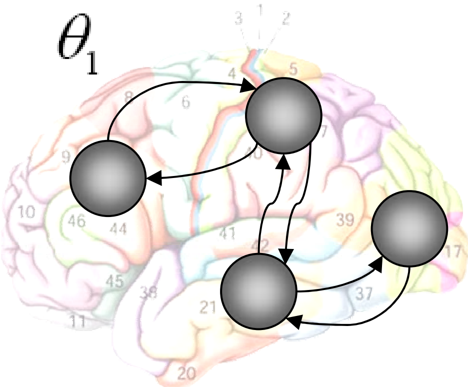
$\vdots$

θ_2

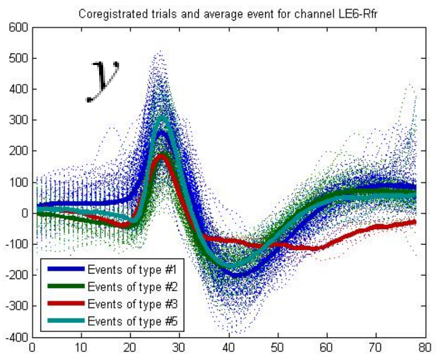


$$p(\theta_2|\theta_3, m)$$

θ_1



$$p(\theta_1|\theta_2, m)$$



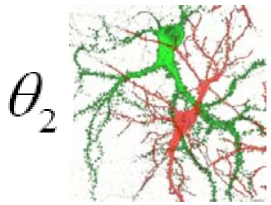
$$p(y|\theta_1, m)$$

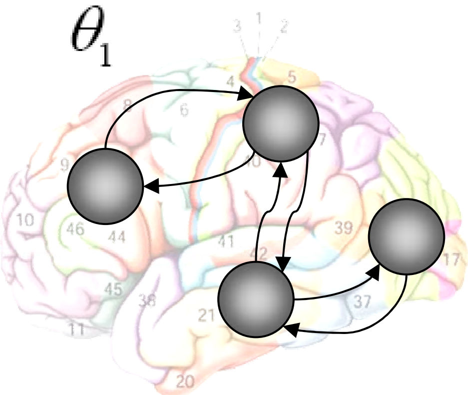
hierarchy

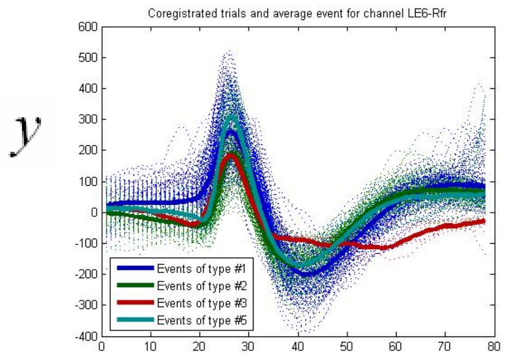
causality

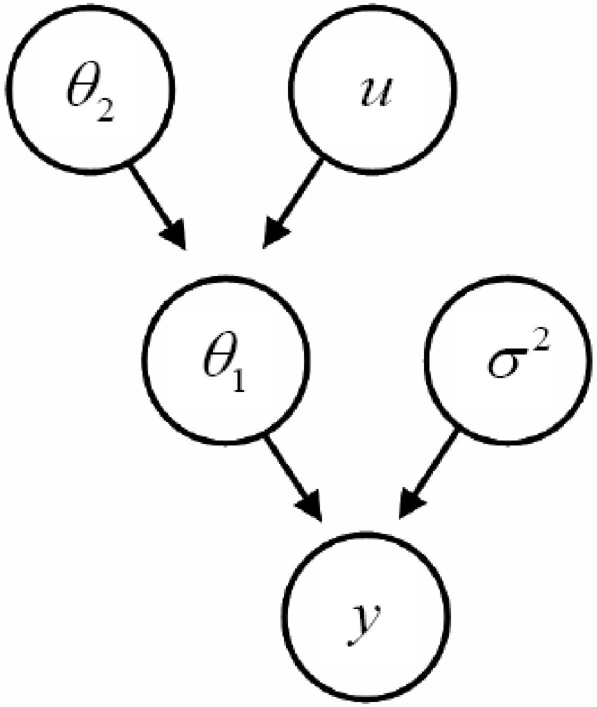
Hierarchical models

directed acyclic graphs (DAGs)









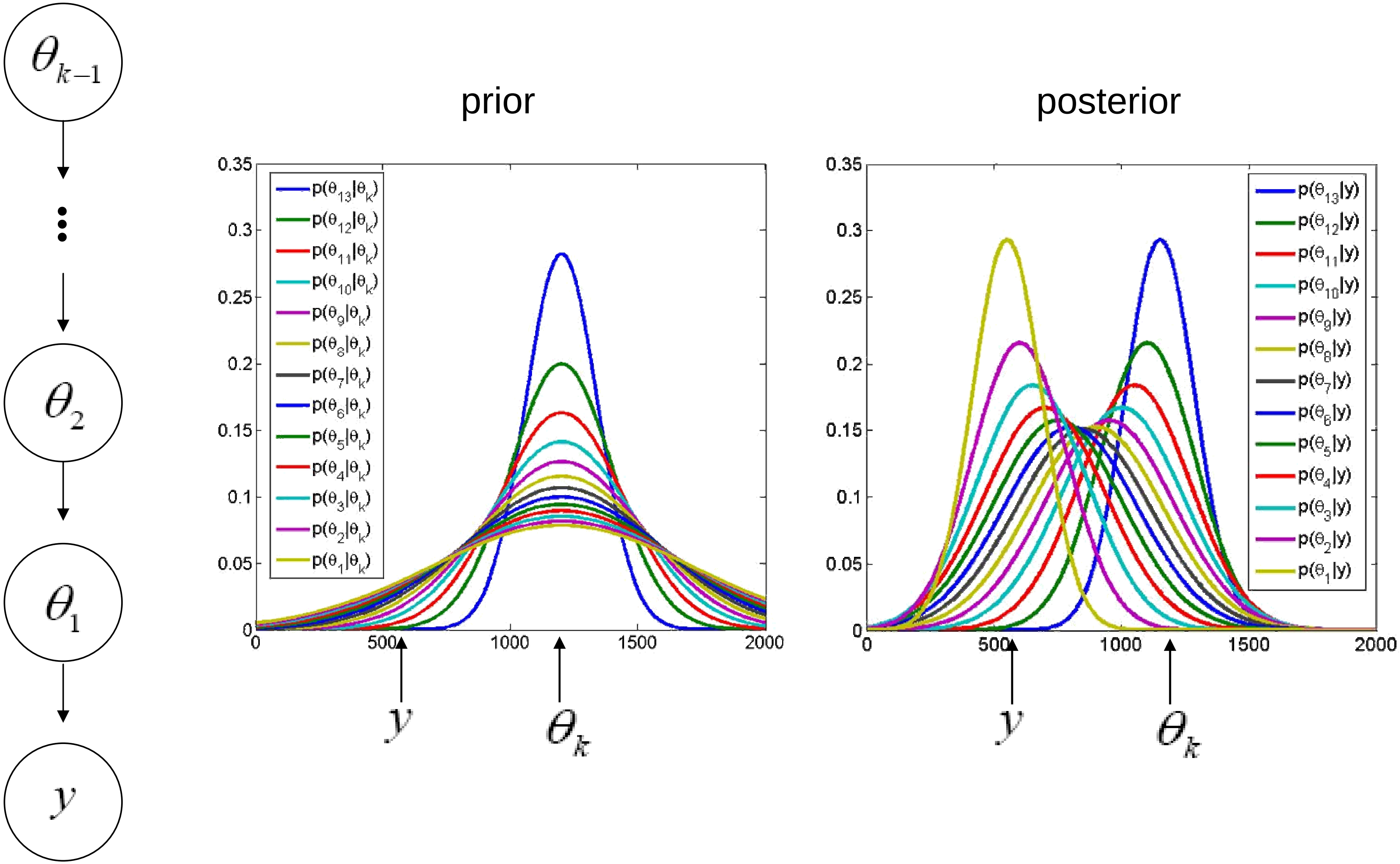
$$p(\theta_1 | \theta_2, u, m)$$

$$p(y | \theta_1, \sigma^2, m)$$

$$p(\theta | m) = \prod_j p(\theta_j | par(\theta_j), m)$$

Hierarchical models

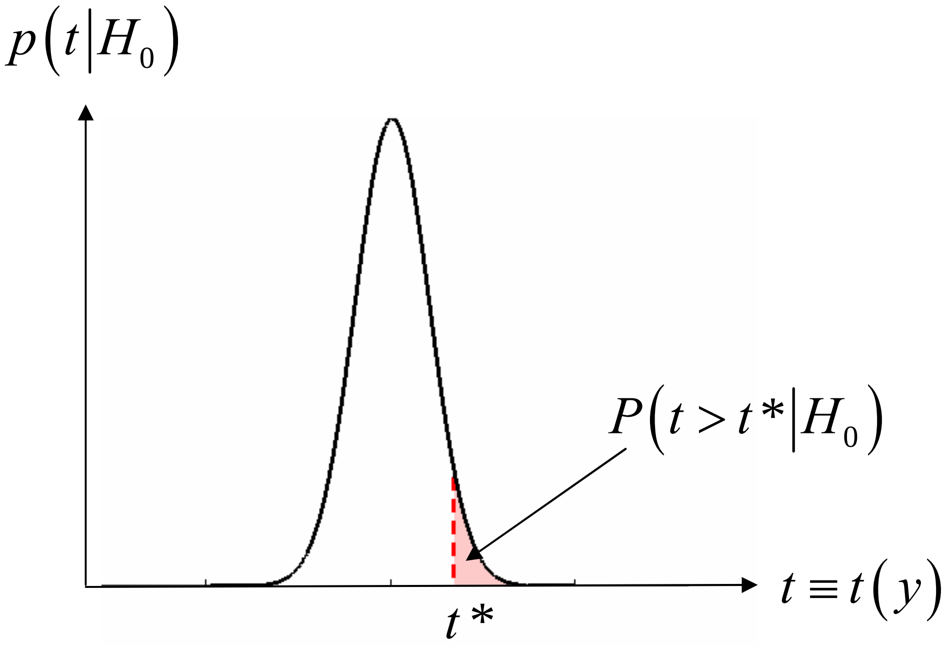
univariate linear hierarchical model



Frequentist versus Bayesian inference

a (quick) note on hypothesis testing

- define the null, e.g.: $H_0 : \theta = 0$

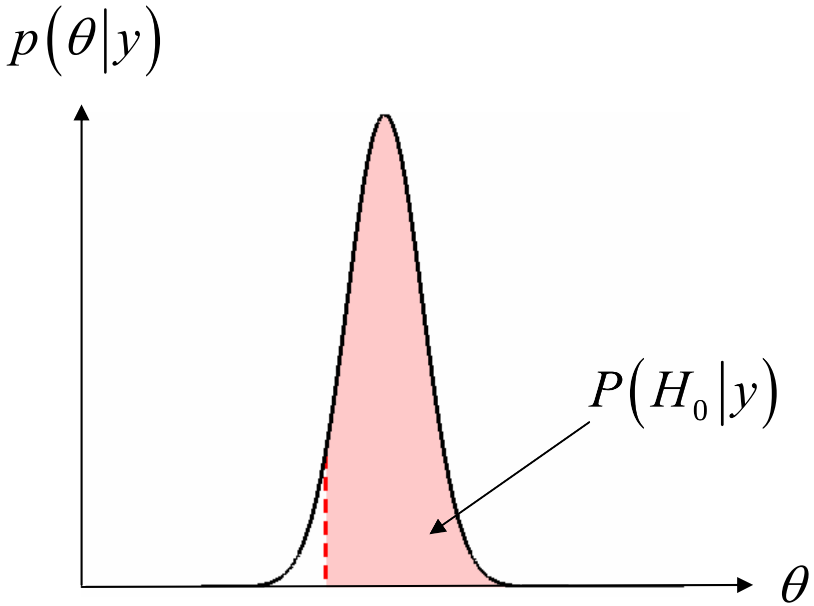


- estimate parameters (obtain test stat.)
- apply decision rule, i.e.:

if $P\left(t > t^* \middle| H_0\right) \leq \alpha$ then reject H0

classical SPM

- invert model (obtain posterior pdf)



- define the null, e.g.: $H_0 : \theta > 0$
- apply decision rule, i.e.:

if $P\left(H_0 \middle| y\right) \geq \alpha$ then accept H0

Bayesian PPM

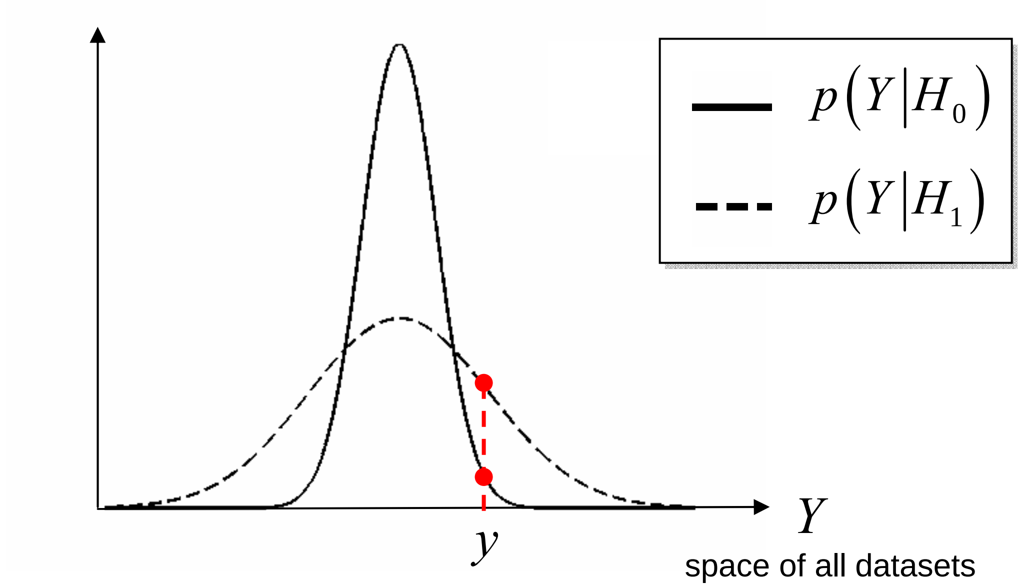
Frequentist versus Bayesian inference

Bayesian inference: what about point hypothesis testing?

- define the null and the alternative hypothesis *in terms of priors*, e.g.:

$$H_0 : p(\theta | H_0) = \begin{cases} 1 & \text{if } \theta = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$H_1 : p(\theta | H_1) = N(0, \Sigma)$$



space of all datasets

- invert both generative models (obtain both model evidences)
- apply decision rule, i.e.:

$$\text{if } \quad \frac{P(H_0|y)}{P(H_1|y)} \leq 1 \quad \text{then reject H0}$$

Overview of the talk

1 Probabilistic modelling and representation of uncertainty

1.1 Bayesian paradigm

1.2 Hierarchical models

1.3 Frequentist versus Bayesian inference

2 Numerical Bayesian inference methods

2.1 Sampling methods

2.2 Variational methods (EM, VB)

3 SPM applications

3.1 aMRI segmentation

3.2 fMRI time series analysis with spatial priors

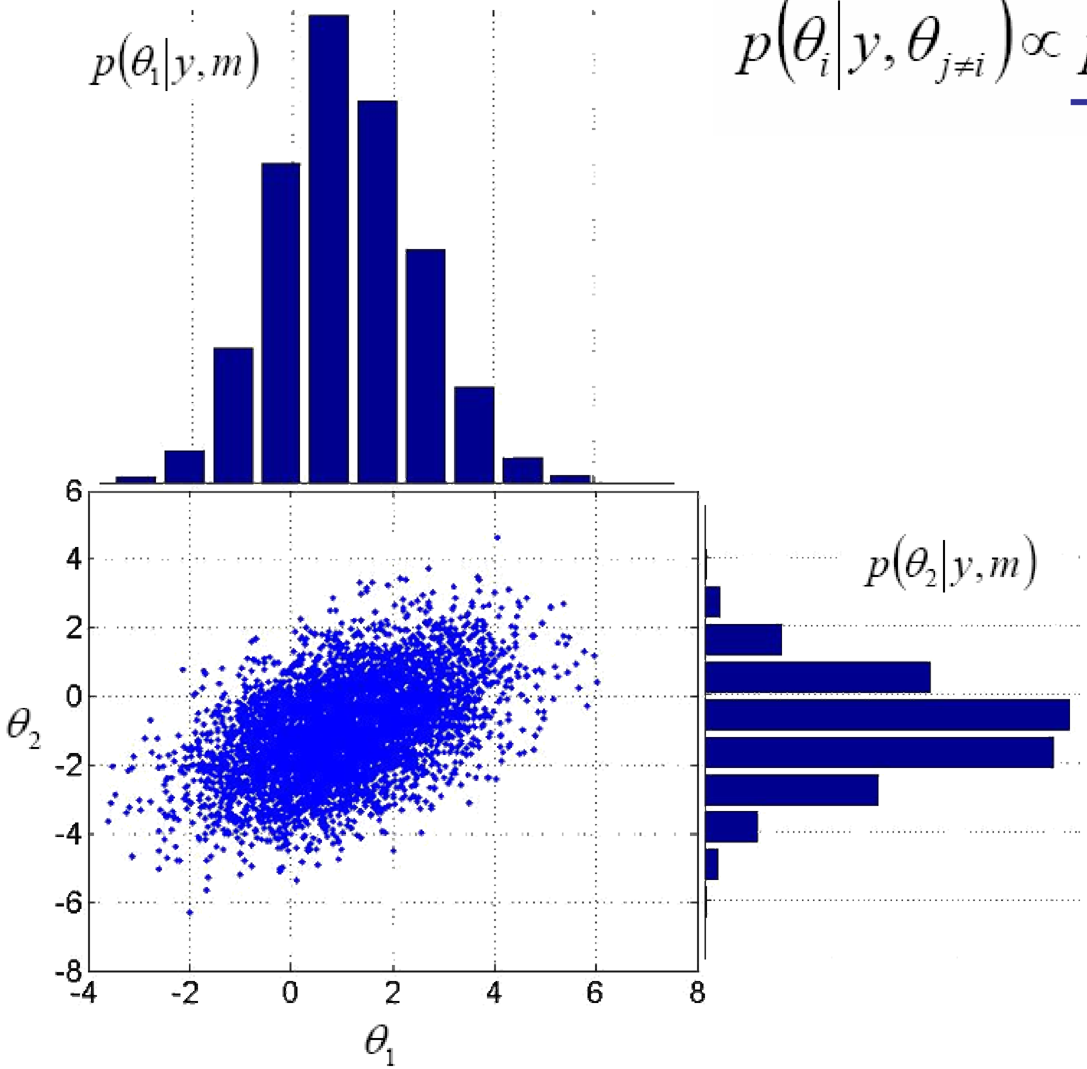
3.3 Dynamic causal modelling

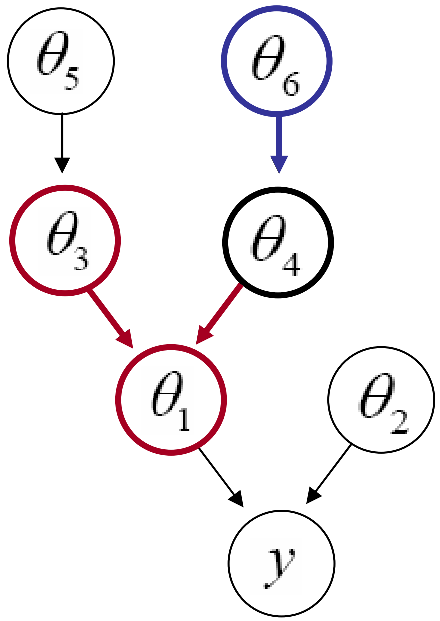
3.4 EEG source reconstruction

Sampling methods

MCMC example: Gibbs sampling

$$p\left(\theta_i \middle| y, \theta_{j\neq i}\right) \propto \underline{p\left(\theta_i \middle| par(\theta_i)\right)} \underline{\prod_{j=ch(i)} p\left(\theta_j \middle| par(\theta_j)\right)}$$



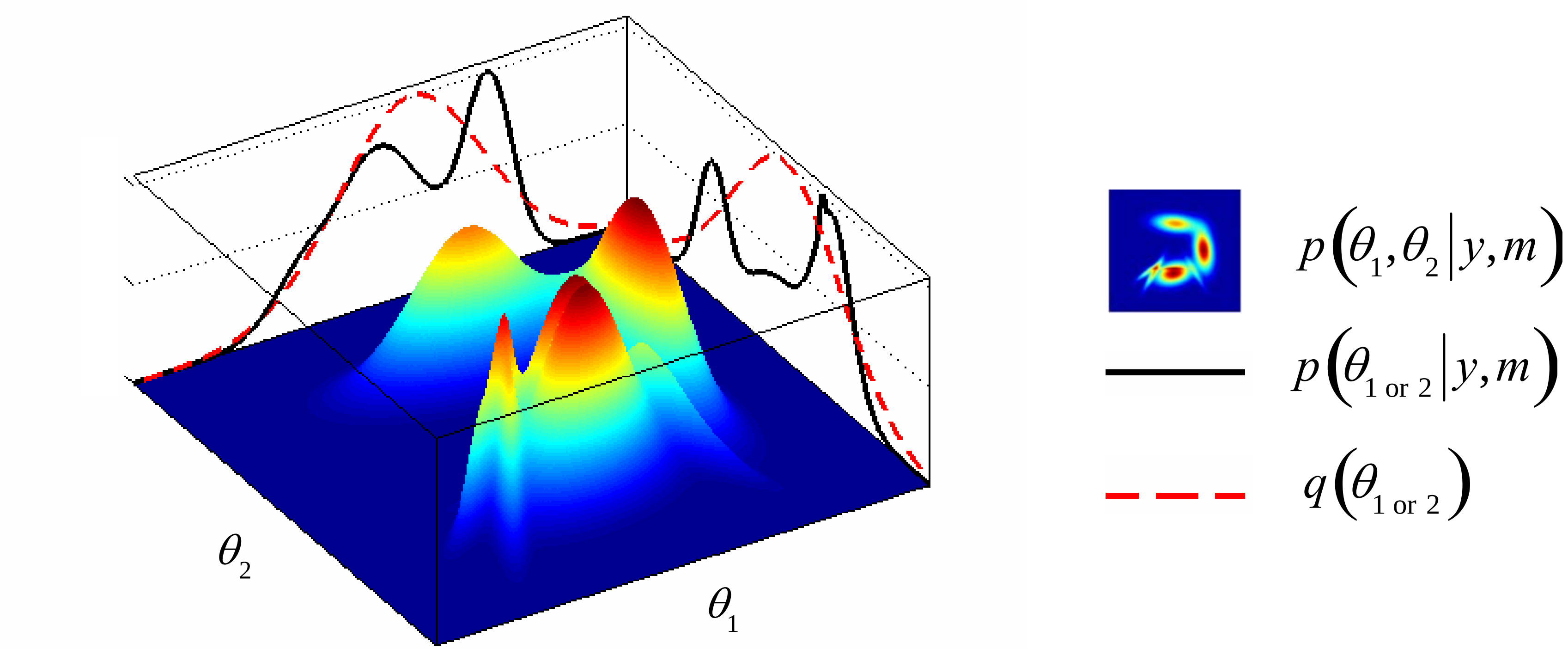


$$\frac{1}{N}\sum_{n=1}^{N} p\left(y \middle| \theta^{(n)}, m\right) \approx p\left(y \middle| m\right)$$

Variational methods

$$\ln p(y|m) = \underbrace{\langle \ln p(\theta, y|m) \rangle_q + S(q)}_{\text{free energy } F(q)} + D_{KL}\left(q(\theta); p(\theta|y,m)\right)$$

VB/EM/ReML: find (iteratively) the “variational” posterior $q(\theta)$ which maximizes the free energy $F(q)$ under some approximation



Overview of the talk

1 Probabilistic modelling and representation of uncertainty

1.1 Bayesian paradigm

1.2 Hierarchical models

1.3 Frequentist versus Bayesian inference

2 Numerical Bayesian inference methods

2.1 Sampling methods

2.2 Variational methods (EM, VB)

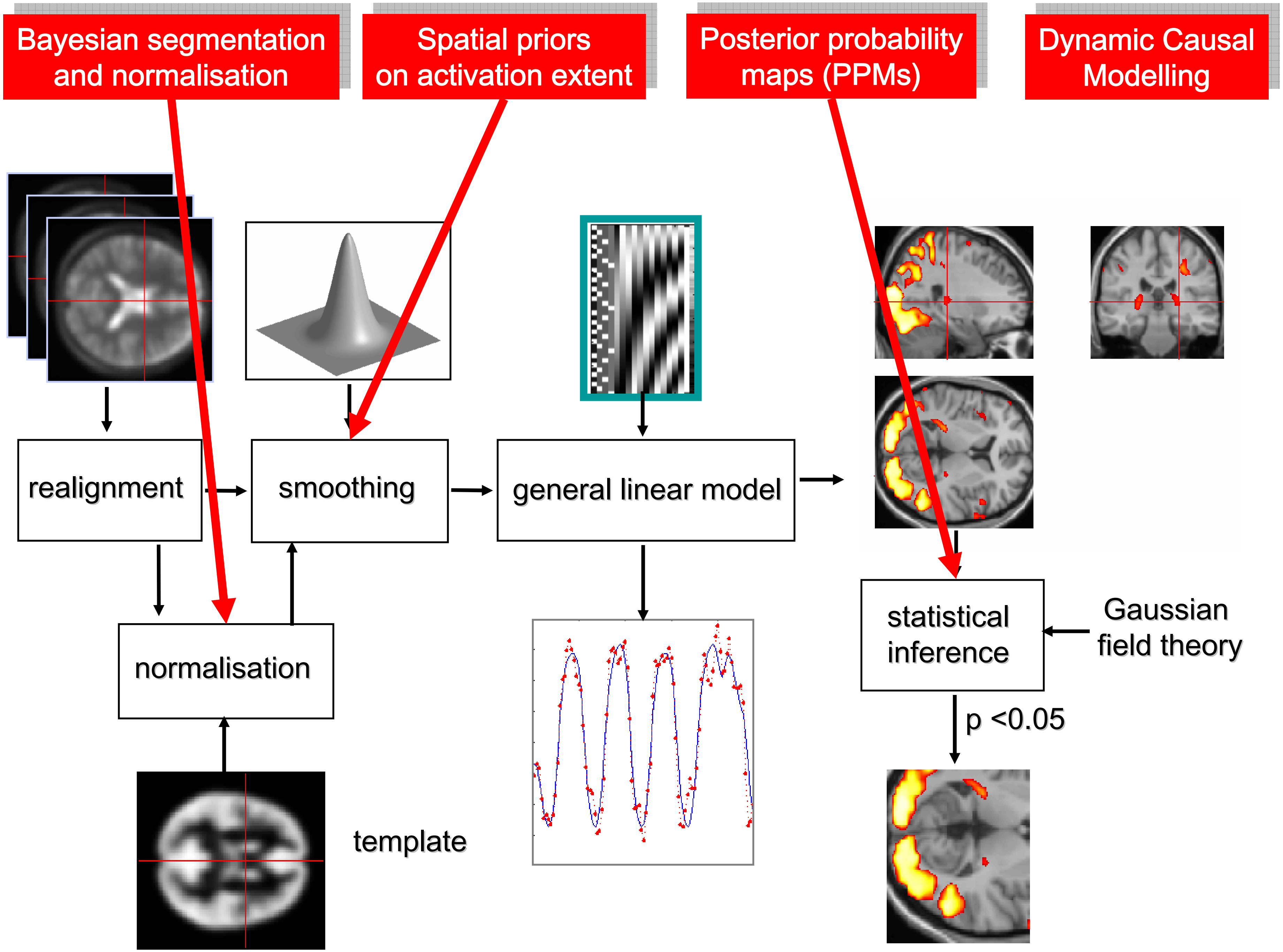
3 SPM applications

3.1 aMRI segmentation

3.2 fMRI time series analysis with spatial priors

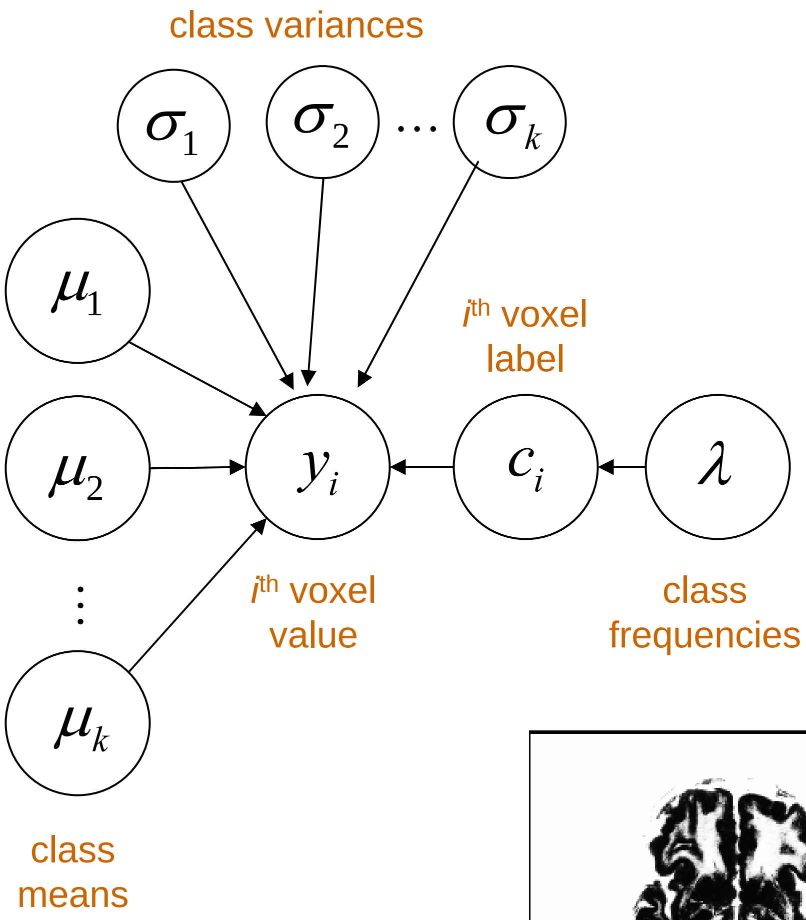
3.3 Dynamic causal modelling

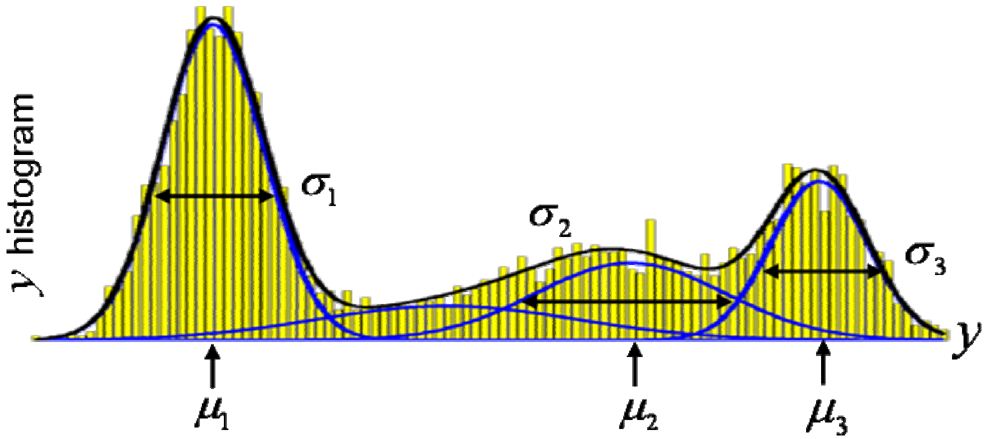
3.4 EEG source reconstruction

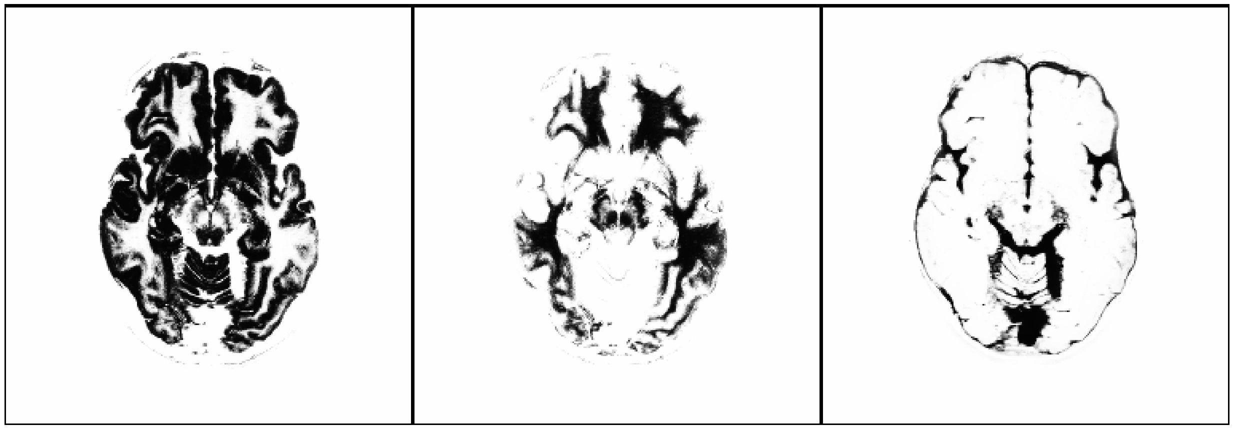


Bayesian segmentation and normalisation
Spatial priors on activation extent
Posterior probability maps (PPMs)
Dynamic Causal Modelling
realignment
smoothing
general linear model
normalisation
template
statistical inference
Gaussian field theory
p <0.05

aMRI segmentation





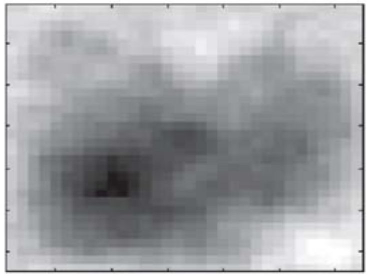


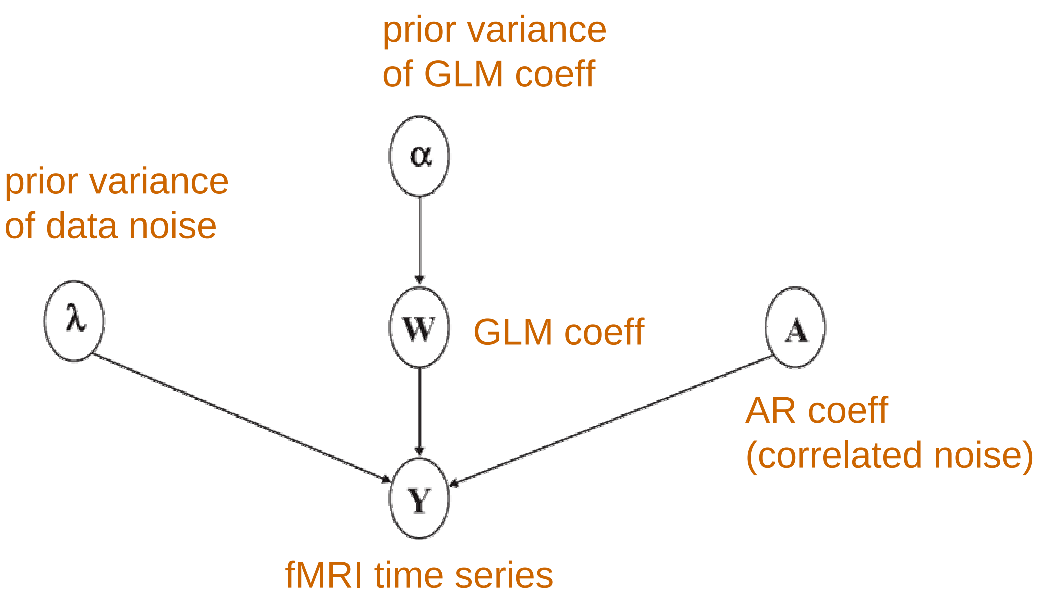
grey matter | white matter | CSF

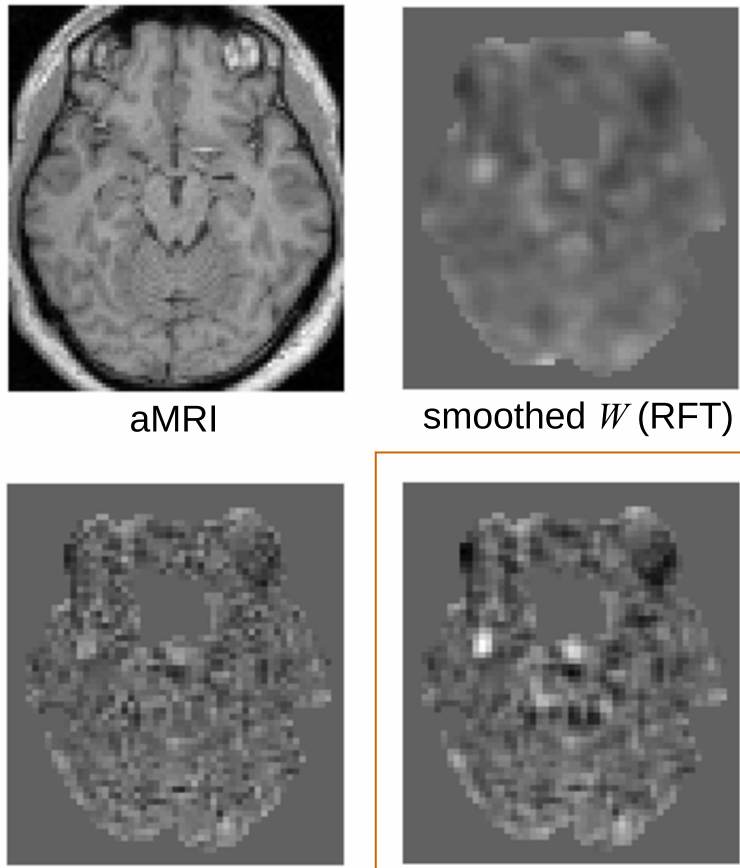
[Ashburner et al., Human Brain Function, 2003]

fMRI time series analysis with spatial priors

		1		
	2	-8	2	
1	-8	20	-8	1
	2	-8	2	
		1		







aMRI

smoothed W (RFT)

ML estimate of W

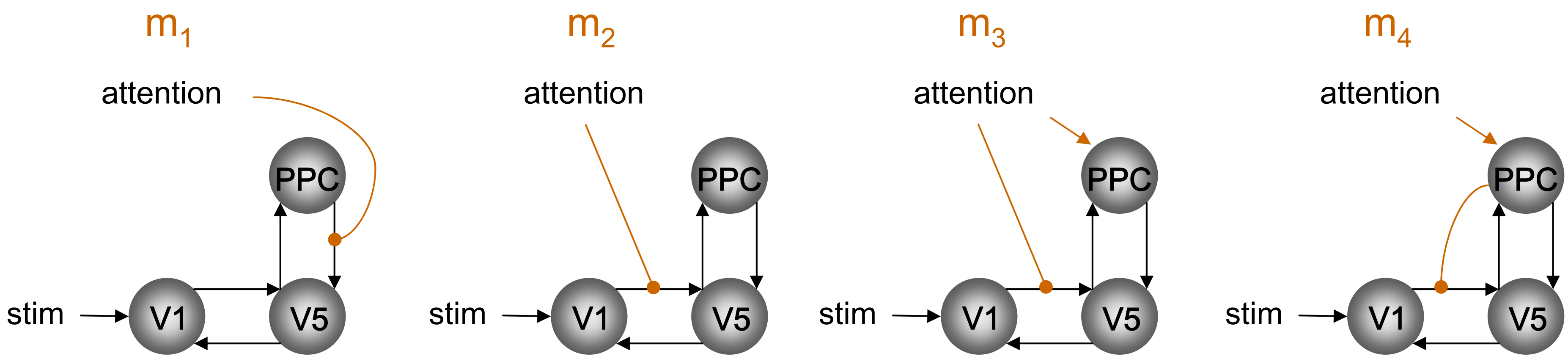
VB estimate of W

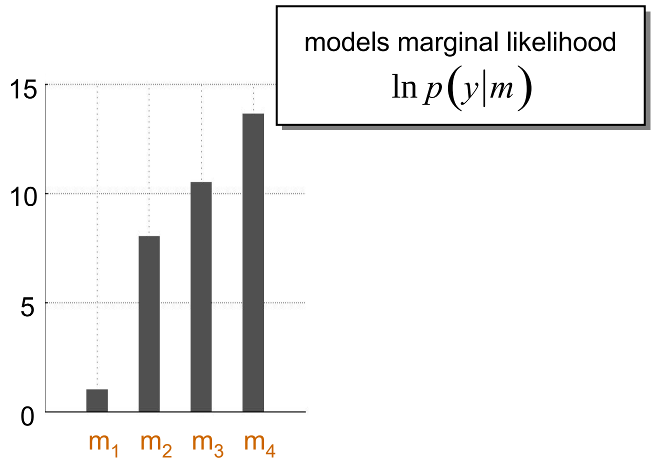
PPM

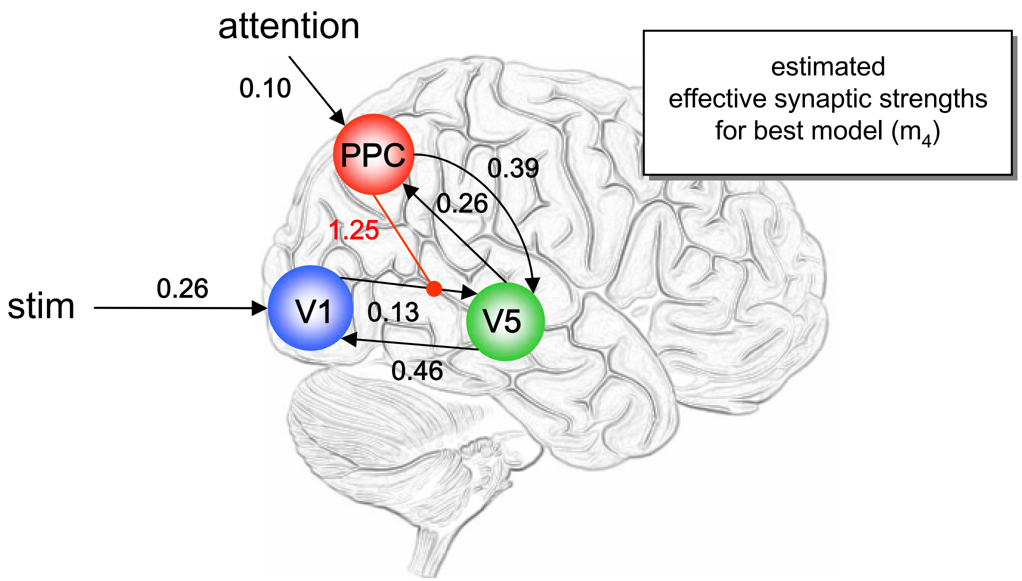
[Penny et al., Neuroimage, 2004]

Dynamic causal modelling

model comparison on network structures





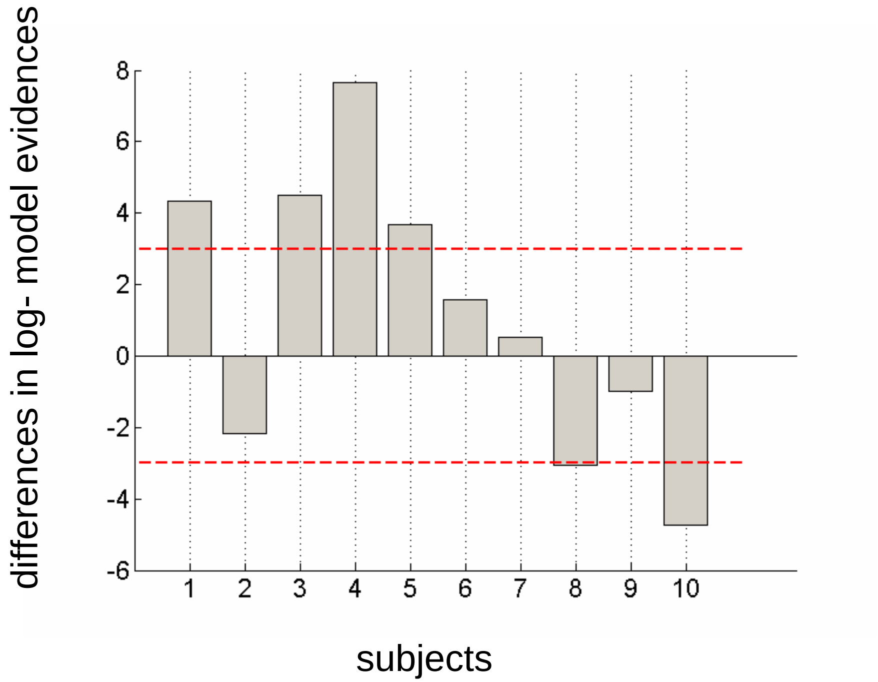


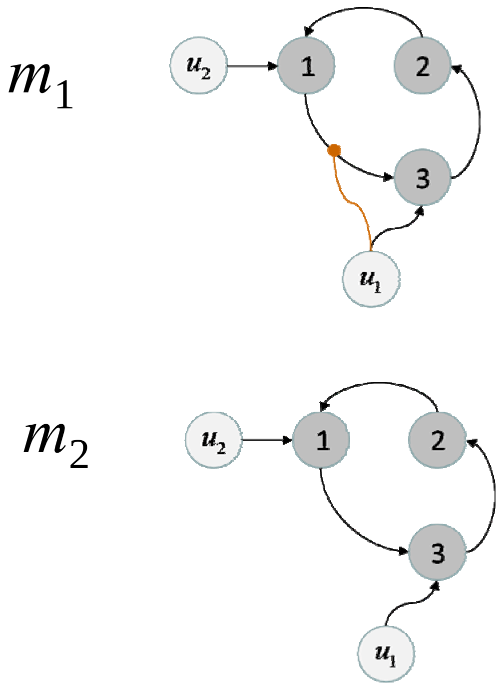
[Stephan et al., Neuroimage, 2008]

Dynamic causal modelling

model comparison for group studies

$$\ln p(y|m_1) - \ln p(y|m_2)$$





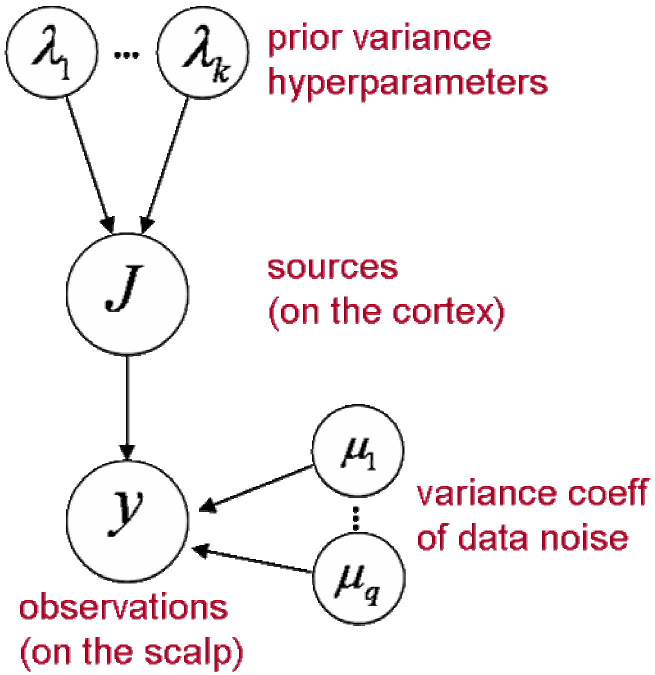
fixed effect — assume all subjects correspond to the same model

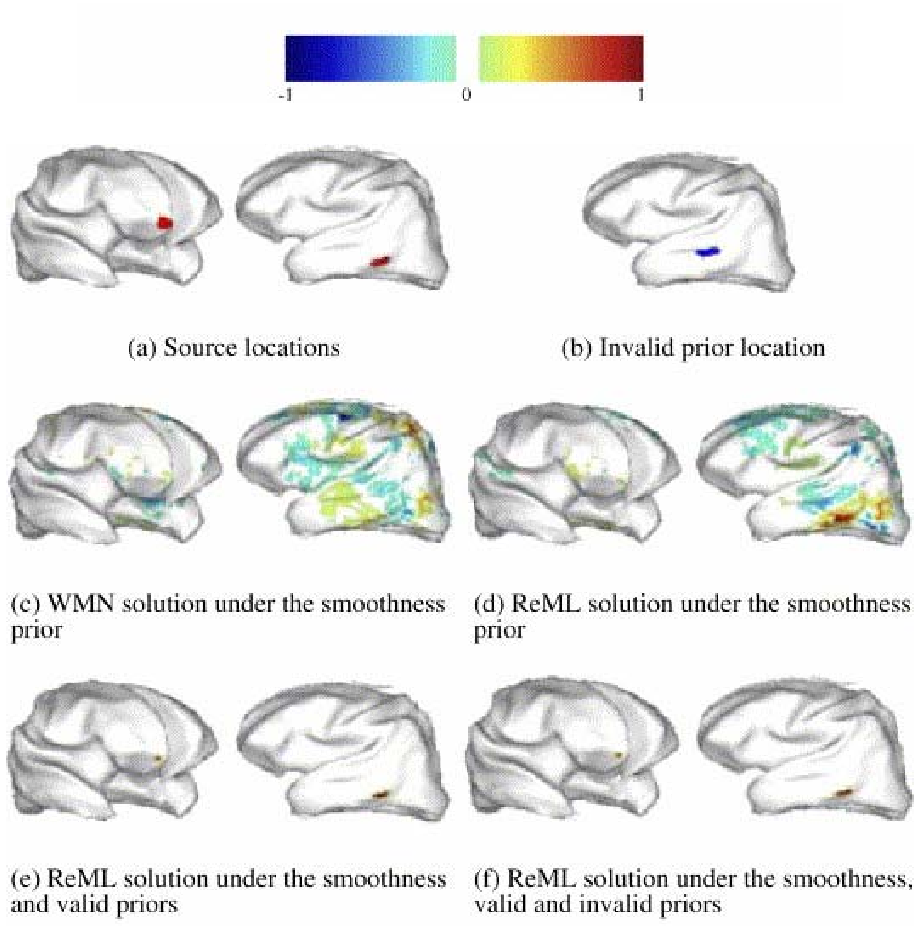
random effect — assume different subjects might correspond to different models

[Stephan et al., Neuroimage, 2009]

EEG source reconstruction

finessing an ill-posed inverse problem



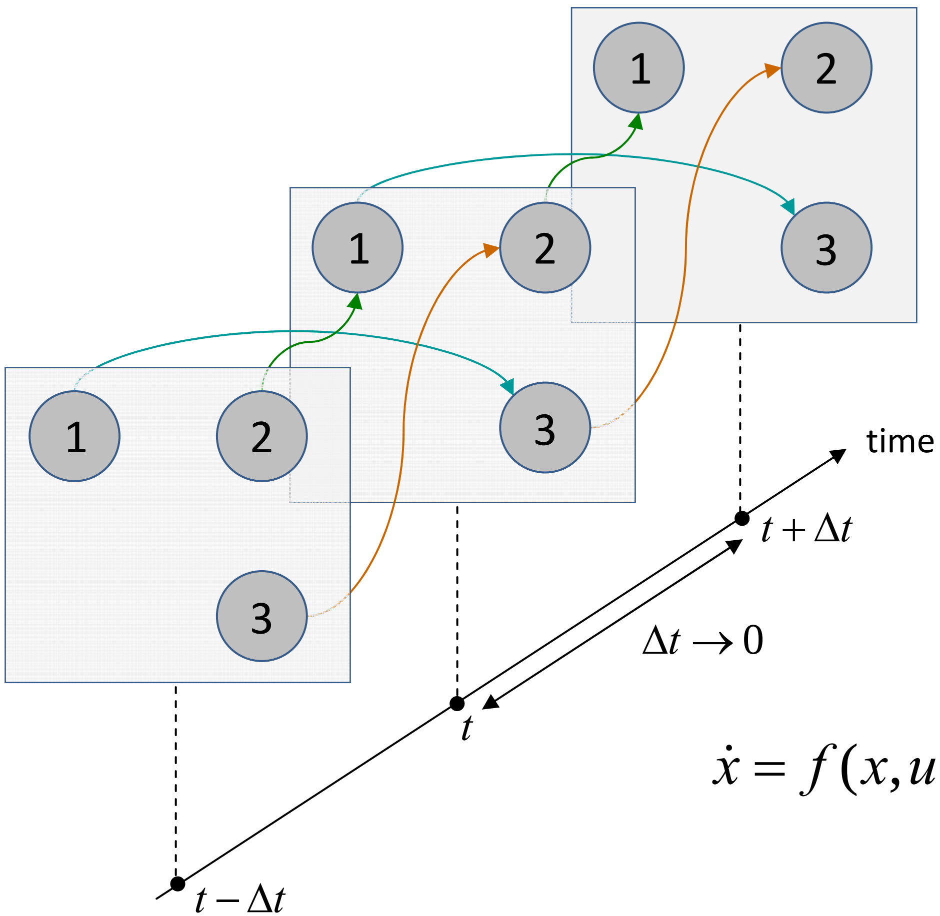


[Mattout et al., Neuroimage, 2006]

I thank you for your attention.

DCMs and DAGs

a note on causality



$$\dot{x} = f(x,u,\theta)$$