Bayesian Model Selection and Averaging

SPM for MEG/EEG Course

Ulrich Stoof
Contents

Introduction and model inversion
Dynamic Causal Modelling (DCM)

Comparing models
Bayesian Model Comparison and Selection (BMS)

Rapidly evaluating models
Bayesian Model Reduction (BMR)

Investigating the parameters
Bayesian Model Averaging (BMA)

Multi-subject analysis
Parametric Empirical Bayes (PEB)
Contents

- Introduction and model inversion
  - Dynamic Causal Modelling (DCM)

- Comparing models
  - Bayesian Model Comparison and Selection (BMS)

- Rapidly evaluating models
  - Bayesian Model Reduction (BMR)

- Investigating the parameters
  - Bayesian Model Averaging (BMA)

- Multi-subject analysis
  - Parametric Empirical Bayes (PEB)
DCM for ERP – Examples in this Presentation

Dynamic Causal Modelling

Physiological Models
- Event-related potentials (ERP)
- Cross-spectral densities (CSD)

Phenomenological Models
- fMRI
- Induced responses (IND)
- Phase coupling (PHA)

Adapted from Bernadette van Wijk, SPM for MEG/EEG Course, Principles of Dynamic Causal Modelling
Mismatch Negativity (MMN) and Roving Paradigm

Design and responses elicited in a roving paradigm
Adapted from Garrido et al. (2008), Figure 1
Garrido et al. 2008, doi.org/10.1016/j.neuroimage.2008.05.018

Model specification in a MMN paradigm
Copied from Garrido et al. (2007), Figure 1
Garrido et al. 2007, doi.org/10.1016/j.neuroimage.2007.03.014

- MMN is an event-related potential (ERP) component evoked by detectable violations in acoustic regularity
- Roving paradigms are characterised by sporadic frequency changes of a repeating tone

A1 - left and right primary auditory cortex
STG - left and right superior temporal gyrus
IFG - right inferior frontal gyrus
Bayesian Framework, Forward and Inverse Problems

$p(Y|\theta, m)$
Likelihood

Forward Problem

$p(Y|m)$
Evidence

Inverse Problem

$p(\theta|Y, m)$
Posterior

DCM rests on estimating certain conditional probabilities

Garrido et al. (2008), doi.org/10.1016/j.neuroimage.2008.05.018
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
DCM Structure, Symbols used in this Presentation

- Model $m$
- Prior probability $p(\theta|m)$
- Data $Y$
- Posterior probability $p(\theta|Y,m)$
- Free energy $F \approx \log p(Y|m)$
- Inverted model

DCM analysis provides a score for model likelihood and posterior parameter estimates

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
Priors determine Model Structure and Solutions

Priors restrict parameters to a specific search space to achieve realistic (interpretable) solutions.
## Contents

- Introduction and model inversion
  - Dynamic Causal Modelling (DCM)

- Comparing models
  - Bayesian Model Comparison and Selection (BMS)

- Rapidly evaluating models
  - Bayesian Model Reduction (BMR)

- Investigating the parameters
  - Bayesian Model Averaging (BMA)

- Multi-subject analysis
  - Parametric Empirical Bayes (PEB)
Inverting Models individually and Free Energy Scoring

The functional anatomy of the MMN: A DCM study of the roving paradigm

Marta I. Garrido a,c,*, Karl J. Friston a, Stefan J. Kiebel a, Klaas E. Stephan a, Torsten Baldeuweg b, James M. Kilner a

a Welcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, UK
b Developmental Cognitive Neuroscience, Institute of Child Health, University College London, UK
c Department of Psychology, University of California, Los Angeles, USA

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Garrido et al. (2008), doi.org/10.1016/j.neuroimage.2008.05.018

Models can be scored against each use using free energy
Model Comparison using log Bayes Factor

Bayes factor

\[ B_{ij} = \frac{p(Y|m = i)}{p(Y|m = j)} \]

Free energy approximates log model evidence

\[ F \approx \log p(Y|m) \]

Log Bayes factor is approximately the differences of free energies

\[ \log B_{ij} = \log p(Y|m = i) - \log p(Y|m = j) \approx F_i - F_j \]

Bayes factor can be interpreted as evidence for a model / hypothesis, e.g., \( \log B > 3 \) suggests strong evidence and a posterior probability of 95%

Interpretation of Bayes factors

| \( B_{ij} \) | \( p(m = i|y) \) (%) | Evidence in favor of model \( i \) |
|------------|----------------------|-----------------------------|
| 1–3        | 50–75                | Weak                        |
| 3–20       | 75–95                | Positive                    |
| 20–150     | 95–99                | Strong                      |
| ≥150       | ≥99                  | Very strong                 |

Bayes factors can be interpreted as follows. Given candidate hypotheses \( i \) and \( j \), a Bayes factor of 20 corresponds to a belief of 95% in the statement ‘hypothesis \( i \) is true’. This corresponds to strong evidence in favor of \( i \).

Copied from Raftery et al. (1995)


Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
Evaluating large Model Spaces

Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions

Phillips et al. (2015), doi.org/10.1523/JNEUROSCI.5095-14.2015

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
Bayesian Model Selection based on Posterior Probabilities

Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions

Holly N. Phillips, Alejandro Blenkmann, Laura E. Hughes, Tristan A. Bekinschtein, and James B. Rowe

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Bayesian Model Selection based on Posterior Probabilities

Transformation of Bayes factors into posterior probabilities using Bayes rule

$$p(m = i | Y) = \frac{p(Y | m = i)}{p(Y)}$$

$$= \frac{1}{1 + \exp(-\log(B_{ij})}$$

Phillips et al. (2015), doi.org/10.1523/JNEUROSCI.5095-14.2015
Avoiding Evidence Dilution / Structuring Model Space

\[ p(f|Y) = \sum_{m \in S_f} p(m|Y) \]

Structuring the model space by grouping models into families helps to avoid evidence dilution.

Adapted from Will Penny, SPM for fMRI Course, DCM Advanced - Part I: Model Selection
Contents

- Introduction and model inversion
  Dynamic Causal Modelling (DCM)

- Comparing models
  Bayesian Model Comparison and Selection (BMS)

- Rapidly evaluating models
  Bayesian Model Reduction (BMR)

- Investigating the parameters
  Bayesian Model Averaging (BMA)

- Multi-subject analysis
  Parametric Empirical Bayes (PEB)
Evaluating complex, multi-level Model Spaces

Example of complex model space with 64 models: 8 between and 8 within regional variations

Fitzgerald et al. (2019), doi.org/10.1101/768846
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
Bayesian Model Reduction (BMR) Procedure

Individual model inversions

\[ m_1 \rightarrow \cdots \rightarrow m_{64} \]

Bayesian model reduction

\[ m_1 \rightarrow \cdots \rightarrow m_{64} \]

BMR allows estimation of parameters and free energy from a single inverted (full) model.
Theoretical Basis for BMR

These (approximate) equalities mean one can evaluate the posterior and evidence of any reduced model, given the posteriors of the full model. In other words, 

\[ F[\tilde{P}(\theta) : P(\theta)] \approx \ln \tilde{P}(y) \]

allows us to skip the optimization of the reduced posterior \( \tilde{Q}(\theta) \) and use the optimized posterior of the full model to compute the evidence (and posterior) of the reduced model directly.

BMR requires an inverted ‘full model’ and a set of structurally identical ‘reduced models’

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
BMR enables Exploration of entire Model Spaces

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
Contents

- Introduction and model inversion
  Dynamic Causal Modelling (DCM)

- Comparing models
  Bayesian Model Comparison and Selection (BMS)

- Rapidly evaluating models
  Bayesian Model Reduction (BMR)

- Investigating the parameters
  Bayesian Model Averaging (BMA)

- Multi-subject analysis
  Parametric Empirical Bayes (PEB)
3.2. Model comparison

As anticipated (Garrido et al., 2008) in healthy controls the Combination model outperformed both the Forward and Backward models with exceedance probability of 89%. The Combination model assumes that the MMN response emerged from changes in all bidirectional extrinsic as well as in intrinsic connections. In patients the optimal model included modulation via the MMN of the intrinsic connections and only the forward connections, which was the Forward model. The exceedance probability for this model was 44%, surpassing the exceedance probabilities of the other two tested models.

Dima et al. (2012), doi.org/10.1016/j.schres.2011.12.024

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging

Bayesian Model Averaging (BMA)

\[ p(\theta | Y) = \sum_{m} p(\theta | m, Y) \cdot p(m | Y) \]

- Parameter posterior prob.
- Parameter probability
- Model probability

Controls

Patients

Dima et al. (2012), doi.org/10.1016/j.schres.2011.12.024

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
Contents

- Introduction and model inversion
- Dynamic Causal Modelling (DCM)
- Comparing models
- Bayesian Model Comparison and Selection (BMS)
- Rapidly evaluating models
- Bayesian Model Reduction (BMR)
- Investigating the parameters
- Bayesian Model Averaging (BMA)
- Multi-subject analysis
- Parametric Empirical Bayes (PEB)
Many scientific questions centre on group comparisons in connectivity parameters.

Rosch et al. (2019), doi.org/10.1016/j.bpsc.2018.07.003, GitHub: Ketamine DCM
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
First and Second Level Modelling

\[ \theta^{(2)} = \eta + \varepsilon^{(3)} \]

Priors on second level parameters

Second level

\[ \theta^{(1)} = \Gamma^{(2)}(\theta^{(2)}) + \varepsilon^{(2)} \]

Between-subject error

Second level (linear) model

First level

\[ y = \Gamma_i^{(1)}(\theta^{(1)}) + \varepsilon^{(1)} \]

Measurement noise

DCM for subject \( i \)

PEB is a hierarchical modelling approach incorporating first and second level

Adapted from Richard Rosch and Peter Zeidman, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
Modelling Steps: DCM to reduced PEB

\[ m_1 \quad m_2 \quad m_3 \quad m_4 \quad m_5 \]

\[ s_1 \quad s_2 \quad s_3 \quad s_4 \quad s_5 \quad s_6 \]

Modelling first and second level within a Bayesian context using SPM procedures and functions

Rosch et al. (2019), doi.org/10.1016/j.bpsc.2018.07.003, GitHub: Ketamine_DCM

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
PEB Example: Effect of ketamine on (Intrinsic) Connectivity

1st Level
- DCM

2nd Level
- PEB Design Matrix
- Model Comparison
- Model Average

Rosch et al. (2019), doi.org/10.1016/j.bpsc.2018.07.003, GitHub: Ketamine_DCM
Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
PEB Advantages and Applications

- Conveys uncertainty about parameters from the subject level to the group level
- Can improve first level parameters estimates
- Can be used to ...
  - ... compare specific reduced PEB models (switching off combinations of group-level parameters)
  - ... or to search over nested models (BMR)
- Prediction (leave-one-out cross validation)
References and additional Material
Additional Resources

Will Penny’s advanced DCM lecture slides
Penny: DCM advanced, SPM Course Slides

Lecture by Stefan Frässle on Bayesian model selection and averaging
Fraessle: BMS and BMA

Tutorial for PEB by Peter Zeidman
Zeidman: DCM-PEB Example

PEB Paper (Friston et al., 2015)
Bayesian model reduction and empirical Bayes for group (DCM)

10 Simple Rules for Group studies before PEB (Stephan et al., 2010)
Ten simple rules for dynamic causal modeling

Worked example using PEB with code by Natalie Adams
Adams: PEB Example

Adapted from Richard Rosch, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging
GLM of Connectivity Parameters

\[ \theta^{(1)} = X \theta^{(2)} + \epsilon^{(2)} \]

- **Unexplained between-subject variability**
- **Design matrix (covariates)**
- **Group level parameters**

\[ \theta^{(1)} = \begin{bmatrix} \theta_{1}^{(1)} \\ \theta_{2}^{(1)} \\ \vdots \\ \theta_{6}^{(1)} \end{bmatrix} \]

\[ X = \begin{bmatrix} X_{1} \\ X_{2} \\ \vdots \\ X_{6} \end{bmatrix} \]

\[ \theta^{(2)} = \begin{bmatrix} \theta_{1}^{(2)} \\ \theta_{2}^{(2)} \\ \vdots \\ \theta_{6}^{(2)} \end{bmatrix} \]

- **Between-subjects effects**
- **Group average connection strength**
- **Effect of group on the connection**
- **Effect of age on the connection**

Adapted from Peter Zeidman, SPM for MEG/EEG Course, Bayesian Model Selection and Averaging