M/EEG Source Reconstruction

Saskia Helbling

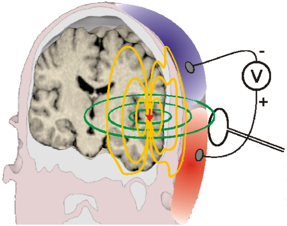
With many thanks to Rik Henson and all the others who contributed to the slides

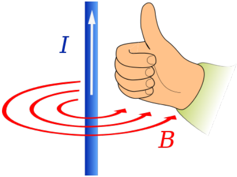
M/EEG SPM course, May 12, 2015

Outline

Primary intracellular currents give rise to volume currents and a magnetic field

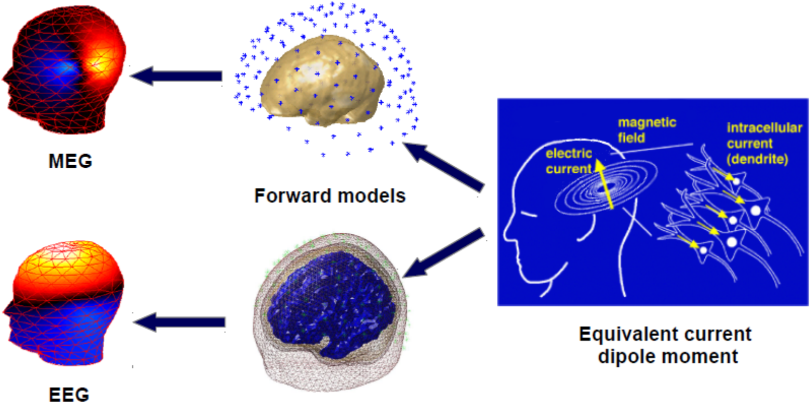
- Volume currents yield potential differences on the scalp that can be measured by EEG
- MEG measures the changes in the magnetic field generated by an electric current (Sarvas 1987, Hämäläinen 1993)
- These magnetic fields are mainly induced by primary currents based on excitatory activity (Okada et al. 1997)





Baillet: MEG consortium

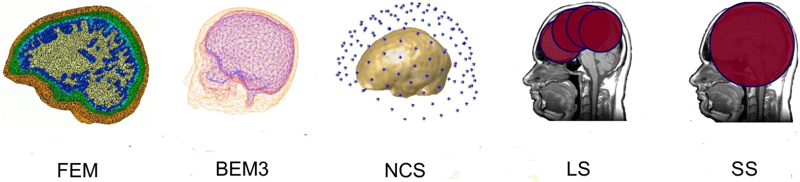
Forward models predict the M/EEG surface signals to current dipoles in the brain



Forward models can be described by leadfield matrices

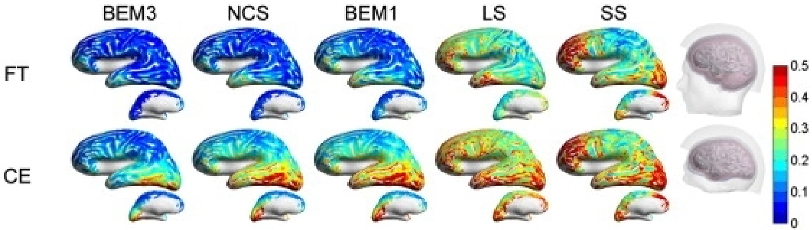
- Sources J are mapped to channels by subject-specific leadfield matrices L: $Y = LJ + \varepsilon$, with data Y, with noise ε
- Leadfield matrix L depends on:
  - the type/location/orientation of sensors
  - the geometry of the head
  - the conductivity of head tissues (in particular for EEG)
  - the source space we are looking at (e.g. cortical surface or volumetric image)

Headmodels show different degrees of complexity



- The simpler models are not sufficient to predict the electric potential differences at the scalp
- Complex models are (1) computationally more expensive and (2) require more prior knowledge about the anatomy and conductivity values and might be prone to approximation errors

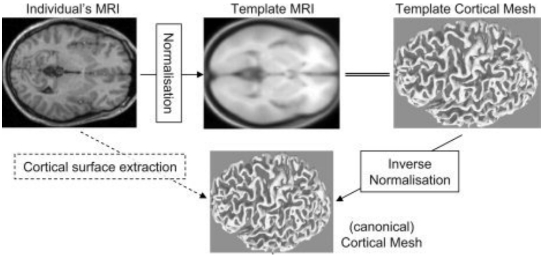
MEG also may benefit from using more complex headmodels



Stenroos, Neuroimage 2014

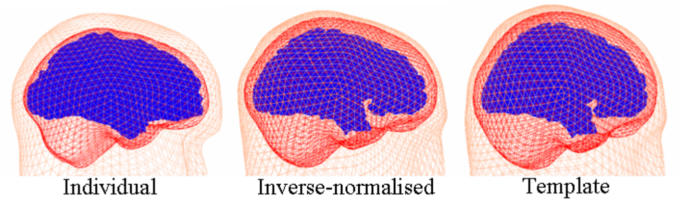
Canonical meshes

- Extracting the headmodel surfaces from MRI can be prone to approximation errors
- The cortical surface in particular is difficult due to the convoluted nature of the brain (but: Freesurfer)
- Rather than extract surfaces from individuals MRIs, we can warp template surfaces from an MNI brain based on spatial (inverse) normalisation



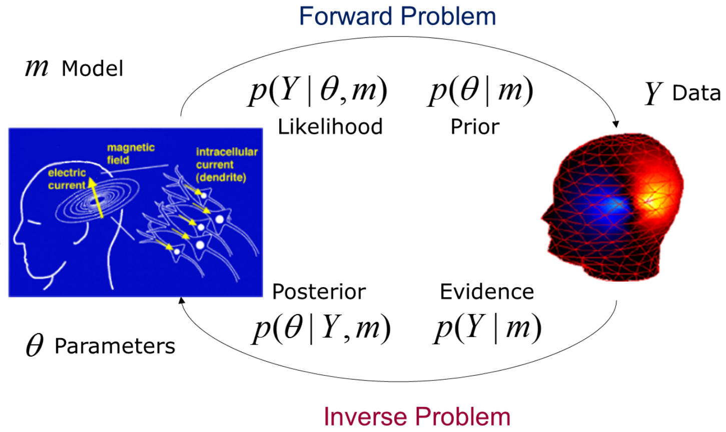
Canonical meshes II

- By using canonical meshes we have a one-to-one mapping across subjects which can be used for group statistics and group-inversion schemes
- Allows for multi-model data integration as source solutions live in MNI space



Mattout et al. 2007

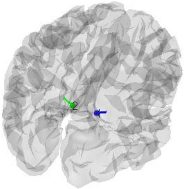
The forward and inverse problem from a Bayesian perspective



1. Forward Models for M/EEG
2. Variational Bayesian Dipole Estimation (ECD)
3. Distributed Parametric Empirical Bayes Estimation
4. Multi-modal and multi-subject integration

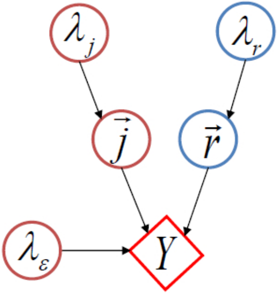
Equivalent Current Dipole Solutions for a small number of cortical current sources

- For small number of Equivalent Current Dipoles (ECD) the inverse problem is linear in the orientation of the sources, but non-linear in location: $Y = L(r)j + \varepsilon$
- Standard ECD approaches iterate location/orientation (within a brain volume) until fit to sensor data is maximised (i.e, error minimised)
- But there remains the issue of local minima and the question of how many dipoles we should use?
- And how can we incorporate prior knowledge?



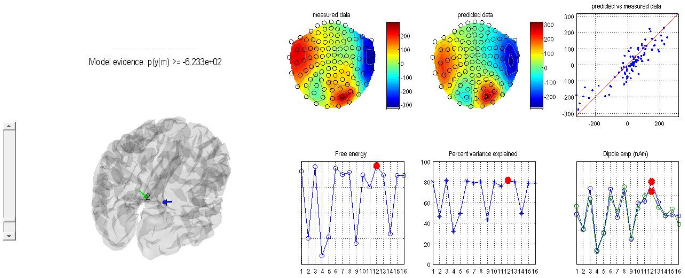
Generative model for the Variational Bayesian ECD approach

- Forward model for a few dipolar sources: $Y = L(r)j + \varepsilon$
- The locations r and moments j are drawn from normal distributions with precisions γ_s and γ_w
- ε is white observation noise with precision γ_ε
- These precisions in turn are drawn from a prior gamma distribution
- We assume that the probabilities factorise: $p\left(Y, r, j, \lambda_r, \lambda_j, \lambda_\varepsilon \mid m\right) = p\left(Y \mid r, j, \lambda_r, \lambda_\varepsilon\right) p\left(\lambda_\varepsilon \mid m\right) p\left(r \mid \lambda_r, m\right) p\left(\lambda_r \mid m\right) p\left(j \mid \lambda_j, m\right) p\left(\lambda_j \mid m\right)$



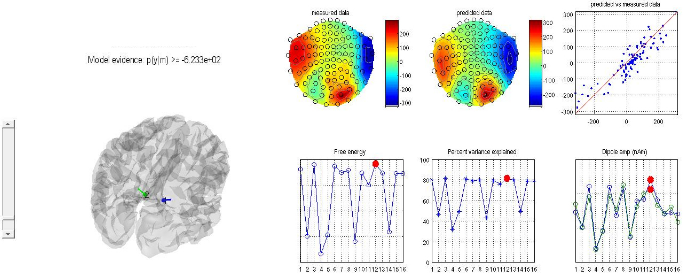
Use a Variational Bayesian scheme with the Free Energy as a cost function

- Use a multi-start procedure to avoid being stucked in local minima
- Compare evidence for models with different number of dipoles or with different priors (Kiebel et al., 2008)



Use a Variational Bayesian scheme with the Free Energy as a cost function

- Use a multi-start procedure to avoid being stucked in local minima
- Compare evidence for models with different number of dipoles or with different priors (Kiebel et al., 2008)



A linear forward model for distributed M/EEG source analysis

- Given p sources fixed in location (e.g, on a cortical mesh) and orientation, the forward model turns linear: $\mathbf{Y} = \mathbf{LJ} + \mathbf{E}$ with $\mathbf{E} \sim N\left(\mathbf{0}, \mathbf{C}^{(e)}\right)$
- But: there are many more possible sources than sensor data. We thus need some form of regularisation to solve this under-determined problem

The classical L2 or weighted minimum norm approach

- $\mathbf{Y} = \mathbf{LJ} + \varepsilon$ with $\varepsilon \sim N\left(\mathbf{0}, \mathbf{C}^{(e)}\right)$
- $\mathbf{J} = argmin\left\{ \|\mathbf{C}_e^{-1/2}(\mathbf{Y} - \mathbf{LJ})\|^2 + \lambda \|\mathbf{WJ}\|^2 \right\}$
- Tikhonov solution: $\mathbf{J} = \left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{L}^T\left[\mathbf{L}\left(\mathbf{W}^T\mathbf{W}\right)^{-1}\mathbf{L}^T + \lambda\mathbf{C}_\varepsilon\right]^{-1}\mathbf{Y}$

Weighting matrices

$W = I$	Minimum Norm
$W = DD^T$	Loreta ($D =$Laplacian)
$W = diag\left(L^TL\right)^{-1}$	Depth-weighted
$W_p = diag\left(L_p^T C_y^{-1} L_p\right)^{-1}$	Beamformer

Philipps et al., 2002

And its Parametric Empirical Bayes (PEB) generalisation

Hierarchical linear model

$Y = LJ + E_e$ with $\mathbf{E_e} \sim N(\mathbf{0}, \mathbf{C_e})$ and $C_e = n \times n$ sensor error covariance matrix
$J = 0 + E_j$ with $\mathbf{E_j} \sim N(\mathbf{0}, \mathbf{C_j})$ and $C_j = p \times p$ source prior covariance matrix

Bayesian terms

Likelihood: $p(Y \mid J) = N(LJ, C_e)$
Prior: $p(J) = N(0, C_j)$
Posterior: $p(J \mid Y) \propto p(Y \mid J)\, p(J)$

Maximum A Posteriori estimate

MAP estimate: $\hat{J} = C_j L^T \left[L C_j L^T + C_e\right]^{-1} Y$
For $C_j = \left(W^T W\right)^{-1}$, this corresponds to the classical weighted minimum norm solution: $\hat{J} = (W^T W) L^T \left[L (W^T W) L^T + \lambda C_e\right]^{-1} Y$

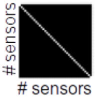
See Phillips et al (2005), Neuroimage

Sensor and source level covariance components

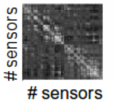
Covariance priors C are specified as the sum of sensor and source components Q_i, weighted by hyperparameters λ_i: $C = \sum_i \lambda_i Q_i$

Sensor components, $Q_i^{(e)}$ (error):

"IID" (White noise)

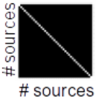


Empty room:

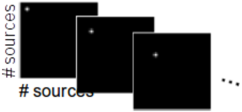


Source components, $Q_i^{(j)}$:

"IID" (Minimum Norm):

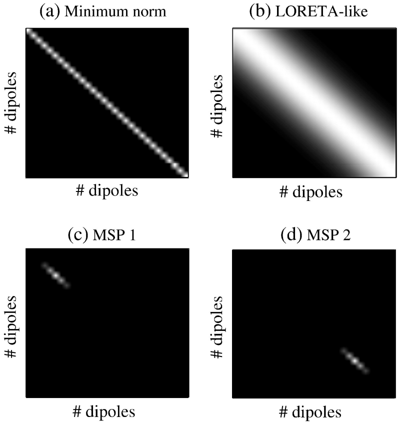


Multiple Sparse Priors (MSP):

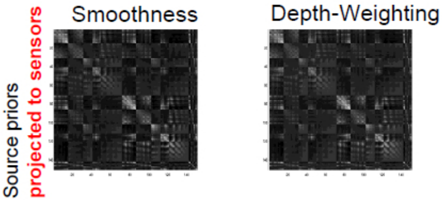


Friston et al., 2008

Source covariance matrices



Hyperpriors



- When multiple Q's are correlated, estimation of hyperparameters λ can be difficult (e.g. local maxima), and they can become negative (improper for covariances)
    - impose positivity on hyperparameters: $\alpha_i = ln(\lambda_i) \Longleftrightarrow \lambda_i = exp(\alpha_i)$
    - impose shrinkage hyperpriors: $p(\alpha) \sim N(\eta, \Omega)$ with $\eta = -4$ and $\Omega = aI$, $a = 16$
- Uniformative priors are 'turned off': $\alpha_i \to -\infty \Longleftrightarrow \lambda_i \to 0$

Negative Free energy provides a cost function for the hyperparameters

- Obtain Restricted Maximum Likelihood (ReML) estimates of the hyperparameters λ by maximising the variational (negative) Free energy F: $\hat{\lambda} = \max_{\lambda} p(Y \mid \lambda) = \max_{\lambda} F$
- A final expectation step gives us Maximum A Posteriori (MAP) estimates of source parameters J: $\hat{J} = \max_{\lambda} p\left(J \mid Y, \hat{\lambda}\right) = \max_{\lambda} F$
- Maximal Free energy F approximates the Bayesian log model evidence for a model m, and in doing so takes accuracy and complexity of the model into account
- Note: a large set of priors does not necessarily mean that the model has a large complexity (as the according hyperparameters may have eliminated the respective components)

Negative Free energy provides a cost function for the hyperparameters

- Obtain Restricted Maximum Likelihood (ReML) estimates of the hyperparameters λ by maximising the variational (negative) Free energy F: $\hat{\lambda} = \max_{\lambda} p(Y \mid \lambda) = \max_{\lambda} F$
- A final expectation step gives us Maximum A Posteriori (MAP) estimates of source parameters J: $\hat{J} = \max_{\lambda} p\left(J \mid Y, \hat{\lambda}\right) = \max_{\lambda} F$
- Maximal Free energy F approximates the Bayesian log model evidence for a model m, and in doing so takes accuracy and complexity of the model into account
- Note: a large set of priors does not necessarily mean that the model has a large complexity (as the according hyperparameters may have eliminated the respective components)

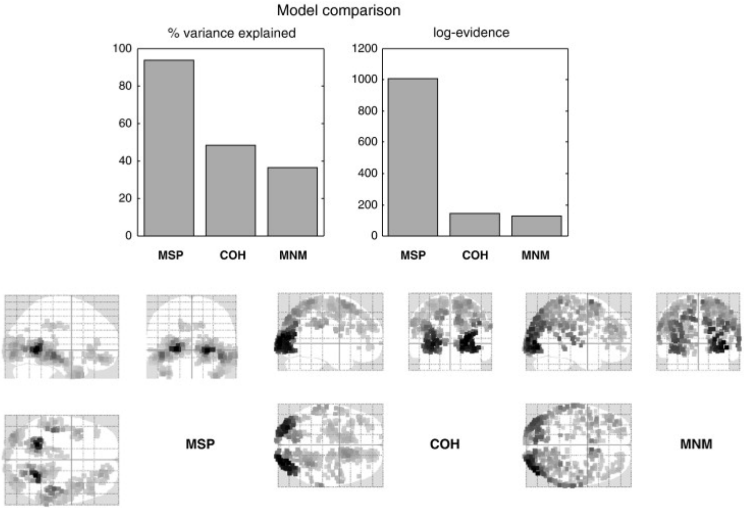
Negative Free energy provides a cost function for the hyperparameters

- Obtain Restricted Maximum Likelihood (ReML) estimates of the hyperparameters λ by maximising the variational (negative) Free energy F: $\hat{\lambda} = \max_{\lambda} p(Y \mid \lambda) = \max_{\lambda} F$
- A final expectation step gives us Maximum A Posteriori (MAP) estimates of source parameters J: $\hat{J} = \max_{\lambda} p\left(J \mid Y, \hat{\lambda}\right) = \max_{\lambda} F$
- Maximal Free energy F approximates the Bayesian log model evidence for a model m, and in doing so takes accuracy and complexity of the model into account
- Note: a large set of priors does not necessarily mean that the model has a large complexity (as the according hyperparameters may have eliminated the respective components)

Negative Free energy provides a cost function for the hyperparameters

- Obtain Restricted Maximum Likelihood (ReML) estimates of the hyperparameters λ by maximising the variational (negative) Free energy F: $\hat{\lambda} = \max_{\lambda} p(Y \mid \lambda) = \max_{\lambda} F$
- A final expectation step gives us Maximum A Posteriori (MAP) estimates of source parameters J: $\hat{J} = \max_{\lambda} p\left(J \mid Y, \hat{\lambda}\right) = \max_{\lambda} F$
- Maximal Free energy F approximates the Bayesian log model evidence for a model m, and in doing so takes accuracy and complexity of the model into account
- Note: a large set of priors does not necessarily mean that the model has a large complexity (as the according hyperparameters may have eliminated the respective components)

Multiple Sparse Priors



Reducing the computational load

- The computation of the Maximum Free Energy is made computationally feasable by using the sensor rather than the source level covariance matrix:
  $\Sigma = \lambda_0 Q_\varepsilon + \sum_{i=1}^{N_q} \lambda_i L C^{-1} L^T$
- Project data to spatial and temporal subspaces by means of Singular Value Decomposition
  $\Rightarrow$ decreases computational cost of the gradient ascent
  $\Rightarrow$ increases the signal to noise ratio by removing redundancy in the data and reducing noise
- Reduce number of hyperparameters to be estimated in each step by applying a heuristic optimisation step

Reducing the computational load

- The computation of the Maximum Free Energy is made computationally feasable by using the sensor rather than the source level covariance matrix:
  $\Sigma = \lambda_0 Q_\varepsilon + \sum_{i=1}^{N_q} \lambda_i L C^{-1} L^T$
- Project data to spatial and temporal subspaces by means of Singular Value Decomposition
  $\Rightarrow$ decreases computational cost of the gradient ascent
  $\Rightarrow$ increases the signal to noise ratio by removing redundancy in the data and reducing noise
- Reduce number of hyperparameters to be estimated in each step by applying a heuristic optimisation step

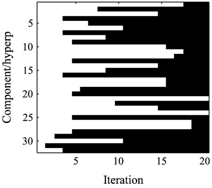
Reducing the computational load

- The computation of the Maximum Free Energy is made computationally feasable by using the sensor rather than the source level covariance matrix:
  $\Sigma = \lambda_0 Q_\varepsilon + \sum_{i=1}^{N_q} \lambda_i L C^{-1} L^T$
- Project data to spatial and temporal subspaces by means of Singular Value Decomposition
  $\Rightarrow$ decreases computational cost of the gradient ascent
  $\Rightarrow$ increases the signal to noise ratio by removing redundancy in the data and reducing noise
- Reduce number of hyperparameters to be estimated in each step by applying a heuristic optimisation step

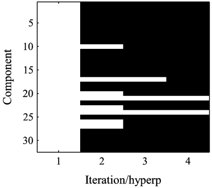
Automatic relevance detection and Greedy Search

- ARD: each covariance component has an associated hyperparameter; components are pruned during the optimisation scheme
- GS: performs a single to many optimisation of hyperparameters and splits sets of hyperparameters recursively
- In SPM you can either use one of these heuristics alone or apply a final optimisation step on the ARD, GS and error covariance matrices to avoid local minima

(a) ARD optimisation of hyperparameters



(b) GS optimisation of hyperparameters



PEB Summary

- Allows for multiple priors in the form of covariance components to the extent that hundreds of sparse priors (MSP), multiple error components or multiple fMRI priors can be incorporated
- The parametric empirical Bayes approach automatically "regularises" the inverse problem in a principled fashion
- Furnishes estimates of model evidence which allow us to evaluate the different priors: on the level of sensor or source covariances, but also for the forward models (Henson et al. 2009, Lopez et al., 2012)

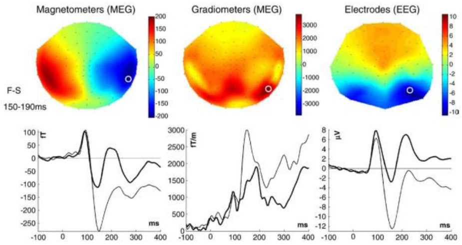
PEB Summary

- Allows for multiple priors in the form of covariance components to the extent that hundreds of sparse priors (MSP), multiple error components or multiple fMRI priors can be incorporated
- The parametric empirical Bayes approach automatically “regularises” the inverse problem in a principled fashion
- Furnishes estimates of model evidence which allow us to evaluate the different priors: on the level of sensor or source covariances, but also for the forward models (Henson et al. 2009, Lopez et al., 2012)

PEB Summary

- Allows for multiple priors in the form of covariance components to the extent that hundreds of sparse priors (MSP), multiple error components or multiple fMRI priors can be incorporated
- The parametric empirical Bayes approach automatically "regularises" the inverse problem in a principled fashion
- Furnishes estimates of model evidence which allow us to evaluate the different priors: on the level of sensor or source covariances, but also for the forward models (Henson et al. 2009, Lopez et al., 2012)

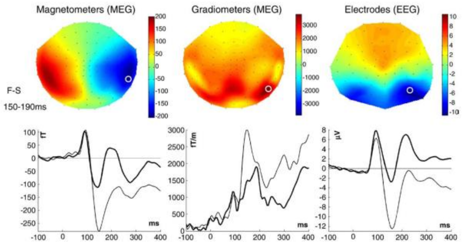
MEG and EEG data fusion



Henson et al., 2009

- Different MEG sensors and EEG are sensitive to different source configurations and hence can provide (partly) independent information on the underlying sources
- Rescale concatenated data and leadfield matrices to accommodate different scaling and measurement units across the different sensor-types
- We apply the same source covariance priors across all modalities, but sensor covariance priors are assumed to differ

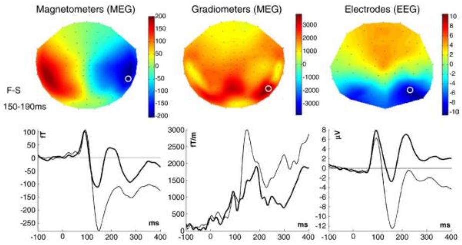
MEG and EEG data fusion



Henson et al., 2009

- Different MEG sensors and EEG are sensitive to different source configurations and hence can provide (partly) independent information on the underlying sources
- Rescale concatenated data and leadfield matrices to accommodate different scaling and measurement units across the different sensor-types
- We apply the same source covariance priors across all modalities, but sensor covariance priors are assumed to differ

MEG and EEG data fusion

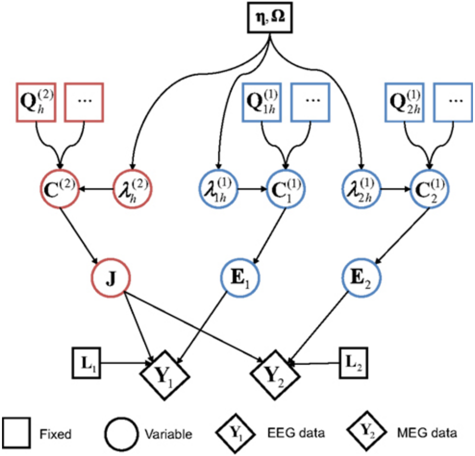


Henson et al., 2009

- Different MEG sensors and EEG are sensitive to different source configurations and hence can provide (partly) independent information on the underlying sources
- Rescale concatenated data and leadfield matrices to accommodate different scaling and measurement units across the different sensor-types
- We apply the same source covariance priors across all modalities, but sensor covariance priors are assumed to differ

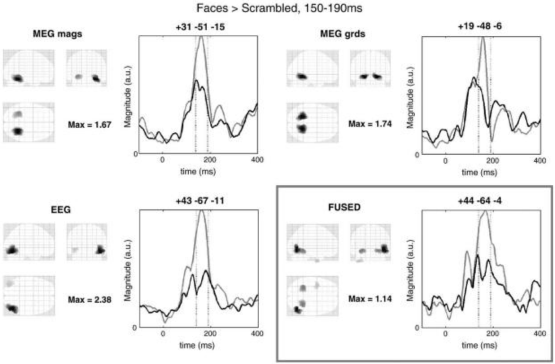
MEG and EEG data fusion

Symmetric multimodal (E+MEG) fusion model



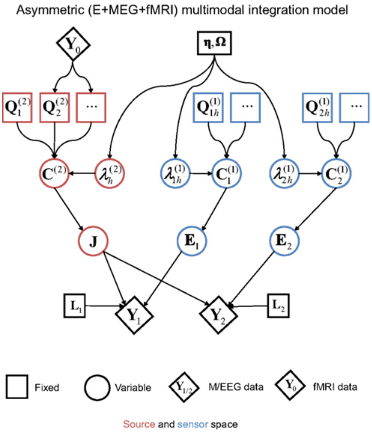
Source and sensor space

MEG and EEG data fusion



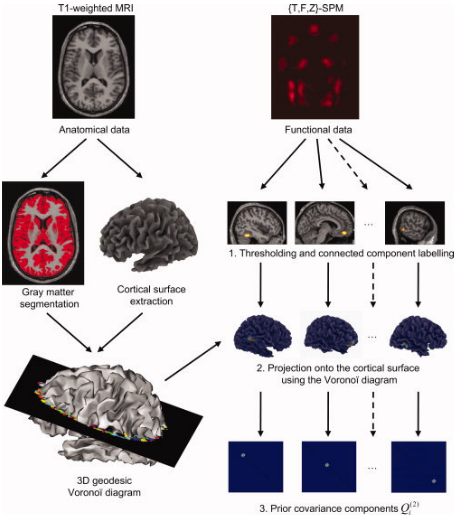
- The maximal sources recovered from fusion were a plausible combination of the ventral temporal sources recovered by MEG and the lateral temporal sources recovered by EEG (Henson et al., 2009, Neuroimage)

Using fMRI priors in M/EEG source reconstruction

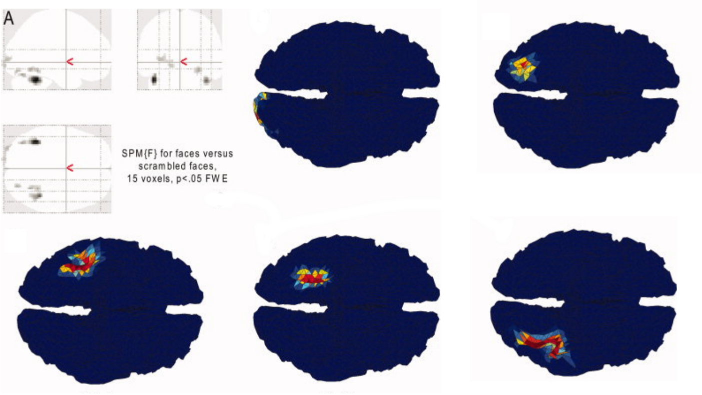


Henson et al. (2011) Frontiers in Human Neurosci

Convert fMRI clusters into covariance components

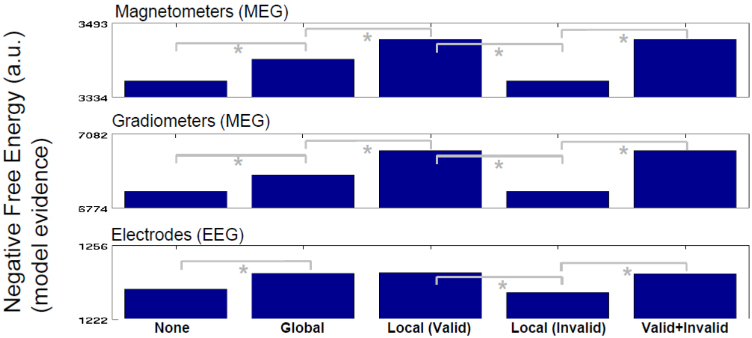


Using fMRI priors from a face perception task



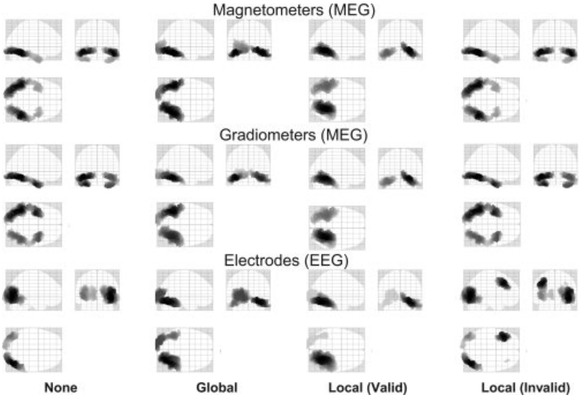
Henson et al., 2010, HBM

Valid fMRI priors increase the log model evidence

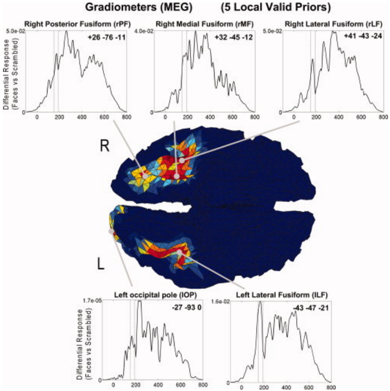


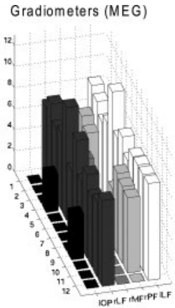
From Henson et al., 2010

fMRI priors counteract superficial bias of L2-norm solutions



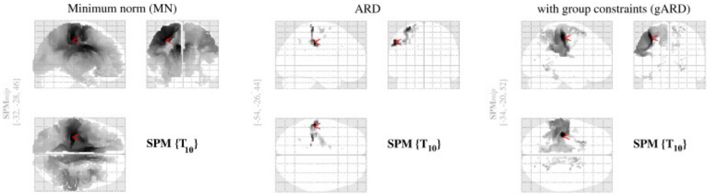
fMRI priors affect the variance but not the exact time course





Bayesian group inversion

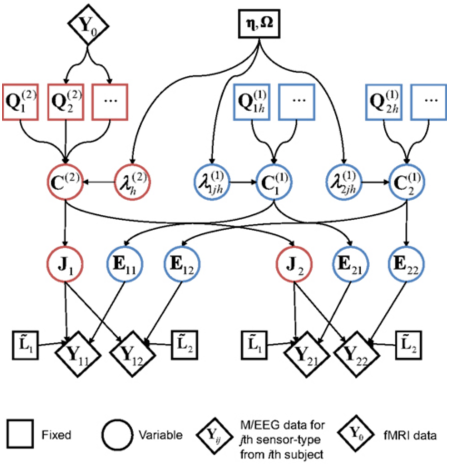
- Concatenate data across subjects and re-aligned leadfields to an average leadfield
- Common source priors $C^{(j)} = \sum \lambda_k^{(j)} Q_k^{(j)}$, but subject specific sensor level priors $C_i^{(e)} = \sum \lambda_{ik}^{(e)} A_i Q_k^{(e)} A_i^T$, with alignment matrices A_i
- Group increases the detection of differences at the group or between subject level without applying additional smoothing



Litvak & Friston (2008) Neuroimage

Bayesian group inversion

Multisubject fusion model



Source and sensor space

Summary

- SPM offers a range of standard forward models (via FieldTrip) and supports canonical headmodels
- Offers unique Bayesian approaches to inversion and allows us to compare different sets of prior assumptions by using the log model evidence
  - Variational Bayesian ECD
  - A PEB approach to distributed imaging (e.g. MSP, MNE, Bayesian Beamformer)
- PEB framework offers a natural way to conduct multi-subject and multi-modal integration

Main references

- Friston et al. (2008) Multiple sparse priors for the M/EEG inverse problem
- Henson et al. (2007) Population-level inferences for distributed MEG source localization under multiple constraints: application to face-evoked fields
- Henson et al. (2010) A Parametric Empirical Bayesian framework for fMRI-constrained MEG/EEG source reconstruction
- Henson et al. (2011) A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration.
- Kiebel et al. (2008) Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG
- Lopez et al. (2014) Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM
- Mattout et al. (2007) Canonical Source Reconstruction for MEG
- Phillips et al. (2005) An empirical Bayesian solution to the source reconstruction problem in EEG