# Statistical Models for the analysis of ASL and BOLD Magnetic Resonance modalities to study brain function and disease

Aina Frau Pascual

Communauté
UNIVERSITÉ Grenoble Alpes

# THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques appliqués**

Arrêté ministérial : 25 Mai 2016


Présentée par

**Aina Frau Pascual**

Thèse dirigée par **Florence Forbes**
et codirigée par **Philippe Ciuciu**

préparée au sein **de l'Inria et du Laboratoire Jean Kuntzmann**
et de **Ecole Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**


# Statistical models for the analysis of ASL and BOLD Magnetic Resonance modalities to study brain function and disease

Thèse soutenue publiquement le **19 Decembre 2016**,
devant le jury composé de :

**M. Thomas Nichols**
Professor, University of Warwick, Rapporteur
**Mme. Patrícia Figueiredo**
Assistant professor, Instituto Superior Técnico Lisboa, Rapporteur
**M. Jean-Yves Tourneret**
Professeur des universités, Université de Toulouse, Examinateur
**M. Jan Warnking**
Chargé de recherche, Grenoble Institute de Neuroscience, Examinateur
**Mme. Florence Forbes**
Directeur de recherche, INRIA Grenoble Rhône-Alpes, Directeur de thèse
**M. Philippe Ciuciu**
Directeur de recherche, CEA Saclay, Co-Directeur de thèse

# *Abstract*

Functional and perfusion imaging modalities are closely related since they both measure, directly or indirectly, blood flow in the brain. Functional Magnetic Resonance Imaging (fMRI) using the blood oxygen level dependent (BOLD) contrast exploits the magnetic properties of blood (oxy- and deoxyhemoglobin) to measure local changes in blood oxygen concentration in the brain. The neurovascular coupling allows us to infer brain function from fMRI images. Perfusion MRI images the cerebral vascular system by directly measuring blood flow. In particular, Arterial Spin Labeling (ASL) does not need contrast agents; it uses spins of endogenous water protons as a tracer instead. Usually ASL is used to probe the basal perfusion at rest. However, in the recent years, it has also been used as a functional imaging modality (as fASL) by tracking task-related perfusion changes. In contrast to the standard BOLD fMRI, results are quantitative, making this type of data attractive for use in clinical research.

This thesis focuses on the investigation of the fASL modality and the development of new methods to analyze it. As previously done for BOLD data, a Bayesian framework is proposed for the analysis of fASL data. It provides a way of modeling activation values and both hemodynamic and perfusion response functions as probabilistic variables in the so-called joint detection estimation (JDE) framework. Bayesian models use a priori knowledge in the estimation of unknown parameters through the specification of probability distributions. In this work, we exploit this feature to incorporate physiological information to make the estimation more robust. In particular, we use physiological models based on the balloon model to derive a link between hemodynamic and perfusion responses and we turn this link into a prior distribution to regularize the estimation of the responses. A Markov Chain Monte Carlo solution with prior physiological knowledge has been first proposed for the estimation of the quantities contained in the fMRI signal. Since the computational cost of this algorithm is very high, we then reformulate the problem to use a variational expectation maximization approach that provides a much faster algorithm with similar results. The use of priors and constraints in this setting is also more straightforward.

These methods have been evaluated on two different datasets using event-related and block designs with very simple experimental tasks. We show the performance of the methods investigated in comparison to standard methods at the subject and group levels. Experimental results show the utility of using physiological priors for improving the recovery of a perfusion response function. They also demonstrate that BOLD fMRI achieves better sensitivity to detect evoked brain activity as compared to fASL although fASL gives a more localized activation, which is in line with the existing literature. From the results, we discuss the impact of the modelling of spatial correlation, as well as the impact of the estimation of temporal responses.

This work proposes new methodological contributions in the study of a relatively new fMRI modality that is functional ASL, and puts it into perspective with the existing techniques. Thus, we provide new tools for the neuroscientific community to study and understand brain function. These tools have been implemented in python in the PyHRF package.

# *Résumé*

Les modalités d'imagerie fonctionnelle et de perfusion sont étroitement liées car les deux mesurent, directement ou indirectement, le débit sanguin cérébral. D'une part, en utilisant le contraste BOLD (Blood-Oxygen-Level-Dependent), l'imagerie fonctionnelle par résonance magnétique (IRMf) exploite les propriétés magnétiques du sang (oxy et désoxyhémoglobine) pour y mesurer les changements locaux de concentration en oxygène: ce couplage neurovasculaire permet de déduire le fonctionnement du cerveau à partir des images IRMf. D'autre part, l'IRM de perfusion reflète le fonctionnement du système vasculaire cérébral en mesurant directement le débit sanguin cérébral. En particulier, l'IRM du marquage de l'eau artérielle (ASL) n'a pas besoin d'agents de contraste: le traceur est remplacé par des spins de protons endogènes d'eau. Habituellement l'ASL est utilisée pour mesurer la perfusion basale au repos. Toutefois, ces dernières années, il a également été utilisé comme une modalité d'imagerie fonctionnelle (comme la fASL) en mesurant les variations de perfusion cérébrale induites par la réalisation de tâches cognitives. Contrairement à l'IRMf standard basée sur le contraste BOLD, les résultats sont quantitatifs, ce qui rend ce type de données intéressantes pour son utilisation dans la recherche clinique.

Cette thèse porte sur l'étude de la modalité fASL et sur le développement de nouvelles méthodes pour l'analyser. Comme précédemment réalisé pour les données BOLD, un cadre bayésien est développé pour l'analyse des données fASL. Il fournit un moyen de modéliser les valeurs d'activation et les fonctions de réponse hémodynamique et de perfusion en tant que variables probabilistes dans l'approche de Détection-Estimation Conjointe. Les modèles bayésiens utilisent une connaissance a priori pour l'estimation des paramètres inconnus à travers la spécification de distributions de probabilité. Dans ce travail, nous exploitons cette fonction pour incorporer au modèle des informations physiologiques, afin de rendre l'estimation plus robuste. En particulier, nous utilisons des modèles physiologiques basés sur le modèle de ballon pour obtenir un lien entre les réponses hémodynamiques et de perfusion, puis nous utilisons ce lien dans une distribution a priori pour régulariser l'estimation des réponses. En utilisant information physiologique a priori, une solution de type Markov Chain Monte Carlo (MCMC) a été proposée pour l'estimation des quantités contenues dans le signal IRMf. Étant donné que le coût de calcul de cet algorithme est très élevé, nous reformulons le problème pour utiliser une approche variationnelle (VEM) qui fournit un algorithme beaucoup plus rapide avec des résultats similaires. Dans ce cadre, l'introduction d'information a priori et de contraintes est également plus simple.

Ces méthodes ont été évaluées sur deux ensembles de données différentes en utilisant des paradigmes événementiels et du bloc, pour des tâches cognitives très simples. Nous montrons les bonnes performances des méthodes proposées par rapport aux mèthodes standards, au niveau des sujets et du groupe. Les résultats expérimentaux montrent que les probabilités a priori physiologiques améliorent l'estimation d'une fonction de réponse de perfusion. Ces résultats démontrent également que le contraste BOLD a une meilleure sensibilité pour la détection de l'activité cérébrale évoquée que fASL, bien que la fASL donne une activation plus localisée, ce qui est conforme à la littérature existante. A partir de ces résultats, nous discutons l'impact de la modélisation de la corrélation spatiale, ainsi que l'impact de l'estimation des réponses temporelles.

Ce travail propose de nouvelles contributions méthodologiques pour l'étude de la fASL, et les met en perspective avec les techniques existantes. Ainsi, nous proposons de nouveaux outils pour la communauté neuroscientifique, mis en œuvre en python dans le package PyHRF, pour étudier et comprendre le fonctionnement du cerveau.

# *Acknowledgements*

My first word of gratitude goes to my advisors Florence Forbes and Philippe Ciuciu, for their guidance and work together during this thesis. I have learnt a lot from you both and you have grounded my future works in research. Thanks for giving me the opportunity of working with you. I also want to express my gratitude to Thomas Nichols and Patrícia Figueiredo for reviewing this work and for giving thorough and constructive feedback, and to Jan Warnking and Jean-Yves Tourneret for accepting being part of my jury and for all your comments.

I would also like to thank Bertrand Thirion for having me in the Parietal team and for all the feedback given, that made this work become better. Thanks also to Gaël Varoquaux, for being the happy person that makes everyone feel part of the team. Thanks to Stéphane Girard and Jean-Baptiste Durand, for their feedback and help when needed, and for the good mood and the nice talks. Thanks also to Jean-Michel Morel and Toni Buades for bringing me to Paris four years ago, when this adventure started.

I would like to express my gratitude to everyone that I could collaborate with. I want to thank everyone involved in the pyHRF project, but especially Jan Warnking and Michel Dojat, for the good and always constructive feedback, and for the joint work. I cannot forget to thank Thomas Vincent, who actively participated in some parts of this work, and from whom I learnt a lot. Thanks to Thomas Perret, who also participated very actively in some parts of this work. Thanks for the good mood, the good work, and the technical lessons. I would also like to thank Salma Bougacha for the feedback, for the process-asl toolbox that made everything easier, and for the nice talks. I want to thank Pablo Mesejo for the work together, for our research and non-research discussions, and for the good mood and good environment that you create. I want to thank Andrés Hoyos and Michael Eickenberg for our collaboration, with lots of learning from each other, for all the fun and time spent together. Special thanks go to Michael, for reading and giving feedback over most of this document. Thanks also for the everyday and altruistic help to Elvis Dohmatob, Alex Abraham, Pascal Rubini and Jaime Arias.

I want to thank the Mistis and Parietal teams, for the nice time spent together, the work, the feedback, the conferences, summer schools and retreats, the coffees, lunches, dinners and after-work beers, the running, hiking, skiing, climbing and swimming together: Pablo Mesejo, Thomas Perret, Alexis Arnaud, Clément Albert, Brice Olivier, Thibaud Rahier, Julyan Arbel, Jean Michel Bécu, Emeline Perthame, Pascal Rubini, Jaime Arias, Gildas Mazo, Nourou Sylla, Alessandro Chiancone, Karina Ashurbekova, Priscillia Previtero, Andres Hoyos, Michael Eickenberg, Salma Bougacha, Elvis Dohmatob, Daria La Rocca, Ana Luisa Pinho, Mehdi Rahim, Olivier Grisel, Loïc Esteve, Kamalakar Reddy, Alexandre Abraham, Carole Lazarus, Fabian Pedregosa, Viviana Siless, Yannick Schwartz, Denis Engemann, Danilo Bdzok, Solveig Badillo, Jérôme Dockes, Nicolas Chauffert, Konstantin Shmelkov, Virgile Fritsch. Thanks also to the assistants for their help in any administrative issue: Régine Bricquet, Myriam Étienne, Carol Pinto, Marion Ponsot, Helen Pouchot.

Thanks to everyone at Neurospin and Inria for the every day coffees, for the lunches, for the pingpong and babyfoot matches, for the good *pots*, for the pretreat, for the running together, for the afterwork beers, for the dinners and parties, for the GoT and game nights: Urielle Thoprakarn, Gregory Operto, Hugo Dary, Clémence Pinaud, Pedro Pinheiro, Yi-Chen Lin, Valentina Borghesani, Darinka Trübutschek, Martín Pérez-Guevara, Elodie de Spéville, Parvaneh Adibpour, Catherine Champseix, Mathilde Bouin, Antoinette Jobert, Karla Monzalvo, Leila Makkaoui, Olga Domanova, Sandrine Lefranc, Mathieu Dubois, Antoine Grigis, Nader Cheaib, Yann LePrince, Samuel David, David Goyard, Denis Rivière, Yann Cointepas, Edouard Duchesnay, Vincent Frouin, Robin Cherbonnier, José Ignacio Requeno, Aurora Scabia, Elisa Castaldi, Lola de Hevia, Clara Fischer, Baptiste Gauthier, Murielle Fabre, Audrey Fontaine, Lucie Bezombes, Sofía Zea, Claudia Mesejo, Keerthi Kurma, Poornima Puri, Gianni Delannoye.

I would also like to thank the people I met in Cachan for the lunches, dinners, coffees and parties, for the poker nights, for the museum visits, for the small trips, for the many times spent together: Hernan, Lara, Matías, Estíbaliz, Nicola, Martin, Aleksandra, Simon, Miquel, Claudia, Sandra, Cesar, Mahsa, Vincent, Christoph, and many other people.

Finally, I would like to thank my family, for being always there: Ana, Laura, Xavi, Belen, Juana, Jorge, Gregorio, Agapita, Maria, and especially to my father Toni, our pillar, who always pushed me to do whatever I wanted to do.

# *Contents*

# 1 Introduction

Neuroimaging techniques allow the *in vivo* study of brain function. In particular, functional Magnetic Resonance Imaging (fMRI) measures the changes induced by cerebral activity. The most widely used fMRI modality is BOLD for blood-oxygen-level-dependent [Ogawa et al., 1990] and, as the name indicates, it reflects the changes in oxygen concentration in the blood. When there is brain activity, oxygen is consumed and its concentration in the blood decreases; then there is an inflow of oxygenated blood to replenish the tissue, increasing blood oxygen concentration. Deoxygenated blood causes magnetic distortions that can be measured by an MRI scanner. For this reason, it is said that BOLD is an indirect measure of cerebral activity: it is a complex reflection of underlying physiological changes in oxygen consumption, and cerebral blood flow and volume [Buxton et al., 2004].

BOLD is one of the most used functional imaging modalities because it is non-invasive, non-ionizing, and it gives access *in vivo* to brain activity with a relatively high spatial resolution. To detect evoked activity, it relies on the contrast between different experimental conditions or between an experimental condition and a baseline, and in subsequent statistical analysis to assess the significance of this activation. BOLD does not give access to true physiological parameters such as cerebral blood flow or cerebral blood volume, but rather measures a mixture of these quantities that is difficult to untangle (except in some specific experimental setup where one can measure the cerebral metabolic rate of oxygen $CMRO_2$). For these reasons, BOLD is a very interesting tool in neuroscience, but in general it is not widely used for clinical applications. It cannot detect chronic changes in the baseline states [Buxton, 2013], as it is the case for normal ageing [Gauthier et al., 2013, Fabiani et al., 2014] and pathologies like Alzheimer's disease [Cantin et al., 2011] or Stroke [Attyé et al., 2014, Krainik et al., 2005]. In order to introduce functional imaging in clinical practice there is a need for accurate estimates of physiological parameters as cerebral blood flow (CBF) and volume (CBV), oxygen extraction fraction (OEF) or cerebral metabolic rate of oxygen ($CMRO_2$). This can be achieved with perfusion MRI techniques, since they image the vascular system by directly measuring blood flow. In particular, Arterial Spin Labeling (ASL) [Williams et al., 1992, Detre et al., 1992] does not need contrast agents: it uses spins of endogenous water protons as a tracer instead. ASL, as opposed to BOLD, provides a direct and more localized quantitative measurement of the CBF in absolute values, since the link between ASL and CBF has been described in the literature [Buxton et al., 1998b, Alsop et al., 2015]. This allows a

direct comparison between subjects, experiments and pathological/non-pathological population groups.

ASL is non-invasive, as opposed to other perfusion techniques, and highly reproducible [Wang et al., 2011b, Grade et al., 2015]. For these reasons, it is an emerging technique in clinical practice [Golay and Guenther, 2012] for its use in cerebrovascular diseases (e.g. Stroke), dementia [Wolk and Detre, 2012, Detre et al., 2012] or neuro-oncology [Grade et al., 2015]. The fact that it is non invasive makes it especially suitable for pediatric populations [Wang et al., 2003, Detre et al., 2012]. It has also been used in pharmacological MRI for drug studies [Chen et al., 2011, Detre et al., 2012]. When used in the context of functional MRI, ASL measures baseline CBF and CBF responses to stimuli [Liu and Wong, 2005, Mumford et al., 2006, Raoult et al., 2011].

This thesis focuses on the investigation of functional ASL (fASL) modality and the development of new methods for its analysis. This modality could give valuable information on brain function and so far has not been fully exploited in clinical research. Compared to the standard BOLD fMRI, the signal-to-noise ratio (SNR) and thus the spatio-temporal resolution of fASL data are limited, making its analysis a substantial challenge [Yang, 2002, Liu and Brown, 2007].

A Bayesian framework is proposed for the analysis of fASL data. It provides a way of modelling activation levels and hemodynamic and perfusion response functions as probabilistic variables in the so called Joint Detection Estimation (JDE) framework [Makni et al., 2008, Vincent et al., 2010, Chaari et al., 2013]. This way, we can account for differences in these quantities across the brain and obtain a quantification of their uncertainty. Bayesian models use *a priori* knowledge to compensate for the ill-posedness of the inverse JDE problem when it is formulated in the maximum likelihood sense. We can therefore incorporate physiological models such as the Balloon model to make the estimations more robust. In this work, it has been done through the derivation of a link between hemodynamic and perfusion responses from physiological models as the Balloon model [Buxton et al., 1998b, Friston et al., 2000], and its use as a prior in our Bayesian setting.

A Markov Chain Monte Carlo (MCMC) solution as in [Vincent et al., 2013a] has been first proposed in this thesis for the estimation of the quantities contained in the ASL fMRI signal. The first contribution of this thesis is the study of the use of physiological models as prior knowledge to facilitate the estimation of task-related perfusion and its temporal response, which is especially difficult to recover [Vincent et al., 2013a]. However, the computational load of this algorithm makes the estimation very time-consuming. For this reason, in a second contribution the problem is recast in a Variational Expectation Maximization (VEM) framework, which provides a much faster algorithm with similar results. The use of priors and constraints in this setting is also more straightforward.

These methods have been evaluated in two different datasets: AINSI, which uses event-related design, and HEROES, which uses block design. The acquisition of the HEROES dataset was performed in the context of this thesis and it contains BOLD, fASL and ASL data. Both datasets rely

on very simple tasks (visual, auditory and motor) as a way to validate the proposed methods in well known brain regions. With these datasets we show the performance of the methods investigated in comparison to standard methods at the subject and group levels.

This thesis proposes new methodological contributions in the study of a relatively new fMRI modality that is functional ASL, and puts it into perspective with the existing techniques. Thus, we provide new tools for the neuroscientific community to study and understand brain function. These tools have been implemented in python in the PyHRF package, an open-source package available on github[1].

[1] github.com/pyhrf/pyhrf

## 1.1 Manuscript structure

In the first chapter, we give an insight into functional and perfusion MRI, and what we want to measure with them. We first introduce the way the brain functions at a cognitive level and at a vascular level, and how this is linked through the neurovascular coupling. We also introduce the models that have been proposed in the literature to explain the neurovascular coupling. We summarize the existing neuroimaging techniques, focusing on the ones that are going to be used in the context of this thesis: BOLD fMRI and functional ASL.

In the second chapter, the classical methods used for the analysis of functional MRI are presented. We focus on the general linear model (GLM) and the statistical tests performed on the results. We also introduce the methods for HRF estimation. The same methods are used in the analysis of ASL fMRI, by adapting them to the particularities of the ASL signal.

In the third chapter, we introduce Bayesian inference and the methods used in the context of this thesis: Markov Chain Monte Carlo (MCMC) and Variational Expectation Maximization (VEM). As an application of these methods, we introduce the already existing Joint Detection Estimation (JDE) approach for the analysis of BOLD fMRI data.

In the fourth chapter, we introduce the JDE approach for ASL data analysis in its MCMC solution. The proposed model considers the inclusion of a physiological prior derived from the extended Balloon model and the hemodynamic model. Results on the AINSI dataset, with event-related design ASL data, are presented.

In the fifth chapter, a VEM solution to the JDE approach is presented for the analysis of functional ASL. The MCMC and VEM corresponding solutions are also compared, using the AINSI dataset.

In the sixth chapter, we present different versions of the Balloon and hemodynamic models, and try to evaluate their impact as well as the impact of the physiological parameters when injecting them in the JDE physiological prior for the analysis of ASL data.

In the seventh chapter, the HEROES dataset, a dataset acquired in the context of this thesis, is presented and it is analysed using the tools developed and explained in the previous chapters. Results are compared to the results found using classical methods.

## 1.2 Summary of contributions

In the context of this thesis, several contributions have been made in the analysis of BOLD and ASL fMRI data. The implementations can be found in the PyHRF toolbox[2].

[2] github.com/pyhrf/pyhrf

### Hemodynamically informed parcellation of BOLD fMRI

The analysis of BOLD fMRI data with the Joint Detection Estimation (JDE) approach needs the previous parcellation of the brain in hemodynamically homogeneous parcels. Usually we use functional atlases or we compute a functional parcellation of the brain from preprocessed fMRI time series using dedicated tools [B. Thirion et al., 2006, Thirion et al., 2014].

In this work, we attempt a fast hemodynamically based parcellation for its use prior to JDE inference. We propose a two-step approach consisting first of hemodynamics feature extraction, in which a general linear model (GLM) is used to discriminate hemodynamic information, followed by a parcellation of these features. Since there is a lack of hemodynamic information in the non-active voxels, the idea is to enforce grouping these uncertain voxels with neighbouring activating voxels. We want to avoid parcels that split active from non-active voxels in two classes because in the JDE prior model we assume both activation states in a parcel.

This contribution can be found in section 4.7.1.

### BOLD VEM multiple-session extension of the JDE approach

The fast solution of the JDE approach for BOLD fMRI presented in [Chaari et al., 2013] uses a variational expectation maximization (VEM) algorithm and considers a single session of BOLD data. [Chaari et al., 2013] shows the faster performance of this algorithm with respect to the Markov Chain Monte Carlo (MCMC) approach presented in [Makni et al., 2008, Vincent et al., 2010], with similar results. In fMRI, usually several sessions are acquired for the same subject to be able to compare them or combine them. In [Badillo et al., 2013b], a multiple-session extension of the JDE approach has been proposed to analyse several sessions together. The solution proposed uses MCMC and considers that the response levels have a mean value per condition and a common variance between sessions.

In the context of this thesis, a VEM solution of this extension has been implemented following the considerations made in [Badillo et al., 2013b]. Experimental results have shown that the solution of the multiple-session VEM is not very different from the average of the results computed with single session VEM. For this reason, we propose a heteroscedastic version of the multiple-session VEM. It amounts to considering session-specific variances. The goal is to be able to weight the importance of the different sessions so as to diminish the contribution of any potential noisy session to the final parameter estimates.

This contribution can be found in section 4.8.

### Derivation of a physiological link between PRF and HRF functions

The Balloon model [Buxton et al., 1998b], extended in [Friston et al., 2000], is a physiological model that links neuronal and vascular processes by considering the capillary as a balloon that dilates under the effect of blood flow variations. [Buxton et al., 1998b] proposed an expression that links the BOLD response to the physiological quantities in the Balloon model. By linearizing the system of equations that this model describes, we derive a linear operator that links perfusion and hemodynamic responses. The linear operator that links perfusion and hemodynamic responses gives us a new tractable tool for analyzing the ASL signal.

This contribution[3] can be found in section 5.2.

[3] This contribution was started by J. Sloboda, as part of her master thesis.

### Physiologically informed ASL MCMC solution

[Vincent et al., 2013a] proposes a JDE approach for the analysis of ASL data using MCMC. Perfusion component results using this approach were not completely satisfying, due to the fact that ASL is a very noisy signal and its perfusion component is really small. In this work, we introduce the linear operator linking hemodynamic and perfusion components as *a priori* knowledge in the ASL JDE approach presented in [Vincent et al., 2013a]. Different ways of introducing this prior are investigated.

This contribution can be found in chapter 5.

### Physiologically informed ASL VEM solution

The MCMC solution of JDE proposed in [Vincent et al., 2013a] is very computationally demanding. For this reason, following the lead of [Chaari et al., 2013] in the analysis of BOLD data, a variational expectation maximization (VEM) solution is derived for functional ASL data. This solution gives a similar performance with a much lower computational cost. The use of physiological priors and the introduction of constraints in the norm of the response have also been investigated for this solution.

This contribution can be found in chapter 6.

### Physiological models comparison for the analysis of ASL fMRI data

The physiological linear operator derived from the extended Balloon model contains a set of parameters that have to be fixed. We usually fix them relying on physiological parameters described in the literature. Moreover, different versions of the extended Balloon model exist. In this work, we compare different physiological models and parameters to see their impact in the JDE estimates, when using the physiological prior in the JDE MCMC algorithm. We observe that the results are the same no matter which model or set the parameters we use. However, the setting chosen impacts the convergence of the solution. This allows us to conclude that the set of physiological parameters has more impact than the model version being used. For the data used, we conclude that the best set of parameters are the ones proposed in [Khalidov et al., 2011]. This information is valuable in itself and in future work we would like to be able to estimate these physiological parameters from the data as in [Mesejo et al., 2015, 2016], and

use them to improve the convergence of JDE.

This contribution can be found in chapter 7.

### Validation of the methods developed

A data set with a mini-block experimental design was acquired and analysed in the context of this thesis. We analyse the performance of the JDE VEM method for BOLD and ASL, and we put it into perspective with the classical GLM approach, at subject and group levels. In this context, we show the impact of the spatial modelling in JDE through a Markov random field with respect to the spatial smoothing in GLM. We also analyse the effect of the modelling of temporal responses in JDE. Cerebral blood flow quantification is also investigated.

This contribution can be found in chapter 8.

## 1.3 Publication list

- A. Frau-Pascual, T. Vincent, F. Forbes, and P. Ciuciu. *"Hemodynamically informed parcellation of cerebral fMRI data"*. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2079–2083. IEEE, 2014a.

- A. Frau-Pascual, T. Vincent, J. Sloboda, P. Ciuciu, and F. Forbes. *"Physiologically informed Bayesian analysis of ASL fMRI data"*. In Bayesian and grAphical Models for Biomedical Imaging, pages 37–48. Springer, 2014.

- A. Frau-Pascual, F. Forbes, and P. Ciuciu. *"Variational physiologically informed solution to hemodynamic and perfusion response estimation from ASL fMRI data"*. In Pattern Recognition in NeuroImaging (PRNI), 2015 International Workshop on, pages 57–60. IEEE, 2015a.

- F. Forbes, A. Frau-Pascual, P. Ciuciu. *"Méthode d'approximation variationnelle pour l'analyse de données d'IRM fonctionnelle acquises par Arterial Spin Labelling"*. Colloque du Groupe d'Études du Traitement du Signal et des Images (GRETSI), Sep 2015, Lyon, France.

- A. Frau-Pascual, F. Forbes, and P. Ciuciu. *"Comparison of stochastic and variational solutions to ASL fMRI data analysis"*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, pages 85–92. Springer, 2015b.

- A. Frau-Pascual, F. Forbes, and P. Ciuciu. *"Physiological models comparison for the analysis of ASL fMRI data"*. In Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, pages 1348–1351. IEEE, 2015.

# 2 Introduction to perfusion and functional imaging

The study of the human brain has been one of the challenges of neuroscience since the 19th century, and still is. Evidence of this is the large number of initiatives that have been created during the last years to study brain function: Human Brain Project, the BRAIN initiative, Blue Brain Project...

The brain is the centre of the central nervous system and it centralizes the control over the other organs of the body. It is composed by neurons and glia (see figure 2.1), cells that provide support to the neurons. Neurons' signal is transmitted by electrical impulses along their axons in the form of electrochemical pulses called action potentials. They communicate with other cells via synapses, membrane-to-membrane structures that permit to pass an electrical or chemical signal between the cells. Neuron bodies together with capillary blood vessels form gray matter tissue and neuron axons form white matter tissue. Gray matter is associated with processing and cognition.

The brain uses huge amounts of energy for an organ of its size, and for this reason the blood supply never stops. A lack of blood would lead to the damage or death of tissue. The arterial blood arriving to the brain comes from the two internal carotid arteries and the two vertebral arteries. They meet in the Willis polygon (see figure 2.2), a circle that then gives rise to other arteries that bifurcate and arrive to all regions of the brain. There is also a venous drainage system to drain non oxygenated blood and other damaging substances.

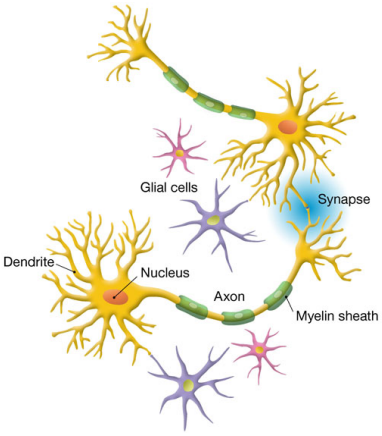



Figure 2.1: Neuron structure: nucleus, dendrites, axon. They communicate through synapse. Glia cells provide support to neurons.

http://www.fmcpaware.org

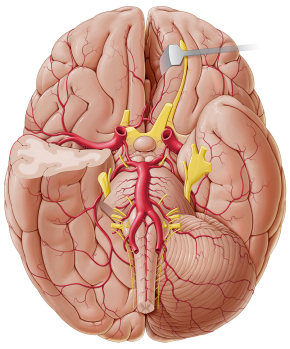


Figure 2.2: The circle of Willis allows all parts of the brain to receive blood even if the supply from one of the major arteries is compromised.

www.kenhub.com

## 2.1 Neurovascular coupling

Most of the energy that the brain consumes is spent in the firing of action potentials, that trigger neuro-transmission from one neuron to another, and in synaptic activity. Some energy is also spent in inhibitory activity to stabilize the neuron membrane potential [Buxton, 2013]. All this energy comes from the transformation of adenosine triphosphate (ATP) when it loses a phosphate. ATP is generated by an oxidative metabolism of glucose: glucose and oxygen are transformed into $CO_2$ and ATP [Buxton, 2013]. That is the reason why glucose and oxygen supply is crucial for brain function. If the supply is not adequate brain tissue can die and, since the oxygen concentration in tissue is quite low, the supply has to be continuously

maintained. Glucose and oxygen are delivered by blood flow, and with it $CO_2$ is also diluted. Cerebral blood flow (CBF) is the volume of arterial blood delivered to the tissue element by minute (ml/100g/min).

The brain controls the increase in blood flow to active regions with neurovascular coupling mechanisms, that can be different depending on the amount of energy and how it is used. As opposed to what it was thought, there is no metabolic signal that triggers blood flow depending on blood oxygen or $CO_2$ concentration [Attwell et al., 2010]. The way it is controlled is through neurotransmitter-mediated signalling [Attwell et al., 2010]. Neurons and astrocytes may act on arteriole smooth muscle cells (see Fig. 2.3), releasing molecules that dilate or constrict them to increase or decrease blood flow. Pericytes are thought to control the diameter of capillaries [Huneau et al., 2015].

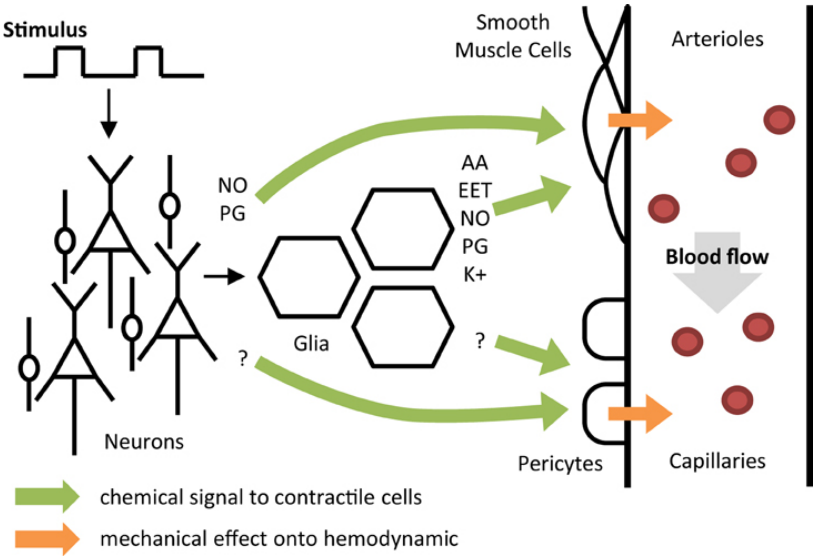



Figure 2.3: Scheme representing neurovascular coupling. Colored arrows represent the relationship between different activities: neurons, glia, smooth muscle cells, pericytes, and blood flow. Molecules that they release to increase or decrease blood flow: NO, nitric oxide; PG, prostaglandin; AA, arachidonic acid; EET, epoxyeicosatrienoic acids; K, potassium. Acknowledged from [Huneau et al., 2015].

Neurovascular coupling is an active area of study and the exact mechanisms are still unclear. [Iadecola and Nedergaard, 2007] provides a review on neurovascular coupling. Evidence suggests [Huneau et al., 2015] that neurovascular coupling can be modified in cerebrovascular and degenerative disorders [Cantin et al., 2011], in intracranial and extracranial vascular disorders [Hamzei et al., 2003] or after ischemic stroke [Krainik et al., 2005].

### 2.1.1 Oxygen metabolism

Oxygen plays a key role in brain function since it is combined with glucose to create ATP, which is a source of energy. For this reason, oxygen consumption ($CMRO_2$) is used as an indicator of energy consumption and therefore brain activity. Since oxygen is difficult to carry, it is bind to hemoglobin and carried through the vasculary system. In blood, there is an equilibrium between oxygen dissolved in plasma and oxygen bind to hemoglobin. When oxygen from the plasma diffuses into the tissue, this equilibrium is broken and hemoglobin compensates by leaving oxygen into the plasma [Buxton, 2013]. This way, oxyhemoglobin (hemoglobin carrying

oxygen) becomes deoxyhemoglobin (hemoglobin not carrying oxygen), and the ratio of oxygenated and deoxygenated blood changes.

Since deoxyhemoglobin is paramagnetic, it alters the magnetic susceptibility of blood. The difference in susceptibility between blood vessels and the surrounding tissue creates local magnetic field distortions [Ogawa et al., 1990].

The fraction of oxygen molecules delivered to the capillary bed is the oxygen extraction fraction (OEF). According to [Buxton et al., 2004], it relates to the other quantities in steady state as $CMRO_2 = OEF \cdot [O_2]_a \cdot CBF$, where $[O_2]_a$ is the arterial oxygen concentration. With neural activation, the increase in CBF is approximately twice the oxygen consumption, which causes a decrease in OEF, assuming that $[O_2]_a$ is kept constant.

## 2.2 Imaging perfusion and brain function

Functions associated with the different areas of the brain have been studied for a while now [Brodmann, 2007] (see figure 2.4).

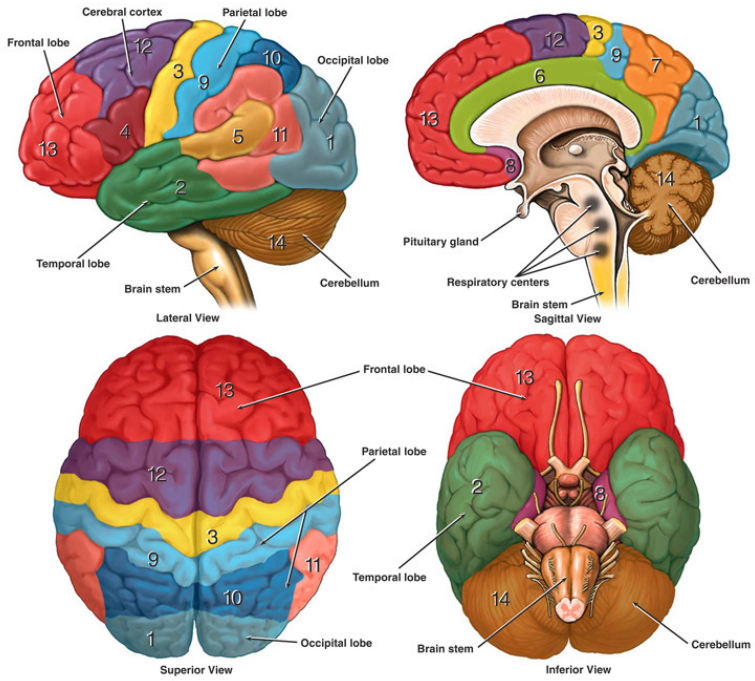



Figure 2.4: Functions associated to different areas of the brain: (1) visual; (2) association; (3) motor function as voluntary movement; (4) Broca's area for control of muscles of speech; (5) auditory; (6) emotional; (7) sensory association; (8) olfactory; (9) sensory; (10) somatosensory association; (11) Wernicke's area for written and spoken language comprehension; (12) motor function as eye movement and orientation; (13) higher mental functions as planning, emotions, judgement, creativity...; (14) motor functions in the cerebellum as coordination, balance, equilibrium and posture.

Credit: Nucleus Medical Art, Inc./Getty Images. http://dana.org

The development of neuroimaging techniques allowed the neuroscientific community to study brain function *in vivo*, in the healthy and pathological conditions. Since brain function is related to blood oxygen supply, the access to blood perfusion (the arrival of blood supply to a tissue) with neuroimaging is also an important tool for brain research. Different imaging techniques have been developed following different principles, and they are called functional or perfusion imaging depending on which principle they use. Because of the relationship between function and blood oxygen supply, perfusion and functional imaging are closely related, and this complementarity can be used to access more accurate measurements.

Some of the mostly used functional imaging techniques are sensitive to indirect changes in blood oxygenation, or measure direct or indirectly electrical activity:

- **Near-infrared spectroscopy (NIRS)** monitors blood hemoglobin levels and detects changes in its concentration associated with neural activity through the determination of optical absorption coefficients. The absorption of light in the near-infrared region of the electromagnetic spectrum ( $\sim 700 - 2500$ nm) reflects blood hemoglobin levels. It is non-invasive, non-ionizing, and the equipment is not expensive. However, it is difficult to interpret, requires a complex calibration, and measures signals only close to the outer layer of the cortex.

- **Electroencephalography (EEG)** records electrical activity of the brain with electrodes placed along the scalp. It measures voltage fluctuations resulting from ionic current within the neurons of the brain. It is usually non-invasive, although invasive electrodes can be also used in specific applications (stereotactic EEG). It has a high temporal resolution but a poor spatial resolution (order of ms and cm). The currents near the scalp are easier to detect. Its acquisition can be combined with fNIRS as illustrated in figure 2.5.

Figure 2.5: fNIRS and EEG can be acquired simultaneously.

nirx.net

- **Magnetoencephalography (MEG)** records magnetic fields produced by neural electrical activity using very sensitive magnetometers and gradiometers (figure 2.6). It has a high temporal resolution (order of ms) and a low spatial resolution (order of cm), although better than in EEG. EEG and MEG can be measured simultaneously.

Figure 2.6: Magnetoencephalography (MEG) acquisition. At Neurospin, CEA Saclay.

http://i2bm.cea.fr

- **BOLD functional MRI (fMRI)** measures blood-oxygen-level dependent (BOLD) changes, which are indirectly related to brain activity. It is non-invasive, it has relatively good spatial resolution (order of mm) and it has access to all brain regions. However, the temporal resolution is not very high (order of s) compared to EEG or MEG.

Other imaging techniques can get quantitative results by measuring physiological quantities such as brain perfusion: the arrival of blood supply to a tissue. Perfusion quantification has been proven to be useful for clinical use, in the study and treatment of vascular diseases. Several perfusion imaging techniques are available for the human brain, although most of them need the use of a contrast agent and are therefore invasive.

- **Positron emission tomography (PET)** detects, using a gamma camera (see figure 2.8), pairs of gamma rays emitted indirectly by a tracer (a positron-emitting radionuclide) that is injected in the subject being scanned. Tissue metabolic activity can be measured using fludeoxyglucose (FDG) as tracer. The concentration of this tracer, and therefore the level of gamma rays, reflects brain activity since concentration of blood also increases locally with brain activity (figure 2.7). PET provides a quantitative measure of physiology and it is not sensitive to small movement, but it has very low spatio-temporal resolution (order of min-hour, and 10 mm) and it is a ionizing imaging modality. Usually, the tracer is injected once and it lasts around one hour

in the case of FDG before the effect decreases. It can be also injected continuously in a lower dose, although it is less common.

- **Single-photon emission computed tomography (SPECT)** detects, using a gamma camera, gamma rays emitted by a radio-tracer previously injected in the subject. In functional brain imaging, the gamma-emitting tracer used (99mTc-HMPAO) is absorbed by brain tissue in a manner proportional to blood flow, and blood flow reflects brain activity. As in PET, the radio-tracer is injected once for a functional experiment. The resolution is similar to PET.
- **Computed tomography (CT) perfusion** consists of the dynamic sequential scanning of a pre-selected region of the brain during the injection of a bolus of iodinated contrast material as it travels through the vasculature.
- **Dynamic susceptibility contrast imaging (DSC-MRI)** consists of the injection of Gadolinium contrast agent and the posterior fast acquisition of T2*-weighted magnetic resonance images, dominated by transversal relaxation (see section 2.3). The Gadolinium reduces the T2* intensity in tissues depending on its local concentration. It can be used to measure cerebral blood volume and blood flow. DSC-MRI was used for the first functional MRI experiment ever done [Kwong et al., 1992].
- **Dynamic contrast-enhanced MR perfusion (DCE-MRI)** consists of the injection of Gadolinium contrast agent and the posterior acquisition of fast T1-weighted MR images, dominated by longitudinal relaxation (see section 2.3). It reflects tissue perfusion, vessel permeability, and extravascular-extracellular space (space between neuron and vessel). Spatial and temporal resolutions of DCE-MRI and DSC-MRI are of the order of mm-cm, and minutes, respectively.
- **Arterial spin labelling (ASL)** uses magnetically labelled blood as an endogenous tracer while acquiring T2*-weighted images. Magnetically labelled and non-labelled volumes can be subtracted to measure perfusion. Many volumes are acquired and averaged to get a clear measure of perfusion, that can be quantitative. Compared to other perfusion techniques, it is non-invasive and non-ionizing. Vasoreactivity experiments can be carried out to study CBF changes [Nöth et al., 2006, Pollock et al., 2009, Krainik et al., 2013, Villien et al., 2013]. The functional version of ASL, functional ASL (fASL), adds the task performance while the acquisition of images is done. Spatial and temporal resolutions are lower than in BOLD imaging, although in the same order of magnitude.

In this work, we focus on BOLD and fASL functional MRI modalities.

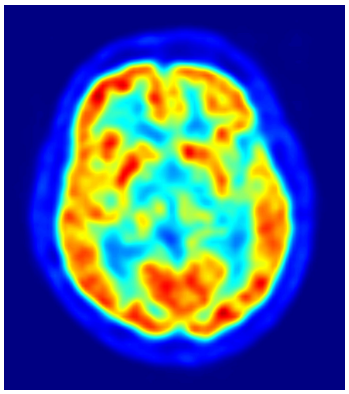


Figure 2.7: Positron emission tomography (PET) scan.

wikipedia.org

Figure 2.8: PET and CT scanner that can be found at the Service Hospitalier Frédéric Joliot (SHFJ).

http://i2bm.cea.fr

## 2.3 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) uses nuclear magnetic resonance (NMR): a proton inside a magnetic field aligns its spin with the magnetic field vector and can absorb and re-emit electromagnetic radiation (a radio-frequency or RF pulse) at the precessing[1] frequency. This frequency is

[1] precess: change in the orientation of the rotational axis of a rotating body

called Larmor frequency and it depends on the static magnetic field and the nucleus itself through the gyromagnetic ratio: $w_0 = \gamma B_0$, where $B_0$ is the magnetic field vector (see figure 2.9) and $\gamma$ the gyromagnetic ratio of the nucleus. The gyromagnetic ratio can be computed as $\gamma = \frac{q}{2m}$, where $q$ is its charge and $m$ is its mass. The proton emits energy until it aligns again with the magnetic field. This process is called relaxation.

In MRI, a large cylindrical magnet creates a magnetic field around the subject, that is placed inside (see MRI scanner in figure 2.10). Then, radio waves are sent and their echo signals are collected and used to construct an image. A space encoding is used to ensure that each point in space has a specific radio frequency at which the signal will be sent and received.

The strength of the signal depends on the proton density (PD) in the tissue, the longitudinal relaxation time T1 and the transversal relaxation time T2. How these parameters contribute to the image intensity depend on the pulse sequence used: it can highlight different tissue properties and reflect different contrasts. This depends on two acquisition parameters: repetition time (TR) and echo time (TE) of the scan.

- TR is the time between successive RF pulses: if short, protons from tissues with longer T1 will not have fully relaxed before the next measurement and the signal from this tissue will be lower. As a consequence, tissues with shorter T1 will have a higher transverse magnetization amplitude after the next excitation.
- TE is the time at which the signal is measured: if short, the amount of dephasing that can occur in tissues where protons are constrained by structures (white and grey matter) is reduced. If TE is long enough (but not too long or signal disappears) differences in transverse relaxation will alter tissue contrast. This is called T2 effect.

For getting T1-weighted images, a short TR ($<$ 500 ms) and short TE ($<$ 30 ms) are used. For a T2-weighted image, a long TR ($>$ 1500 ms) and long TE ($>$ 90 ms) are used. The transverse magnetization contains effects of macroscopic differences in the magnetic field, and intra- and intermolecular differences in the magnetic field. The second effect corresponds to the T2 value of the tissue, and the combination of both corresponds to the T2* value. T2* value can capture magnetic inhomogeneities as the BOLD effect. Note that only specific NMR pulses allow to perform T2 imaging, namely spin echo (SE) sequences, which are pretty slow. Faster NMR sequences rely on the gradient echo (GE) principle and give you access to the T2* measurements. Figure 2.11 shows how T1-weighted and T2-weighted images look like.

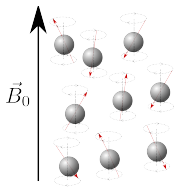



Figure 2.9: Proton spins align with the magnetic field vector $B_0$. They precess at a frequency that depends on this magnetic field $w_0 = \gamma B_0$.

Figure 2.10: 3T MRI scanner that can be found at Neurospin, CEA Saclay.

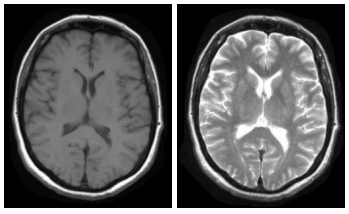


Figure 2.11: T1 (left) and T2-weighted (right) images. The contrast of the tissues is different: in T1-weighted images fat and white matter is bright, and in T2-weighted images grey matter and cefalospinal fluid are bright.

casemed.case.edu

## 2.4 Functional MRI

In functional MRI, brain images are acquired while a subject placed inside the scanner is performing some task or is submitted to some sort of stimulus. One stimulus or task represent a condition in our experimental design. They have to be carefully chosen depending on the test we want to perform. Designs are usually based on the subtraction of different conditions as a way to reflect differences in cognitive processes. We can

also compare a condition with rest, although it is better to use a controlled baseline condition. Designs can be event-related, consisting of very short and isolated stimulus presentation; or block, consisting of the repeated presentation of a condition during a certain amount of time (e.g. $\sim$ 30 sec). Although it is easier to interpret and control event related designs since we can have many stimulus randomized, their statistical power for detecting evoked brain activity is much lower than block designs. Sometimes mini-block ($\sim$ 10 s) designs are used as a compromise between the two.

The term fMRI usually refers to task fMRI. However, there are other modalities as resting state fMRI (rs-fMRI), that measures brain function at rest and does not involve the performance of a task.

### 2.4.1 Blood Oxygen Level Dependent fMRI

Blood Oxygen Level Dependent (BOLD) signal [Ogawa et al., 1990] measures the ratio between oxy and deoxy-hemoglobin. Both hemoglobin states have different magnetic properties and when there is an increase of deoxyhemoglobin in blood, as it is the case when oxygen is consumed due to neural activity, the magnetic susceptibility of blood is altered compared to the surrounding tissue. This creates local magnetic inhomogeneities that decrease BOLD signal. Then an inflow of oxygenated blood, much higher than the oxygen consumed, changes this ratio again and BOLD signal increases (figure 2.12). The oversupply of oxygen leads to a more oxygenated venous blood. That is why we talk about BOLD effect as a venous effect.

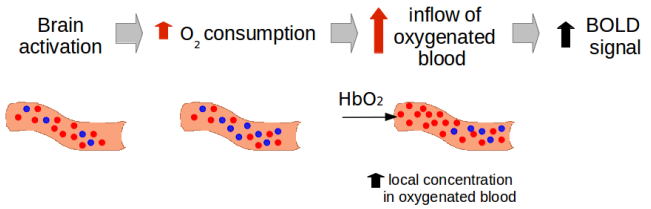



Figure 2.12: The BOLD signal measures the ratio between oxy- and deoxy-hemoglobin in the blood. This ratio changes during brain activity.

BOLD signal changes with respect to baseline depend on simultaneous changes in cerebral blood flow (CBF), cerebral blood volume (CBV) and oxygen consumption ($CMRO_2$), and it is difficult to separate these effects: CBF increase causes a decrease of deoxyhemoglobin; oxygen consumption increases deoxyhemoglobin; venous CBV increase causes an increase of deoxyhemoglobin; and arterial CBV increase causes a decrease of deoxyhemoglobin. For this reason, BOLD is an indirect effect of neural activity. See [Buxton, 2013, Logothetis and Wandell, 2004, Logothetis et al., 2001] for further explanations.

The response that reflects all these simultaneous changes is called the hemodynamic response function. Its canonical shape (figure 2.13) peaks around 5.5 seconds after neural stimulation, followed by an undershoot around 10 seconds after the peak, and then it goes back to baseline. It can have a small dip at the beginning of the response. It is usually characterized by a parametrized function, namely the subtraction of two Gamma functions since [Glover, 1999], adjusted from a set of auditory and sensorimotor activity signals.

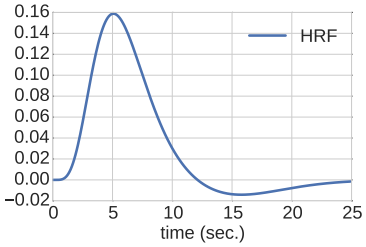



Figure 2.13: Hemodynamic response function generated with the difference of 2 Gamma functions.

In order to detect BOLD signal changes, we need acquisition times smaller than the BOLD changing rates. For that reason, high speed acquisition sequences such as echo planar imaging (EPI) are used. EPI sequence allows the acquisition of a slice over a period of about 60 ms (single slice TR) with a single RF pulse, although this comes at the expense of lower spatial resolution (e.g. $2 \times 2 \times 2$ mm$^3$) compared to traditional and longer sequences. A typical repetition time (TR), here between two consecutive volumes, would be 2.4 seconds for 40 slices acquired in 60 ms each. Using parallel imaging, compressed sensing or simultaneous multislice acquisition, one can decrease the TR.

### 2.4.2 Arterial spin labelling

Arterial Spin Labelling (ASL) [Williams et al., 1992, Detre et al., 1992] uses magnetically labelled blood as an endogenous tracer. The magnetic labeling consists of an inversion of the magnetization of the water spins contained in the blood of a certain region. The labelled region varies depending on the ASL method. When magnetically labelled (or tagged) arterial blood arrives to brain tissue, it causes magnetic disturbances that form the ASL signal. For this reason, in contrast to BOLD, one say that ASL measures an arterial effect. The perfusion signal lies on the difference between a control and a magnetically tagged image ($M_{control}$ and $M_{tag}$ respectively, $M$ for magnetization), that are acquired in a temporally interleaved manner and form a time series. This difference ($\Delta M$) shows the movement of the magnetically tagged molecules [Liu and Brown, 2007] (see also figure 2.14):

$$\Delta M = M_{control} - M_{tag} \tag{2.1}$$

The control image is subtracted to remove the contribution of the static tissue to the tag image. The intensity of the resulting perfusion image can be then transformed into a quantitative regional cerebral blood flow measure. This measure will be therefore comparable between subjects, experiments and pathological/non-pathological conditions. Depending on how the magnetic tagging is done, there are different ASL methods [Liu and Brown, 2007].

*Pulsed ASL (PASL)* [Edelman et al., 1994] uses short radiofrequency pulses to invert the spins in a certain region (see figure 2.15(a)). This method has a high spin inversion efficiency and does not need a high RF power, but it depends on the coverage and $B_1$ homogeneity of the RF field.

*Continuous ASL (CASL)* uses long RF pulses together with a constant gradient field applied to a plane of spins (see figure 2.15(b)). When they are properly adjusted, they cause flow-driven adiabatic inversion [Williams et al., 1992]: the tagging of inflowing spins that move at a certain velocity. During the control image acquisition, an RF pulse is applied outside the brain to account for magnetic distortions caused by the RF pulse. Distortions will be subtracted with the control image. This method allows tagging closer to the acquired region, but needs high RF power.

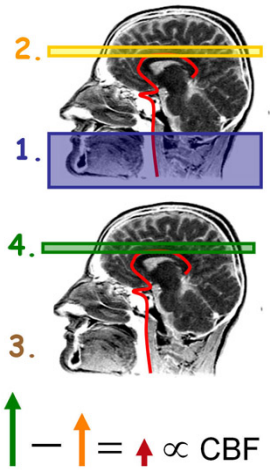



Figure 2.14: ASL acquisition: (1) Magnetic tagging of inflowing arterial blood; (2) Image acquisition; (3)-(4) Acquisition of a control image; (5) $\Delta M$ is proportional to regional CBF.

http://fmri.research.umich.edu

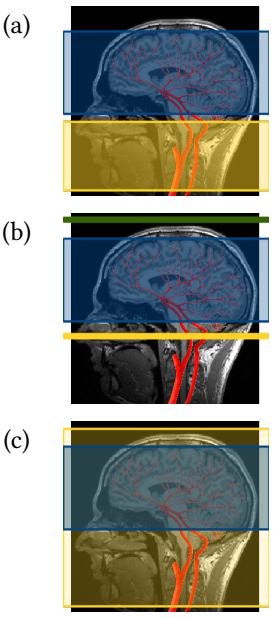



Figure 2.15: ASL tagging methods: (a) pulsed, (b) continuous and (c) velocity-selective. In the images, the blue and yellow rectangles indicate the imaging and tagging regions, respectively. In (b) the green rectangle indicates an RF pulse applied during the control image acquisition.

*Pseudo-continuous ASL (PCASL)* [Garcia et al., 2005] is a compromise between pulsed ASL and continuous ASL. It uses a train of rapidly repeating low tip RF pulses and alternating sign magnetic field gradients.

*Velocity Selective ASL (VS-ASL)* [Wong et al., 2006] uses a RF and a gradient pulse train to invert spins that flow faster than a certain velocity. The tagging is not space-dependent (see figure 2.15(c)).

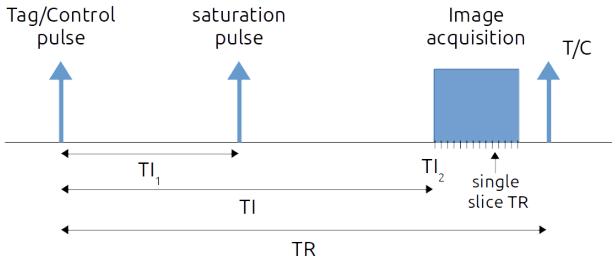



Figure 2.16: Acquisition scheme for pulsed ASL: tag/control pulse creates the bolus; at $TI_1$ there is a saturation pulse and tagged spins keep arriving to the tissue; at $TI_2$ the image is acquired; after a TR there is a new volume acquisition.

In this work, we focus on pulsed ASL [Luh et al., 1999]. In this ASL method, the tagging is usually done in the neck to tag the inflowing blood from the carotid arteries. There is a tag/control pulse (see figure 2.16) that inverts the magnetization of the spins in the tagging region and then a saturation pulse that allows to saturate to zero the tagged spins that are still in the tagged region at time $TI_1$. The tagged spins are perfused to the brain while undergoing relaxation and they arrive to the blood vessels after a transit time ($\Delta t$ in Fig. 2.17). When they arrive to the tissue and they exchange with the spins in the capillary bed, they alter the local tissue's longitudinal magnetization [Aguirre et al., 2005] (spins shift from $T1_{blood}$ to $T1_{tissue}$) and the perfusion signal increases. After the tagging stops ($TI_1$) and the last tagged spins arrive to the vessels ($\tau$ in Fig. 2.17), the signal starts dropping. The tag image is acquired next at that moment ($TI_2$). See figure 2.17.

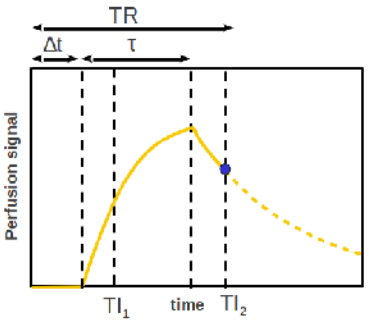



Figure 2.17: Magnetically labelled spins are perfused to the brain until they arrive to the vessels after a transit time $\Delta t$, and the perfusion signal starts increasing. Then the tagging is stopped and the spins continue arriving to the vessels until the last ones arrive at time $\tau$, and the perfusion signal starts dropping. The image is acquired at $TI_2$.

The difference signal between tag and control states at image acquisition time $TI_2$ will be proportional to perfusion. The spins that were saturated after $TI_1$ will not contribute to the signal: their magnetization was set to zero with the saturation pulse, their relaxation will be different to the tagged spins and their signal will be cancelled by subtraction.

Control and tag images are acquired using an EPI sequence and two TR are required to get 1 perfusion measurement by subtraction. Typical value for TR is 3 seconds.

### CBF quantification

For pulsed ASL, the link between CBF and the measured difference $\Delta M$, is described in [Alsop et al., 2015] (based on [Buxton et al., 1998a]) for QUIPS II PASL imaging, as:

$$CBF[mL/100g/min] = \frac{6000\lambda\Delta M}{2\alpha TI_1 M_0}\exp\left(\frac{TI_2}{T_{1,blood}}\right) \tag{2.2}$$

where $\Delta M = M_{control} - M_{tag}$ is the averaged control-tag magnetization; $M_0$ is the relaxed magnetization; $\lambda = 0.9$ mL/g is the averaged brain/blood

partition coefficient (T1 decay correction factor); $\alpha = 0.98$ is the labeling efficiency (tissue water exchange correction factor); $T_{1,blood} = 1650$ ms is the longitudinal relaxation time of the blood at 3T. This model assumes that all the tagged blood is delivered to the target tissue, that there is no outflow of labelled blood, and that the relaxation of the labeled spins is governed by blood T1.

This scaling of the data is very sensitive to time and space changes, and we need to consider the time difference between slices [Nöth et al., 2006]. It is recommended to be performed before any other preprocessings of the data. Other recommendations when dealing with this data can be found in [Alsop et al., 2015].

The relaxed magnetization ($M_0$) can be measured in a separate scan with a PD-weighted image with similar readout to ASL. Dividing by this image, we correct for signal variations caused by RF coil inhomogeneities and differences in transverse relaxation. When the TR is lower than 5 s, we multiply the PD-weighted image by $1/(1-\exp(-TR/T_{1,tissue}))$, where $T_{1,tissue}$ is assumed to be the $T_1$ of gray matter, in order to correct for longitudinal relaxation.

To measure voxelwise $T_1$ values, $T_1$ mapping can be performed. The gold standard for $T_1$ mapping is the Inversion Recovery method [Bydder and Young, 1985]. However, other methods as the Variable Flip Angle (VFA) method [Fram et al., 1987] have lower scan times and can be easily introduced in an ASL sequence. VFA computes T1 value from two gradient echo images acquired with different flip angles but the same TR. The real value of these flip angles is not uniform across brain regions due to differences in the RF field. To account for this, VFA requires a previous B1 correction that can be performed with the Actual Flip Angle (AFI) method [Yarnykh, 2007]. AFI determines the flip angle distribution using two signals with different TR. Other scaling factors as $\lambda$ can be also corrected to account for regional differences in tissue and blood water content.

Typical gray matter CBF values are from 40 to 100mL/100g/min [Alsop et al., 2015]. In pulsed ASL, an underestimation of gray matter CBF values has been reported in the literature [Wang et al., 2011a, Figueiredo et al., 2005].

### 2.4.3 Comparing BOLD and ASL

ASL and BOLD are related because both measure a BOLD effect. In the case of ASL, it is present in the non-subtracted control/tag time series. In ASL we need to perform the tagging before the acquisition and this takes time ($\sim 2$ s). To compensate for this, the time spent in actually acquiring the image is rather short ($\sim 0.5$ s) and this implies a quite low spatial resolution. We also need to take into account that the perfusion in ASL is extracted from the subtraction of two consecutive images (tag and control), so we have one perfusion measure every two acquired images, and there is a time shift between these consecutive images that needs to be considered. Therefore, the spatial and temporal resolution of BOLD is higher than in ASL [Liu and Brown, 2007]. ASL acquisition is also noisier than BOLD acquisition, due to the fact that the TE used for ASL is not optimal for the BOLD contrast

[Ogawa et al., 1993]. This makes the BOLD effect, which is present in both modalities, easier to extract from a BOLD acquisition.

Although ASL faces many challenges, it is interesting because it gives a quantitative measurement of absolute CBF without the need of any external tracer. This means that the measurement is direct and closer to actual neural activity: it measures arterial blood flow. This translates into lower inter-subject and inter-session variability [Tjandra et al., 2005, Leontiev and Buxton, 2007, Raoult et al., 2011, Pimentel et al., 2013], and into a more localized detection of neuronal activity [Luh et al., 2000, Tjandra et al., 2005, Raoult et al., 2011, Pimentel et al., 2013]. BOLD and functional ASL have been compared in experiments involving motor tasks [Raoult et al., 2011, Pimentel et al., 2013], visual tasks [Leontiev and Buxton, 2007], and speech tasks [Kemeny et al., 2005].

A general comparison of BOLD and ASL can be found in [Liu and Brown, 2007, Detre and Wang, 2002].

### 2.4.4 ASL and BOLD simultaneous measures

Simultaneous measurements of BOLD and ASL contrasts can provide valuable information on neural activity and neurovascular coupling. From a series of ASL images, perfusion and BOLD signals can be obtained by subtracting or adding the control and labeled images, respectively, in the same time series [Yang, 2002]. BOLD sensitivity can be increased by acquiring a second longer echo time (TE) in the ASL acquisition, with a longer echo having a TE closer to the T2* value of the tissue (e.g. 35 ms). These time series will be BOLD, but will have the resolution of ASL [Barker et al., 2013]. [Schmithorst et al., 2014] propose a simultaneous BOLD/ASL acquisition method with PCASL by adding one or several acquisitions of BOLD fMRI after the acquisition of the ASL volume. It decreases the temporal resolution of ASL: additional 1.3 s per BOLD acquisition added.

## 2.5 Physiological models to explain brain function

The study of the processes underlying brain function is an active field of research. With neuroimaging techniques, we get access to the living human brain with a relatively poor spatial and temporal resolution. These processes are better studied in animals with more invasive techniques. To better interpret the signals that we get in fMRI, a better understanding of what is happening and therefore of what we are actually measuring is important. Mathematical models have been proposed to model the neurovascular coupling and the oxygen consumption and BOLD effect. The Balloon model [Buxton et al., 1998b] proposes the idea that blood vessels work as a balloon that inflates or desinflates when cerebral blood flow changes as a function of brain activity. Cerebral blood flow and balloon volume changes cause changes in deoxyhemoglobin concentration in blood, that are reflected in the BOLD signal. Although this model is not exact and its validity has been discussed [Drew et al., 2011, Lorthois et al., 2011], it models quite well the response shapes of biophysical parameters. An evaluation can be found in [Blanchard et al., 2011].

The Balloon model does not explain the neurovascular coupling: the process from neural activity to CBF changes. Several models have been proposed [Huneau et al., 2015] to explain the neurovascular coupling and have been validated on BOLD data:

*Friston flow model [Friston et al., 2000]* proposes a neurovascular coupling model to explain how neural activity produces variations in cerebral blood flow. It considers that arteriolar activity is generated by neural responses and that it auto-regulates. Together with the Balloon model, it explains the process from neural activity to BOLD signal change. This model is simple and reflects quite well response shapes, but it does not perfectly explain the physiology behind the neurovascular coupling.

*Buxton flow model [Buxton et al., 2004]* proposes a neurovascular and neurometabolic coupling model assuming a linear convolution of neural activity with a flow response function: a gamma-variate function depending on normalized amplitude, duration and order. Neural activity is modelled linearly linking excitatory and inhibitory activity elicited by stimulation, and nonlinearly considering the existence of a neural baseline activity and a positive neural response. This model is consistent with experimental results assuming nonlinear neural adaptation, but does not create realistic perfusion response time courses.

*Arteriolar compliance model [Behzadi and Liu, 2005]* proposes the inclusion of a neurovascular compartment in the Balloon model that models an arteriolar wall compliance before generating the resulting blood flow. It relates neural activity to a vasoactive signal, and this signal to changes in muscular compliance. This model can explain changes in HRF with aging and changes induced with carbon dioxide.

*Proximal integration model or arterial impulse model [Kim et al., 2013]* assumes that neural activity generates a flow-inducing command that first reaches the nearby capillaries and that propagates later to larger vessels. Although a global model for BOLD fitting has been presented and fitted BOLD signal well, no fitting of flow measures has been reported.

These and other models are compared and put into perspective in [Huneau et al., 2015]. Some are closer to physiology than others, and some are mainly descriptive and try to fit the response shapes, as it is in the case of Friston flow model [Friston et al., 2000]. For simplicity and because we are interested in modelling the CBF and HRF shapes as closely as possible, in this work we focus on this model for neurovascular coupling. We call extended Balloon model the model that combines both Friston flow and Balloon models to explain how neural activity is transformed in the BOLD effect.

### 2.5.1 The extended Balloon Model

The Balloon model was first proposed in [Buxton et al., 1998b] to link neuronal and vascular processes by considering the capillary as a balloon that dilates under the effect of blood flow variations. More specifically,

the model describes how, after some stimulation, the local blood flow $\boldsymbol{f}_{in}(t)$ increases and leads to the subsequent augmentation of the local capillary volume $\boldsymbol{\nu}(t)$. This incoming blood is strongly oxygenated but only part of the oxygen is consumed. It follows a local decrease of the deoxyhemoglobin concentration $\boldsymbol{\xi}(t)$ and therefore a BOLD signal variation. The Balloon model was then extended in [Friston et al., 2000] to include the effect of the neuronal activity $\boldsymbol{u}(t)$ on the variation of some auto-regulated flow inducing signal $\boldsymbol{\psi}(t)$ so as to eventually link neuronal to hemodynamic activity. The global physiological model corresponds then to a non-linear system with four state variables $\{\boldsymbol{\psi}, \boldsymbol{f}_{in}, \boldsymbol{\nu}, \boldsymbol{\xi}\}$ corresponding to normalized flow inducing signal, local blood flow, local capillary volume, and deoxyhemoglobin concentration. Their interactions over time are described by the following system of differential equations:

$$\begin{cases} \frac{d\boldsymbol{f}_{in}(t)}{dt} = \boldsymbol{\psi}(t) \\ \frac{d\boldsymbol{\psi}(t)}{dt} = \eta u(t) - \frac{\boldsymbol{\psi}(t)}{\tau_\psi} - \frac{\boldsymbol{f}_{in}(t)-1}{\tau_f} \\ \frac{d\boldsymbol{\xi}(t)}{dt} = \frac{1}{\tau_m}\left(\boldsymbol{f}_{in}(t)\frac{1-(1-E_0)^{1/\boldsymbol{f}_{in}(t)}}{E_0} - \boldsymbol{\xi}(t)\boldsymbol{\nu}(t)^{\frac{1}{\tilde{w}}-1}\right) \\ \frac{d\boldsymbol{\nu}(t)}{dt} = \frac{1}{\tau_m}\left(\boldsymbol{f}_{in}(t) - \boldsymbol{\nu}(t)^{\frac{1}{\tilde{w}}}\right) \end{cases} \tag{2.3}$$

with initial conditions $\boldsymbol{\psi}(0) = 0, \boldsymbol{f}_{in}(0) = \boldsymbol{\nu}(0) = \boldsymbol{\xi}(0) = 1$. Figure 2.18 depicts these relationships. Lower case notation is used for normalized functions by convention. The system depends on five hemodynamic parameters: $\tau_\psi$, $\tau_f$ and $\tau_m$ are time constants respectively for signal decay/elimination, auto-regulatory feedback from blood flow and mean transit time, $\tilde{w}$ reflects the ability of the vein to eject blood, and $E_0$ is the oxygen extraction fraction. Another parameter $\eta$ is the neuronal efficacy weighting term that models neuronal efficacy variability.

Once the solution of the previous system is found, [Buxton et al., 1998b] proposed the following expression that links the BOLD response $\boldsymbol{h}(t)$ to the physiological quantities considering intra-vascular and extra-vascular components:

$$\boldsymbol{h}(t) = V_0[k_1(1-\boldsymbol{\xi}(t)) + k_2(1-\frac{\boldsymbol{\xi}(t)}{\boldsymbol{\nu}(t)}) + k_3(1-\boldsymbol{\nu}(t))] \tag{2.4}$$

where $k_1$, $k_2$ and $k_3$ are scanner-dependent constants and $V_0$ is the resting blood volume fraction. According to [Buxton et al., 1998b], $k_1 \cong 7E_0$, $k_2 \cong 2$ and $k_3 \cong 2E_0 - 0.2$ at a field strength of 1.5T and echo time $TE =$ 40ms. However, the values of $k_1$, $k_2$ and $k_3$ have been posteriorly revised [Obata et al., 2004] (see chapter 7). The perfusion response function (PRF), namely $\boldsymbol{g}(t)$, is the percentage perfusion signal change: $\boldsymbol{g}(t) = \boldsymbol{f}_{in} - 1$, $\boldsymbol{f}_{in}$ being the normalized perfusion with initial value 1.

[Friston et al., 2000] proposed this set of physiological parameters used:

$$V_0 = 0.02 \quad \tau_\psi = 1.25 \quad \tau_f = 2.5 \quad \tau_m = 1 \quad \tilde{w} = 0.2 \quad E_0 = 0.8 \quad \eta = 0.5.$$

The BOLD response function (HRF, $\boldsymbol{h}$) and perfusion response function (PRF, $\boldsymbol{g}$) generated using these parameters with the physiological model are shown in Fig. 2.19 under the label "Friston 00" (dashed line). The rest of the curves show the effect of changing the physiological parameters:

- $\eta$ is a scaling factor and causes non-linearities above a certain value.

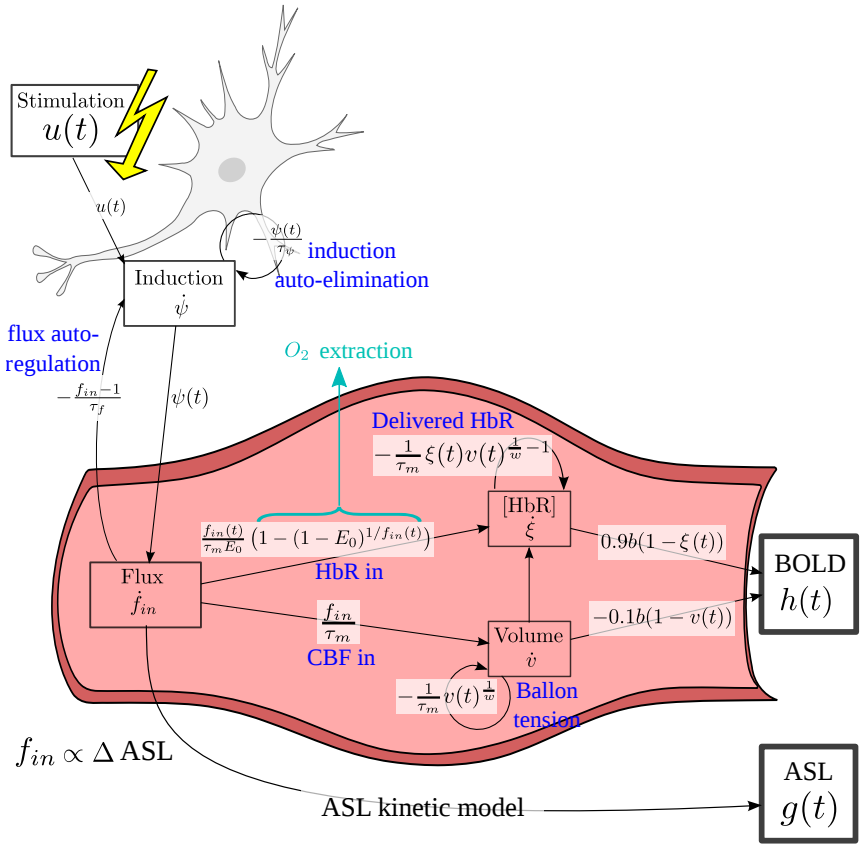



Figure 2.18: Extended Balloon model.

- $\tau_\psi$ controls the signal decay, which is more or less smooth.
- $\tau_f$ is the auto-regulatory feedback and it regulates the undershoot.
- $\tau_m$ is the transit time and it expands or contracts the signal in time.
- $\tilde{w}$ is the windkessel parameter and it models the initial dip and the response magnitude.
- $E_0$ is the oxygen extraction and it impacts the response scale.

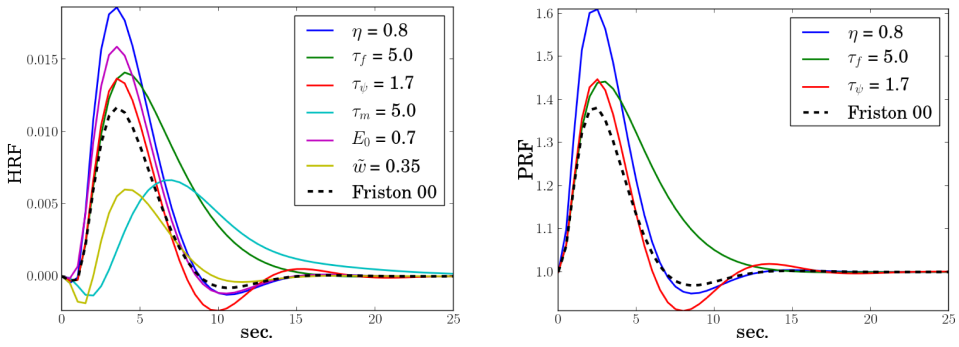



Figure 2.19: Effect of the physiological parameters on the HRF and PRF shapes. The parameters values proposed in [Friston et al., 2000] are used except for one parameter modified as indicated in the plot.

After analysing the behaviour of the model when varying the parameters values, the impact of each parameter was investigated and we concluded that the values proposed in [Friston et al., 2000] seemed reasonable. Several parameter settings have been proposed in the literature [Friston et al., 2000, Khalidov et al., 2011, Havlicek et al., 2011].

In this chapter, we introduced functional and perfusion MRI: What physiological processes we are actually measuring, BOLD and ASL functional MRI modalities put into perspective with other existing neuroimaging techniques, and mathematical models that try to explain the physiological processes underlying brain function and BOLD and blood flow measures. From this point, we will focus on the analysis of BOLD and ASL data.

# 3 Classical methods in fMRI data analysis

FMRI has been used since the 90s after the ground-breaking discovery of the BOLD contrast in [Ogawa et al., 1990]. Over the last two decades, the methods to analyse it have evolved. The gold standard for fMRI analysis is the General Linear Model (GLM) assuming a canonical hemodynamic response function (HRF), and its posterior subject and group level statistical analyses to infer brain activity.

In this chapter, we first introduce the classical methods to analyse BOLD fMRI: preprocessing pipeline, the general linear model, statistical inference, and the modelling of the hemodynamic response function. Then, we introduce the classical methods to analyse functional Arterial Spin Labeling (fASL) data. The statistical methods used for fASL are similar to the ones used for BOLD. The preprocessing pipeline and the design matrix change to deal with the perfusion component of fASL.

## 3.1 Preprocessing BOLD fMRI data

FMRI signal is noisy and has to be preprocessed before performing statistical analysis. There is a consensus in the community about the preprocessing pipeline that one has to apply to BOLD fMRI data before further analysis. A leading software package for data preprocessing is SPM[1] (Statistical Parametric Mapping) [Penny et al., 2011]. The recommended preprocessing steps are:

*Slice timing correction* corrects the effect of the timing difference between slices during the acquisition, since they are not acquired simultaneously. Between the first and the last slices acquired, there is a difference of almost a repetition time (TR). It shifts the signal phase to temporally align all acquired slices to a reference slice, which is usually taken to be the one in the middle of the TR interval and this depends on the slice acquisition order. The signal needs to be shifted accordingly when this is done ($\sim$TR/2). See figure 3.1.

*Realignment* corrects for head motion during the acquisition. Realignment and slice timing correction can be done simultaneously as in [Roche, 2011].

*Mean functional image.* The temporal mean of the BOLD time series is computed to obtain a mean functional image.

[1] http://www.fil.ion.ucl.ac.uk/spm/

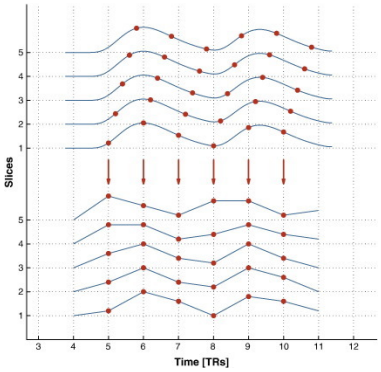



Figure 3.1: Slice timing correction: signal phase is shifted to temporally align all acquired slices to a reference slice. Taken from [Sladky et al., 2011].

*Coregistration.* We align spatially the anatomical image to the mean functional image so they are in the same space. It also reduces the resolution to the one of the target image.

*Segmentation.* The anatomical image is segmented into white matter, gray matter and cerebrospinal fluid probability maps. For each voxel, they indicate the probability of it being part of these types of tissue. We can compute a gray matter mask from those maps.

*Normalization.* A spatial transformation is done from the subject space to a template space. This is important if we want to compare different subjects in the same space (e.g. MNI template).

*Spatial smoothing using a Gaussian filter.* At the subject level, it reduces the effect of the noise and it considers a spatial correlation between voxels. At the group level it reduces coregistration and normalization errors. It also increases the group statistical effect when there is a high variability. However, the smoothing degrades the statistical specificity since it changes the sensitivity-specificity trade-off.

Spatial smoothing can help models that otherwise do not take into account spatial correlation of the signal. If spatial correlation is taken into account in the signal model, then not using smoothed data is better because one would be just losing sensitivity to effects. Ideally, one should take into account spatial correlation in the model. However, these methods tend to be computationally demanding.

Undesired low frequency effects due to scanner drifts or physiology related artefacts are considered in the posterior analysis of the data.

## 3.2 Statistical analysis of BOLD fMRI

The General Linear Model (GLM) has been classically used for statistical analysis of BOLD fMRI. It assumes that the acquired signal in a voxel is the sum of the activity evoked by a number of independent processes. It is based on the following hypotheses:

- Linearity: the length of the response is linearly dependent on the stimulus length.
- Additivity: The effects of the stimulus are added.
- Temporal stationarity: A stimulus always induces the same response.
- Spatial stationarity: the response generated by a stimulus is the same through different voxels and experimental conditions. One usually chooses a "canonical hemodynamic response function" (HRF), which can e.g. be generated as the difference of two Gamma functions [Glover, 1999].

The GLM framework models the BOLD signal with a set of functions called regressors, and calculates their weights using linear regression. We describe, at each voxel, the parameters associated with each regressor that fit in the least squared sense. For every voxel $j$:

$$\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{b}_j \tag{3.1}$$

where $\boldsymbol{y}_j$ is the BOLD signal, $\boldsymbol{X}$ the design matrix (containing the regressors as columns) and $\boldsymbol{\beta}_j$ the unknown weights of the regressors. $\boldsymbol{b}_j$ is the noise in voxel $j$, which is assumed i.i.d. Gaussian if we assume independence or serially correlated, typically following a first order autoregressive (AR) model with a one-timepoint history [Woolrich et al., 2001]. Note that longer history AR models have also been tested to fit the noise component [Harrison et al., 2003].

Each functional regressor from the matrix $\boldsymbol{X}$ (see Fig. 3.2) defines temporal effects that can be observed in the fMRI data: task-related activity, nuisance events, low-frequency signals... The regressors defining task-related activity (figure 3.3) are constructed as the convolution of the hemodynamic response function with the stimulation signal that contains the event stimulus onsets in the paradigm. The regressors defining nuisance events and low-frequency signals are called drift regressors. They account for low frequency ($< 1/120\ Hz$) drifts that fMRI contains: scanner drifts, cardiac/respiratory artefacts and residual movement effects and their interaction with the static magnetic field [Frackowiak et al., 2004] . Drift regressors are orthogonal and they are usually implemented as polynomial or cosine functions. Typical values for polynomial drifts are 5 polynomial functions up to the 5th degree, often normalized to lie within the range of [0, 1] over the full time course. For cosine drifts, the cut-off frequency of $1/128\ Hz$ is used.

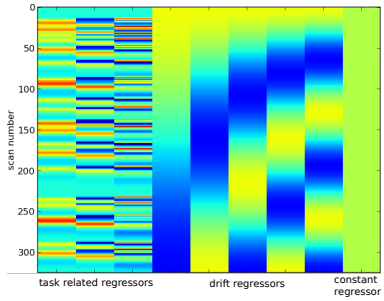



Figure 3.2: Design matrix of a GLM approach using canonical HRF and its first and second derivatives. It is an event-related design with just one experimental condition. The first three columns represent the regressors associated with $\boldsymbol{h}_{can}$, $\boldsymbol{h}'_{can}$ and $\boldsymbol{h}''_{can}$. It is followed by five polynomial drifts and a constant to capture the signal mean.

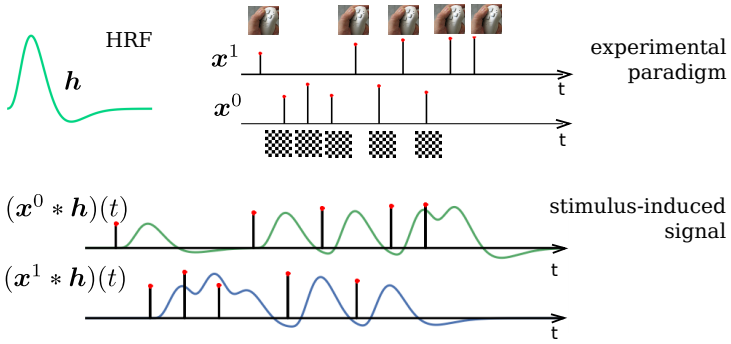



Figure 3.3: The task-related regressor is the convolution of the stimulation signal by a canonical HRF: canonical HRF $\star$ experimental paradigm = stimulus-induced signal. In this figure, the paradigm shown is event-related, with two conditions corresponding to visual and motor tasks.

### 3.2.1 Univariate estimation

The regressor parameters $\hat{\boldsymbol{\beta}}$ are the estimated magnitude of activation for each condition described in columns of $\boldsymbol{X}$. They are calculated by minimizing the squared distance between vector $\boldsymbol{X}\boldsymbol{\beta}_j$ and the measurement timecourse $\boldsymbol{y}_j$. If we assume i.i.d Gaussian noise $\boldsymbol{b}_j \sim \mathcal{N}(0, \sigma_j^2 \boldsymbol{I}_N)$, the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}_j$ reads:

$$\hat{\boldsymbol{\beta}}_j = \boldsymbol{X}^+ \boldsymbol{y}_j, \tag{3.2}$$

where $\boldsymbol{X}^+$ is the Moore-Penrose pseudoinverse. In classical statistics the design matrix $\boldsymbol{X}$ is assumed injective (it maps one to one), making $\boldsymbol{X}^t\boldsymbol{X}$ invertible. It follows that we can write $\boldsymbol{X}^+ = (\boldsymbol{X}^\mathrm{t}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathrm{t}$. The distribution of $\hat{\boldsymbol{\beta}}$ can be written as:

$$\hat{\boldsymbol{\beta}}_j \sim \mathcal{N}(\boldsymbol{\beta}_j, \sigma_j^2 (\boldsymbol{X}^\mathrm{t}\boldsymbol{X})^{-1}) \tag{3.3}$$

where $\sigma_j^2$ is the unknown variance of the error term $\boldsymbol{b}_j$. The estimator $\hat{\boldsymbol{\beta}}_j$ is unbiased since the noise $\boldsymbol{b}_j$ is assumed Gaussian and with zero mean. From the minimization of the residuals $\boldsymbol{r}_j = \boldsymbol{y}_j - \boldsymbol{X}\hat{\boldsymbol{\beta}}_j$ we can compute the noise variance MLE:

$$\hat{\sigma}_j^2 = \frac{\|\boldsymbol{r}_j\|^2}{N} \tag{3.4}$$

Note here that residuals can also be written as $\boldsymbol{r}_j = (\boldsymbol{I}_N - \boldsymbol{X}\boldsymbol{X}^+)\boldsymbol{b}_j$. If we take the expectation:

$$\begin{aligned}\mathrm{E}\big[\hat{\sigma}_j^2\big] &= \frac{\mathrm{E}\big[\|\boldsymbol{r}_j\|^2\big]}{N} = \frac{1}{N}\mathrm{E}\big[\mathrm{tr}\,(\boldsymbol{b}_j^t(\boldsymbol{I}_N - \boldsymbol{X}\boldsymbol{X}^+)\boldsymbol{b}_j)\big] = \frac{\sigma_j^2}{N}\mathrm{tr}\,(\boldsymbol{I}_N - \boldsymbol{X}\boldsymbol{X}^+) \\ &= \frac{\sigma_j^2}{N}(\mathrm{tr}\,(\boldsymbol{I}_N) - \mathrm{tr}\,(\boldsymbol{I}_R)) = \frac{N-R}{N}\sigma_j^2 \end{aligned} \tag{3.5}$$

where $R$ is the rank of $\boldsymbol{X} \in \mathbb{R}^{N\times R}$. The unbiased estimator is then:

$$\hat{\sigma}_j^2 = \frac{\|\boldsymbol{r}_j\|^2}{N-R}. \tag{3.6}$$

### 3.2.2 Statistical testing

Statistical tests check the significance of a given contrast $\boldsymbol{\gamma}^t\boldsymbol{\beta}_j$, a linear combination of the effects associated with the experimental conditions. From the last section, we can see that $\boldsymbol{\gamma}^t\hat{\boldsymbol{\beta}}_j = \boldsymbol{\gamma}^t\boldsymbol{\beta}_j + \boldsymbol{\gamma}^t\boldsymbol{X}^+\boldsymbol{b}_j$ and follows the law $\boldsymbol{\gamma}^t\hat{\boldsymbol{\beta}}_j \sim \mathcal{N}(\boldsymbol{\gamma}^t\boldsymbol{\beta}_j, \sigma_j^2\boldsymbol{\gamma}^t(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{\gamma})$. A standard measure of activation is the z-score of the normal distribution, which tells us how many standard deviations we are from a zero mean. It reads $z = \frac{\boldsymbol{\gamma}^t\hat{\boldsymbol{\beta}}_j}{\sigma_j\sqrt{\boldsymbol{\gamma}^t(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{\gamma}}}$. However, note that the denominator contains $\sigma_j$, the unknown noise variance of the voxel. This makes the pure z-score impossible to estimate. Using the unbiased estimator of the variance $\hat{\sigma}^2$ instead, we obtain a quantity that is t-distributed (due to the unknown variance) instead of Gaussian, which we take into account when making inference by e.g. calculating p-values. T-test or the F-test are common when comparing models. These tests consist in determining what outcomes from a set of random variables $\boldsymbol{\gamma}^{\mathrm{t}}\boldsymbol{\beta}_j$ would lead to a rejection of the null hypothesis for a pre-specified level of significance which is a probability threshold.

#### T-test

A t-test is a statistical hypothesis test in which the test statistic follows a Student t distribution (see figure 3.4) under the null hypothesis. We want to know how different the estimated parameters are to a certain value. The null hypothesis $\boldsymbol{H}_0$ can be defined in our case, for voxel $j$, as:

$$\boldsymbol{H}_0 : \boldsymbol{\gamma}^t\boldsymbol{\beta}_j = 0. \tag{3.7}$$

We could also test alternative hypothesis as $\forall j, \boldsymbol{H}_1 : \boldsymbol{\gamma}^t\boldsymbol{\beta}_j \neq 0$.

As $\boldsymbol{\gamma}^t\hat{\boldsymbol{\beta}}_j$ is a linear combination of random variables following normal laws of unknown variances, we can use a Student law with $N-R$ ($N$ scans, $R$ regressors) degrees of freedom:

$$T_j = \frac{\boldsymbol{\gamma}^t\hat{\boldsymbol{\beta}}_j}{\hat{\sigma}_j\sqrt{\boldsymbol{\gamma}^t(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{\gamma}}} \sim t_{N-R} \tag{3.8}$$

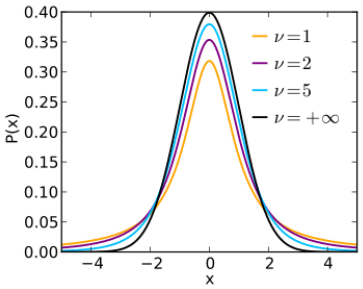



Figure 3.4: T student distribution:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$\Gamma$ being the Gamma function.

en.wikipedia.org

We need to estimate the noise variance $\hat{\sigma}_j$ as previously explained. For every voxel $j$, the statistic is evaluated and we get $\boldsymbol{t} = (t_j)_{j=1:J}$ observed realisations of $T$. We compute the p-value in each voxel as: $\alpha = P(T > u_\alpha | \boldsymbol{H}_0)$, where $u_\alpha$ is the statistical value associated with a level of significance $\alpha$, usually considered 0.05. At each voxel, a value $t_j$ is observed and rejects the null hypothesis if $t_j > u_\alpha$. The *p-value* summarizes the evidence against $H_0$ and corresponds to the probability of observing an extreme value greater than the observed $t_j$, under $H_0$. If the *p-value* is lower than $\alpha$, we reject $H_0$ and the activation of the voxel is considered significant.

### F-test

An F-test is based on Fisher statistic and it allows multiple hypothesis testing. In our case, we can test whether a set of contrasts are all simultaneously different from zero with a null hypothesis:

$$\boldsymbol{H}_0 : \boldsymbol{C}^t \boldsymbol{\beta}_j = 0 \tag{3.9}$$

$\boldsymbol{C}$ being a contrast matrix with k columns. Here we will know that one of the contrast is significant, but we will not know which one.

Under the null hypothesis, the F-statistic follows a Fisher distribution (see figure 3.5):

$$F_j = \frac{\boldsymbol{C}^t \hat{\boldsymbol{\beta}}_j}{k\hat{\sigma}_j \sqrt{\boldsymbol{C}^t (\boldsymbol{X}^t \boldsymbol{X})^{-1} \boldsymbol{C}}} \sim F_{k,N-R} \tag{3.10}$$

This test is useful in the case of modelling a condition with several regressors corresponding to a basis function as it could be the HRF with derivatives, since we can group several regressors in one test and see the effect of a stimulus.

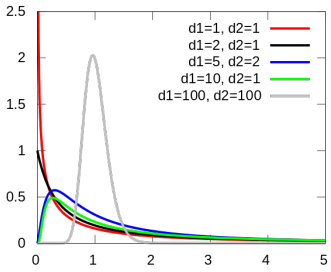



Figure 3.5: F-distribution: $f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1} \, d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x \, \mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$ B being the Beta function.

en.wikipedia.org

### Correction for multiple comparisons

Applying the previously explained statistical tests independently we get brain maps (statistical parametric maps or SPM). However, with the voxel-wise statistics we accept a certain amount of error: $J\alpha$ false positives. That is quite high when we accumulate it through all voxels: if $J$ is $\sim$ 50000, with $\alpha = 0.001$ we have 50 false positives and with $\alpha = 0.05$ we have 2500 false positives. To correct for this effect, we need to consider a global strategy. Two strategies are generally used:

*Family-Wise Error Rate (FWER) control.* FWER is the probability of considering active voxels that are truly inactive. We can apply different corrections:

*Bonferroni correction* assumes voxels independent and corrects the level of activation significance from $\alpha$ to $\alpha/J$. This makes the correction quite rigid, since it does not account for spatial correlation that certainly exists in the brain, and that we also include during data preprocessing when realigning, normalizing or smoothing.

*Random field theory* considers that a smooth statistical map has a lower probability of exceeding a threshold by chance. It is based on

the topological characterization of the statistical map using Euler's characteristic. See [Worsley et al., 1992] for details.

*Resampling methods* as permutation [Holmes et al., 1996] are accurate but computationally demanding. See [Nichols and Holmes, 2002] for details.

A comparative review can be found in [Genovese et al., 2002].

*False Discovery Rate (FDR) control.* FDR is the expected proportion of false positives in voxels considered active. We can correct it by using the Benjamini-Hocheberg procedure [Benjamini and Hochberg, 1995], that consists of an adaptative Bonferroni correction that applies a different correction depending on the significance of the voxel. See also [Genovese et al., 2002].

We can also apply cluster-level correction to get rid of small and unlikely activations. In this case, the clusters (set of active neighbouring voxels) of small size are not considered activated. See [Frackowiak et al., 2004, Nichols and Hayasaka, 2003, Genovese et al., 2002] for further information.

### Group level statistics

Group statistics seek to find significant consistent activations within a group of $S$ subjects with $J$ observations per subject. To perform these statistics, we compute the variability of a population considering fixed or random effects (e.g. effect maps). See figure 3.6.

*Fixed effects analysis* takes into account within-subject variation. It supposes that the $s$th subject mean true effect $w_s$ is fixed and that variation comes from measurement errors (within-subject errors), which are assumed random. The $j$th observed effect in subject $s$ $y_{s,j}$ is assumed to be:

$$y_{s,j} = w_s + e_{s,j}.$$

Within-subject error $e_{s,j}$ is assumed Gaussian with zero mean and variance $\sigma_w^2$. We can obtain the effect distribution by estimating the average effect size

$$\hat{w}_{pop,FFX} = \frac{1}{SJ} \sum_{s=1}^{S} \sum_{j=1}^{J} y_{s,j}$$

and its variance

$$\sigma^2_{FFX} = \frac{\sigma_w^2}{SJ}.$$

Since it does not take into account between-subject variability, fixed effects analysis is specific to the group under study and we can not infer anything about a wider population.

In the context of a General Linear Model, they can be computed by concatenating the data of different subjects and creating a block design matrix with the design matrices of each subject. The effects of interest can then be examined using an augmented contrast vector that considers all subject contrasts [Penny et al., 2011].

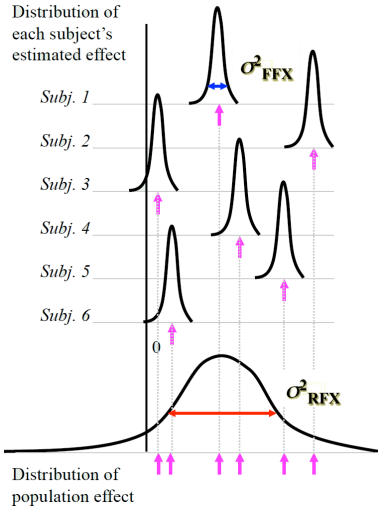



Figure 3.6: Fixed and random effects in group studies.

Taken from slides "Group Modeling for fMRI data" by T. Nichols

*Random Effects analysis* [Penny et al., 2011] considers between-subject variation. It considers that the sth subject true mean effect $w_s$ is also a source of variation, but that the population mean $w_{pop}$ is fixed. Therefore, we have two sources of variability: the between-subject error $z_s$, and the measurement or within-subject error $e_{s,j}$. The observed effect $y_{s,j}$ is then assumed to be a function of:

$$y_{s,j} = w_s + e_{s,j} = w_{pop} + z_s + e_{s,j}.$$

Within-subject error $e_{s,j}$ is assumed Gaussian with zero mean and variance $\sigma_w^2$, which is assumed equal across subjects. Between-subject error $z_s$ is assumed Gaussian too, with zero mean and variance $\sigma_b^2$. Therefore, subject effects $w_s$ are normally distributed around the average population effect $w_{pop}$, with a certain variance $\sigma_b^2$. Here we can generalise the results to the population of interest.

In practice, we can do this analysis through a two-level model. In the first level, we compute subject level statistics and define the effect of interest for each subject with a contrast vector to obtain contrast effect maps. In this level we are considering the variation of the subject samples around the true subject mean: $\bar{w}_s = w_s + e_{s,j}$, with variance $\sigma_w^2/J$.

In the second level, we perform a GLM that implements a one-sample t-test [Penny et al., 2011] with the contrast images of all subjects. We consider the subject mean variation around the population mean: $w_s = w_{pop} + z_s$, with variance $\sigma_b^2$. The population mean is then estimated as

$$\hat{w}_{pop,RFX} = \frac{1}{S}\sum_{s=1}^{S} \bar{w}_s,$$

with mean $E[\hat{w}_{pop,RFX}] = w_{pop}$ and variance

$$\sigma_{RFX}^2 = Var[\hat{w}_{pop,RFX}] = \frac{\sigma_b^2}{S} + \frac{\sigma_w^2}{SJ}.$$

As shown in [Penny et al., 2011], this is equivalent to computing Maximum Likelihood estimates.

Mixed effects analysis combine both fixed and random effects analysis. They account for the within-subject uncertainties, as represented in particular by the estimated variances of the effect estimates, in the analysis. It therefore corrects the group statistical map for higher- order information [Roche et al., 2007].

In neuroimaging, random effect analysis are widely used in the group level. For further detail, see [Penny et al., 2011].

## 3.3 Modelling the hemodynamic response function

The estimation of the hemodynamic response function (HRF) is interesting from a cognitive and clinical point of view, as it changes depending on the area of the brain, and between different subjects, ages and with pathology. This could give valuable information to perform the comparison between

populations (pathological and healthy) or to test the effects of drugs in brain function and vascularization. The HRF can be described by physiological models as the Balloon model, by models not considering physiology, or inferred from the data.

### 3.3.1 From physiological models

We can generate the HRF using physiological models like the Balloon model, already introduced in the previous chapter (see figure 3.7). In this case, we use the set of differential equations to generate the response function. This model uses physiological parameters that have to be measured. Different sets of parameters have been proposed in the literature. We can also generate other physiological responses like blood flow, volume or deoxyhemoglobin concentration.

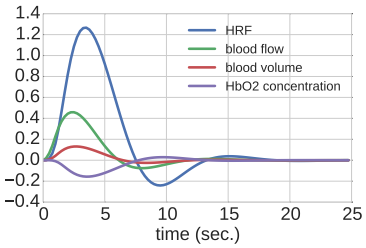



Figure 3.7: Hemodynamic response function and other physiological responses generated with the Balloon model using parameters proposed in [Khalidov et al., 2011].

### 3.3.2 Using the canonical HRF

One can consider a fixed HRF with a canonical shape ($\boldsymbol{h}_{can}$) generated as the difference of two Gamma functions [Glover, 1999] (see figure 3.8) and only modulate the amplitude:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \beta_{0,j}^m (\boldsymbol{h}_{can} * \boldsymbol{x}^m) + \boldsymbol{b}_j \tag{3.11}$$

where $\boldsymbol{y}_j$ is the signal in voxel $j$, $\boldsymbol{x}^m$ is the regressor vector for condition $m$, and $\beta_{0,j}^m$ is the weight of the regressor, corresponding to the amplitude modulation of the HRF.

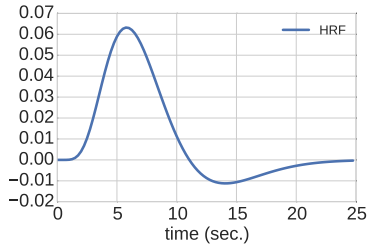



Figure 3.8: Hemodynamic response function generated with the difference of 2 Gamma functions.

#### Introducing function basis

We can use function basis to model HRF variations. One possibility is adding regressors for the temporal derivative and the derivative with respect to the dispersion parameters of the canonical HRF (see figure 3.9). With them, we model delay and dispersion of the response, respectively.

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \beta_{0,j}^m (\boldsymbol{h}_{can} * \boldsymbol{x}^m) + \beta_{1,j}^m (\boldsymbol{h}'_{can} * \boldsymbol{x}^m) + \beta_{2,j}^m (\boldsymbol{h}''_{can} * \boldsymbol{x}^m) + \boldsymbol{b}_j \tag{3.12}$$

where $\beta_{0,j}^m$, $\beta_{1,j}^m$ and $\beta_{2,j}^m$ are the weights of the regressors corresponding to the canonical HRF $\boldsymbol{h}_{can}$, and its first $\boldsymbol{h}'_{can}$ and second $\boldsymbol{h}''_{can}$ derivatives. We can generalize $\boldsymbol{h}_{can}$ as a function basis $(\boldsymbol{h}_c)_{c=1:C}$, where $c$ is every stimulus induced component. In this case, $\boldsymbol{y}_j$:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \sum_{c=0}^{C} (\boldsymbol{\beta}_{c,j}^m (\boldsymbol{h}_c * \boldsymbol{x}^m)) + \boldsymbol{b}_j, \tag{3.13}$$

This method adds some flexibility to the model, but also complexity when inferring activation detection. It requires the use of F-tests to group the activation detected by different regressors.

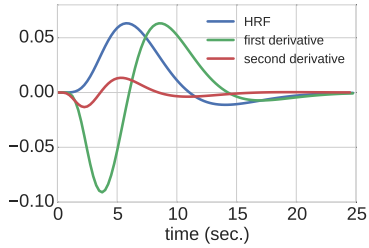



Figure 3.9: Hemodynamic response function and its first and second derivatives.

#### Non-parametric approaches

Here the HRF is entirely estimated each time. One can use Finite Impulse Response (FIR) modelling or a regularized version of it:

*Finite Impulse Response (FIR)* allows to capture the HRF shape by estimating a finite set of coefficients. The signal is modelled as:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \sum_{d=0}^{D} h_{j,d} x_{t-d}^{m} + \boldsymbol{\beta}_{j,0} + \boldsymbol{b}_j \tag{3.14}$$

considering M conditions and with D+1 order FIR filter. $h_{j,d}$ is the HRF value with delay $d$ in voxel $j$, and $x_{t-d}^{m}$ is the regressor matrix for condition $m$ and delay $t-d$. $\boldsymbol{\beta}_{j,0}$ is the effect associated with the constant regressor and captures the mean signal. See figure 3.10.

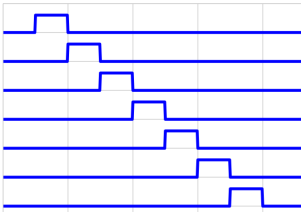


Figure 3.10: In finite impulse response each coefficient of the HRF is captured with a delayed pulse.

*Regularized Finite Impulse Response* embeds a temporal regularization on the HRF shapes. In this method, the HRF estimation is a trade-off between information contained in the data and in a prior, modelled with a Bayesian formalism. Since we know that the hemodynamic response function is temporally smooth, we can introduce a prior to account for this temporal regularization (see figure 3.11):

$$\boldsymbol{h}_j \sim \mathcal{N}(0, v_h \boldsymbol{R}) \text{ and } \boldsymbol{R} = \left( \frac{\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2}{(\Delta t)^4} \right)^{-1} \tag{3.15}$$

Here $\boldsymbol{R}$ is a covariance matrix that provides smoothness to the function. It introduces a constraint on the second order derivative to penalise high variations between neighbouring time points, and therefore recover a smooth shape. $\Delta t$ is the time step of $\boldsymbol{h}$ and $\boldsymbol{D}_2$ the 2nd order finite difference matrix[2]. $\boldsymbol{D}_2$ and $\boldsymbol{R}^{-1}$ have a dimension of $D \times D$ and read:

$$\boldsymbol{D}_2 = \begin{pmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 1 & -2 & 1 \\ 0 & \dots & 0 & 1 & -2 \end{pmatrix} \tag{3.16}$$

$$\boldsymbol{R}^{-1} = \frac{1}{(\Delta t)^4} \begin{pmatrix} 5 & -4 & 1 & 0 & \dots & \dots. & 0 \\ -4 & 6 & -4 & 1 & 0 & \dots & 0 \\ 1 & -4 & 6 & -4 & 1 & \dots & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & \dots & 0 & 1 & -4 & 6 & -4 \\ 0 & \dots & \dots & 0 & 1 & -4 & 5 \end{pmatrix} \tag{3.17}$$

Then, to compute the HRF:

$$\boldsymbol{h}_j^{FIR} = \left( \sum_m (\boldsymbol{X}^m)^{\mathrm{t}} \boldsymbol{X}^m + v_h \boldsymbol{R}^{-1} \right)^{-1} \sum_m (\boldsymbol{X}^m)^{\mathrm{t}} \boldsymbol{y}_j \tag{3.18}$$

See for instance [P. Ciuciu et al., 2003]

[2] The second derivative of $\boldsymbol{h}$ is:

$$\begin{aligned} \frac{\mathrm{d}^2 \boldsymbol{h}}{\mathrm{d}t^2}(t) &= \lim_{\Delta t \to 0} \frac{h_{(d+1)\Delta t} - 2h_{d\Delta t} + h_{(d-1)\Delta t}}{(\Delta t)^2} \\ &\approx \frac{\boldsymbol{D}_2 \boldsymbol{h}}{(\Delta t)^2} \\ &\text{with } d = 1 : D \end{aligned}$$

and if we do the square norm we get the covariance matrix $\boldsymbol{R}$:

$$\left\| \frac{\mathrm{d}^2 \boldsymbol{h}}{\mathrm{d}t^2} \right\|^2 = \boldsymbol{h}^{\mathrm{t}} \left( \frac{\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2}{(\Delta t)^4} \right) \boldsymbol{h}.$$

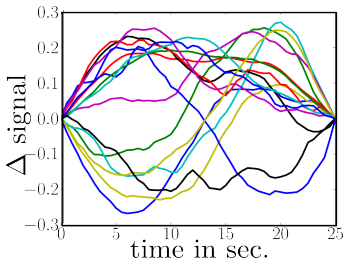



Figure 3.11: Prior distribution of the HRF: Gaussian distribution. The first and last points are 0.

Until this point, this chapter has been dedicated to the introduction to classical methods for the analysis of BOLD fMRI. From here, we dedicate some pages to the existing methods for the analysis of functional ASL data. Some of the preprocessing and methods are similar to the ones explained for BOLD and they will not be explained again to avoid redundancy.

## 3.4 Preprocessing functional ASL data

Due to the acquisition procedure and the structure of the data, ASL can not be preprocessed as BOLD. An example of preprocessing pipeline can be found in [Wang et al., 2008]. In the context of this thesis, the *process-asl* toolbox[3] has been used for the analysis of ASL data. The pipeline reads as follows:

*Scale factor correction.* Because of the T1 decay, signal will be lower depending on the exact time of the slice acquisition time. Scale factor correction allows us to have a nominal perfusion weighting. It is done before spatial preprocessing because then voxels move. One needs to know the order of the slices, all the timings of the acquisition sequence, and the delay of one slice acquisition. It makes the values closer to quantified values. The scale factor that we apply is $\frac{1}{2\alpha TI_1} \exp\left(\frac{TI_2}{T_{1,blood}}\right)$, where $\alpha = 0.98$ is the labeling efficiency; $TI_1$ the time when tagging stops; $TI_2$ the time when the image is acquired; $T_{1,blood} = 1650$ ms is the longitudinal relaxation time at 3T and depends on the slice. See section 2.4.2 for more details.

*Realignment.* This is an important step in ASL because its analysis is based on the subtraction of control/tag. We realign to the first or to the mean volumes.

*Mean ASL image.* We do the temporal mean of the ASL time series.

*Coregistration.* Register high resolution anatomical to low resolution mean ASL image.

*Segmentation.* The anatomical image is segmented into white matter, gray matter and cerebrospinal fluid images.

*Normalization.* We transform to a template space (normally MNI) to be able to compare different subjects in the same space. It is usually better to work at the subject space, since normalization can include errors in the processing.

*Spatial smoothing using a Gaussian filter.* At a subject level, it reduces the effect of the noise and it considers spatial correlation between voxels. At a group level, it minimizes coregistration and normalization errors.

As in BOLD, if spatial correlation between voxels is taken into account in the model, we do not need to apply smoothing before. We will also deal with low frequency drifts in the model. Note here that no slice timing correction is performed. The tagging is the part that takes more time in ASL sequences, and to compensate the actual acquisition time is much lower than a TR ($\sim 0.5$ s). A slice timing correction would not have a huge impact, and it could add noise to the signal.

## 3.5 Statistical analysis of fASL

Functional ASL can be analysed using the same tools as in BOLD, due to the nature of the signal. It has a BOLD or hemodynamic component and a

[3] process-asl https://github.com/process-asl is a python package developed by S. Bougacha under the supervision of P. Ciuciu at Neurospin (CEA Saclay), and reproducing the pipeline developed by J. Warnking at the Grenoble Institute of Neuroscience (GIN). The toolbox uses *nipype* http://nipy.org/nipype/0.10.0/ to create a preprocessing pipeline using SPM. It contains a CBF quantification module for pre- and post-processing of ASL fMRI data, following the recommendations in [Alsop et al., 2015]. CBF quantification has been done in collaboration with J. Warnking and A. Vignaud.

perfusion component. Several methods have been proposed to analyse ASL data: [Liu and Wong, 2005] proposes a signal processing model and several differencing approaches to extract the perfusion component of the signal by subtraction of control and tag images; [Woolrich and Behrens, 2006] proposes a Bayesian approach to make statistical inferences on significance of perfusion and T2* effects. The General Linear Model has also been used in the analysis of fASL data: [Mumford et al., 2006] proposes a GLM approach for ASL and compares this method to differencing approaches [Liu and Wong, 2005]; [Hernandez-Garcia et al., 2010] makes a quantitative analysis of ASL using GLM.

### 3.5.1 Differencing approaches

Differencing approaches consider the control-tag alternation in the time series and preprocesses the data to get a perfusion signal by subtracting volumes. The low temporal resolution of ASL poses a problem in this case. The resolution is already low, and by doing the difference we are dividing it by two. Moreover, control and tag volumes were not acquired at the same time, so they describe different states. Differencing can be done in different ways, and some approaches include interpolation in order to compensate for the low temporal resolution issue. Some differencing approaches can be found in [Liu and Wong, 2005, Mumford et al., 2006]. The difference has to be done in both sides of the equation

$$\tilde{\boldsymbol{D}}_x \boldsymbol{y}_j = \tilde{\boldsymbol{D}}_x \boldsymbol{X} \boldsymbol{\beta}_j + \tilde{\boldsymbol{D}}_x \boldsymbol{b}_j \tag{3.19}$$

where $\tilde{\boldsymbol{D}}_x$ specifies the way this difference is performed:

*Standard pairwise differencing* [4] has size $N/2 \times N$ and makes no interpolation. From two images we get a new one making the subtraction.

*Running subtraction* [5] has size $(N-1) \times N$ and interpolates values. Every image, except the first one, is subtracted to the previous image.

*Surround subtraction* [6] has size $(N-2) \times N$ and interpolates using 3 images. Every data point is computed doing the subtraction of a half of the previous and next images to the present image.

*Sinc subtraction* has size $N \times N$ and uses a sinc (see figure 3.12) to combine neighbouring images in a more elaborate interpolation. It is the differencing method that works the best.

The resulting perfusion time series can be then analysed as though they were BOLD time series. See [Mumford et al., 2006] for extended explanation. The main advantage of these methods is that the noise becomes whiter when the subtraction is done.

### 3.5.2 Non-differencing approaches: GLM

Non differencing approaches analyse the whole time series containing a BOLD effect and the perfusion signal. As in BOLD, GLM can be used to analyse fASL [Mumford et al., 2006, Hernandez-Garcia et al., 2010], but the design matrix has to be modified to account for the perfusion

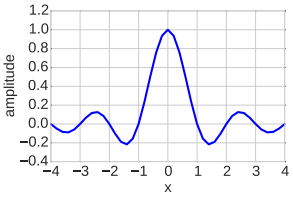



Figure 3.12: $sinc(x) = \frac{sin(x)}{x}$

[4] Pairwise differencing matrix

$$\tilde{\boldsymbol{D}}_1 = \begin{pmatrix} 1 & -1 & & & \dots & 0 \\ & & 1 & -1 & \dots & 0 \\ & & & & \ddots & \end{pmatrix}$$

[5] Running subtraction matrix

$$\tilde{\boldsymbol{D}}_2 = \begin{pmatrix} 1 & -1 & & \dots & 0 \\ & -1 & 1 & \dots & 0 \\ & & & \ddots & \end{pmatrix}$$

[6] Surround subtraction matrix

$$\tilde{\boldsymbol{D}}_3 = \frac{1}{2} \begin{pmatrix} 1 & -2 & 1 & & \dots & 0 \\ & 1 & -2 & 1 & \dots & 0 \\ & & & & \ddots & \end{pmatrix}$$

signal. The ASL regressors in the design matrix will be: baseline MRI, task-related BOLD describing the BOLD effect changes, baseline perfusion corresponding to the control/tag difference at rest, and task-related ASL describing the control/tag difference due to activation. See figure 3.13.

The signal model reads:

$$\boldsymbol{y}_j^{ASL} = \mathbf{X}^{ASL}\boldsymbol{\beta}_j + \boldsymbol{b}_j \tag{3.20}$$

where $\boldsymbol{y}_j^{ASL}$ is the acquired ASL signal, $\mathbf{X}^{ASL}$ the design matrix, $\boldsymbol{\beta}_j$ the weights of the regressors that we are going to estimate, and $\boldsymbol{b}_j$ an error vector.

In the task-related regressors, an HRF can be used for perfusion and BOLD regressors. However, the canonical HRF has been calibrated in BOLD experiments and reflects simultaneous variations of cerebral blood flow, blood volume and changes in cerebral oxygen consumption. It is well known that the perfusion function, reflecting just CBF variation, is a bit different than the hemodynamic response: the perfusion function peaks before and it has a smaller or non-existing undershoot. A modified response peaking at 4 s and without undershoot can be used to consider this in GLM.

From this point, the analysis of functional ASL data is the same as in BOLD. We estimate the effect for each regressor and we perform statistical tests for a subject or a group of subjects. For more details, see [Mumford et al., 2006, Hernandez-Garcia et al., 2010].

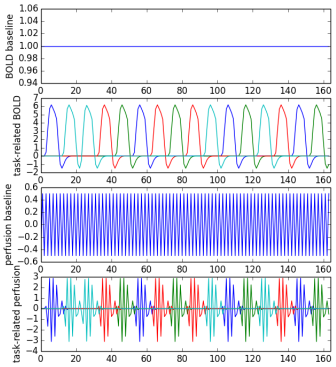



Figure 3.13: The regressors of the design matrix are baseline MRI, activation BOLD, baseline ASL and activation ASL. Different colors represent different experimental conditions. Reproduced from [Hernandez-Garcia et al., 2010].

# 4 Bayesian analysis of functional MRI data

In this chapter we introduce the main Bayesian tools used in our analysis of fMRI data. We present Markov Chain Monte Carlo (MCMC) and Variational Expectation Maximization techniques. The use of these optimization tools to the problems studied in this thesis is detailed in the next chapters.

## 4.1 A light introduction to Bayesian theory

Bayes' theorem describes the probability of an event based on conditions that might affect this event. These conditions are called priors because they are *a priori* information about the events. The probability of the event after considering these conditions is called posterior distribution.

Bayes' theorem combines the sum and product probability rules.

$$\begin{aligned} \text{Product Rule} \quad & p(y,\phi) = p(y|\phi)p(\phi) = p(\phi|y)p(y) \\ \text{Sum Rule} \quad & p(y) = \int_{\phi} p(y,\phi)d\phi \quad \text{which becomes} \\ & p(y) = \sum_{\phi} p(y,\phi) \quad \text{in the discrete case.} \end{aligned}$$

where $p(\phi)$ is the probability of $\phi$, $p(\phi, y)$ the joint probability of $\phi$ and $y$, and $p(\phi|y)$ the probability of $\phi$ conditional to $y$. Combining the sum and product rules, we obtain Bayes' rule: the posterior probability distribution $p(\phi|y)$ is the likelihood $p(y|\phi)$ multiplied by the prior distribution $p(\phi)$, over the evidence $p(y)$. This theorem allows us to go from the probability of observations given some parameters to the probability of the parameters given the observations.

$$p(\phi|y) = \frac{p(y|\phi)p(\phi)}{p(y)} = \frac{p(y|\phi)p(\phi)}{\int_{\phi} p(y|\phi)p(\phi)d\phi}.$$

We have expanded $p(y)$ into an integral over the joint probability $p(y, \phi)$, written as a product of conditional and prior probability. This shows that $p(y)$ acts like a normalizing constant, guaranteeing that the integral over $\phi$ of the right-hand side is 1. Omitting this constant, also called *evidence*, we can establish a proportionality relationship: the posterior distribution is proportional to the likelihood times the prior

$$p(\phi|y) \propto p(y|\phi)p(\phi)$$

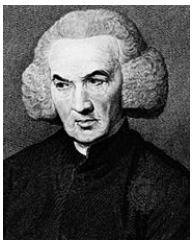


Figure 4.1: Thomas Bayes (1701-1761), in the upper picture, wrote what later became the Bayes' theorem as part of a wider and unpublished work on the probability distribution of a binomial parameter. Richard Price (1723-1791), in the lower picture, modified it and presented it at the Royal Society in 1763, after Bayes' death, and published it in 1764.

wikipedia.org and newworldencyclopedia.org

This is an important remark, since evidence computation is usually difficult.

## 4.2 Bayesian inference

Bayesian inference is a statistical inference method that uses Bayes' theorem to update the probability of a hypothesis as more information becomes available. Since we get a distribution, we can predict values for unseen data points and quantify uncertainty of the prediction. The evaluation of this posterior distribution is not easy when the evidence $p(y)$ is intractable. When this happens, we can approximate the inference using different approaches. Here we focus on two.

*Markov Chain Monte Carlo (MCMC)* is a sampling method: samples are drawn from the posterior distribution and are then used to compute various quantities using averaging. Sampling methods are asymptotically exact, but computationally very expensive.

*Variational Expectation Maximization (VEM)* is an approximate inference method: the posterior distribution is approximated with a simple, computationally tractable expression. This method is much faster than sampling methods. However, it is not necessarily exact asymptotically and optimization can fall into local minima. For this reason, comparing to sampling techniques may be useful for validation.

With these techniques, we can construct point-wise estimates either from $p(\phi|y)$ in the case of using sampling, or from the optimization of the posterior approximation derived using variational EM. We also have access to marginal effects of subsets of parameters $p(\phi_i|y)$ and to posterior predictions $p(\tilde{y}|y)$, where $\tilde{y}$ is the model output given $\tilde{\phi}$ the parameter estimate derived from $p(\phi|y)$. Note that $\phi$ here may also contain missing hidden variables.

## 4.3 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a sampling technique based on Monte Carlo integration using Markov Chains [Gilks et al., 1996].

Monte Carlo integration draws samples from the posterior distribution and then approximates the posterior mean with the sample mean. This sample mean tends to the real posterior mean when samples are independent and the number of samples tends to infinity, by the law of large numbers.

The generation of independent samples is not always feasible and Markov chains can be used to deal with this. A Markov chain is a sequence of random variables $\{X_0, X_1, ...\}$ that considers only the recent samples (e.g. $X_{t-1}$) to generate the current one (e.g. $X_t$) considering a transition probability (e.g. $P(X_t|X_{t-1})$). The chain gradually forgets its initial state and the distribution of the output of the Markov chain converges to a unique stationary distribution. If we discard the first samples, called burn-in period, we ensure that the initialization does not affect the mean. The number of samples discarded will depend on how fast the Markov chain

samples converge to stationary-distribution-like samples (called mixing time). Determining the burn-in period is not computationally feasible in most situations, and visual inspection of the MCMC output is usually used [Gilks et al., 1996].

The mean discarding the burn-in samples is called ergodic mean [Gilks et al., 1996] and it will tend to the stationary distribution mean when the number of samples is sufficiently high. This can be checked by running the algorithm several times and checking the similarity of the results.

We can make the Markov chain stationary distribution be our posterior distribution by using the Metropolis-Hastings (HM) algorithm. The MH algorithm consists in sampling a candidate point from a proposal distribution and accepting this candidate with a certain probability. The acceptance probability depends on the stationary distribution and the proposal distribution. A special case of the MH algorithm is the Gibbs sampling [Roberts, 1996], in which the acceptance rate is always 1 because the proposal distribution is the full conditional distribution of the parameter of interest $p(\phi|y)$. If we have a multivariate distribution with $N$ components $\phi = (\phi_1, ...\phi_n)$, we can sample from the distribution of a component conditioned on all other components already sampled (sample one-at-a-time). For example, to sample component $\phi_s$ in iteration $r$ ($\phi_s^{(r)}$), we use the distribution $p(\phi_s|\phi_1^{(r)}, \ldots, \phi_{s-1}^{(r)}, \phi_{s+1}^{(r-1)}, \ldots, \phi_n^{(r-1)})$. All components are sampled in one iteration and the order in which this is done usually remains unchanged, although it is not necessary [Gilks et al., 1996]. We use the components already sampled in the current iteration.

More detailed information can be found in [Robert and Casella, 2013, Roberts, 1996, Gilks et al., 1996].

## 4.4 Expectation Maximization and Variational Expectation Maximization

Expectation Maximization (EM) algorithm [Dempster et al., 1977] is an iterative method for finding maximum likelihood estimates of parameters $\theta$ with unobserved latent variables $X$. EM can be viewed [Neal and Hinton, 1998] as an alternating maximization procedure of the (*negative*) variational free energy[1] function $\mathcal{F}$. The term free energy comes from thermodynamics and it is a negative cost function that we want to maximize. In the E-step, the posterior of the hidden variable $X$ is computed, and in the M-step the parameters $\theta$ of the model are updated by maximizing the free energy function with respect to the set of parameters. See [Bishop, 2006] for further details.

[1] Sometimes the term *negative* is neglected and we call it *free energy*, probably due to the fact that authors use it as the *lower bound* of the likelihood.

When the posterior density and therefore the E-step are intractable, we can approximate the posterior using Variational Expectation Maximization (VEM) [Jordan et al., 1999]. VEM finds an analytical expression $\tilde{p}(x)$ as similar as possible to the posterior distribution $p(x|y)$. For VEM, the iterative procedure reads

$$\textbf{E-step: } \tilde{p}^{(r)} = \arg\max_{\tilde{p} \in \mathcal{D}} \ \mathcal{F}(\tilde{p}, y, \theta^{(r)}) \tag{4.1}$$

$$\textbf{M-step: } \theta^{(r+1)} = \arg\max_{\theta \in \mathbf{\Theta}} \ \mathcal{F}(\tilde{p}^{(r)}, y, \theta) \tag{4.2}$$

where $r$ is the iteration number. The free energy, for any $\tilde{p} \in \mathcal{D}$ and $\theta$, reads

$$\begin{aligned}\mathcal{F}(\tilde{p}, y, \theta) &= \int \tilde{p}(x) \log p(y, x; \theta) dx - \int \tilde{p}(x) \log \tilde{p}(x) dx \\ &= \mathrm{E}_{\tilde{p}}\big[\log p(y, X; \theta)\big] + \mathcal{I}(\tilde{p}) \end{aligned} \tag{4.3}$$

where $\mathrm{E}_{\tilde{p}}\big[\log p(y, X; \theta)\big]$ is the expected complete log-likelihood under the surrogate distribution and $\mathcal{I}(\tilde{p})$ is the entropy of the surrogate distribution. The free energy can be decomposed into

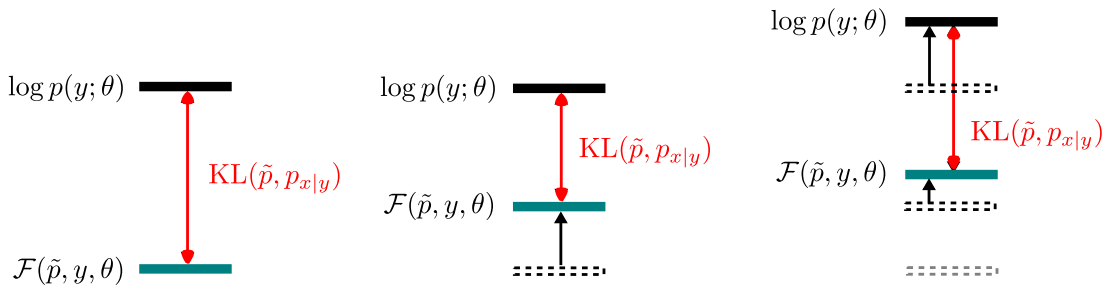
$$\begin{aligned}\mathcal{F}(\tilde{p}, y, \theta) &= \int \tilde{p}(x) \log \frac{p(y, x; \theta)}{\tilde{p}(x)} dx \\ &= \int \tilde{p}(x) \log \frac{p(x|y; \theta) p(y)}{\tilde{p}(x)} dx \\ &= \int \tilde{p}(x) \log p(y) dx - \int \tilde{p}(x) \log \frac{\tilde{p}(x)}{p(x|y; \theta)} dx \\ &= \log p(y) - \mathrm{KL}(\tilde{p}, p_{x|y}) \end{aligned} \tag{4.4}$$

where $\mathrm{KL}(\tilde{p}, p_{x|y})$ is the Kullback Leibler (KL) divergence. The KL divergence quantifies the dissimilarity between the approximate $\tilde{p}(x)$ and the real posterior $p(x|y)$. It is the expectation of the logarithmic difference between $\tilde{p}(x)$ and $p(x|y)$, taking $\tilde{p}(x)$ for the expectation. Although it is unknown, we do know that it is non-negative. Minimizing the KL divergence implies making the approximate posterior more similar to the actual one, and this is equivalent to maximizing the free energy. The free energy becomes a lower bound for the model log evidence:

$$\mathrm{KL}(\tilde{p}, p_{x|y}) \geqslant 0 \tag{4.5}$$

$$\log p(y) = \mathcal{F}(\tilde{p}, y, \theta) + \mathrm{KL}(\tilde{p}, p_{x|y}) \geqslant \mathcal{F}(\tilde{p}, y, \theta) \tag{4.6}$$

See figure 4.2, reproduced from [Beal, 2003, Bishop, 2006].




(a) Initialization (b) Expectation step (c) Maximization step

Figure 4.2: Variational EM minimizes the KL distance in the E step and updates the model log evidence and the free energy in the M step with the new parameters. Figure reproduced from [Beal, 2003, Bishop, 2006].

When we are dealing with several missing variables $X$ in the E step, we can restrict the class of approximate posteriors to the mean field approximation: we consider the posteriors $\tilde{p}(x)$ that factorize into independent distributions $\tilde{p}_i(x_i)$:

$$\tilde{p}(x) = \prod_i \tilde{p}_i(x_i) \tag{4.7}$$

When we consider the mean field approximation, the free energy can be

decomposed into:

$$\begin{aligned}
\mathcal{F}(\tilde{p}, y, \theta) &= \int \prod_i \tilde{p}(x_i) \log \frac{p(y, x; \theta)}{\prod_i \tilde{p}(x_i)} dx \\
&= \int \tilde{p}(x_j) \int \prod_{i \setminus j} \tilde{p}(x_i) \log \frac{p(y, x; \theta)}{\tilde{p}(x_j) \prod_{i \setminus j} \tilde{p}(x_i)} dx_i dx_j \\
&= \int \tilde{p}(x_j) \int \prod_{i \setminus j} \tilde{p}(x_i) \left( \log p(y, x; \theta) - \sum_{i \setminus j} \log \tilde{p}(x_i) \right) dx_i dx_j \\
&\qquad - \int \tilde{p}(x_j) \log \tilde{p}(x_j) dx_j \\
&\propto \int \tilde{p}(x_j) \log \frac{\exp\left( \mathrm{E}_{\tilde{p}_{\setminus j}}[\log p(y, X; \theta)] \right)}{\tilde{p}(x_j)} dx_j \\
&\propto -\mathrm{KL}\left( \tilde{p}_j, \frac{1}{Z} \exp\left( \mathrm{E}_{\tilde{p}_{\setminus j}}[\log p(y, X; \theta)] \right) \right) \qquad (4.8)
\end{aligned}$$

where $Z$ is the normalizing term of the distribution. This shows that we can fix all other densities $\tilde{p}(x_{\setminus j})$ and then approximate the posterior density of a certain variable $\tilde{p}(x_j)$

$$\hat{\tilde{p}}_j = \arg\max_{\tilde{p}_j} F(\tilde{p}, y, \theta) \propto \exp\left( \mathrm{E}_{\tilde{p}_{\setminus j}}[\log p(y, X; \theta)] \right) \qquad (4.9)$$

and by using again Bayes' theorem, this is proportional to

$$\hat{\tilde{p}}_j \propto \exp\left( \mathrm{E}_{\tilde{p}_{\setminus j}}[\log p(x_j | y, X_{\setminus j}; \theta)] \right). \qquad (4.10)$$

We can do this for each hidden variable. In a Bayesian setting we may have, in addition to hidden variables, random parameters. They can be treated as hidden variables in the E step. When we are just dealing with random variables possibly including random parameters, then there is no need for an M-step and this is called Variational Bayes (VB). See [Beal, 2003, Tzikas et al., 2008] for further details. A regular M step as (4.2) appears when parameters are unknown but considered as fixed (frequentist case).

## 4.5 Application to BOLD fMRI: the Joint Detection Estimation model

Bayesian models have been proposed in the analysis of BOLD fMRI in [Woolrich et al., 2004, Penny et al., 2003, 2005, Flandin and Penny, 2007, Gössl et al., 2001, Friston et al., 2002b,a]. As already explained in chapters 2 and 3, BOLD fMRI signal is known to have several components. The component of interest in fMRI is the task related component and it contains the variations that the BOLD signal suffers in response to cerebral stimulation. These variations give rise to the hemodynamic response function (HRF) and the amplitude of the effect gives us a measure of local brain activity. As we have seen in section 3.3, this response function can be modelled. However, the estimation of this function and the detection of activation at the same time is not straightforward.

A joint detection estimation (JDE) approach has been proposed in [Makni et al., 2005, 2008] to be able to jointly estimate the response function

and its amplitude. This method assumes a constant HRF over different areas of the brain, and a varying amplitude in each voxel. For that purpose, it needs an a priori defined parcellation of the cerebral cortex in parcels $\boldsymbol{\mathcal{P}} = (\mathcal{P}_\kappa)_{\kappa=1:}$ with homogeneous hemodynamic properties. The size of each parcel $\mathcal{P}_\kappa$ is typically hundreds of voxels. For a certain voxel $j \in \mathcal{J}$, the generative model of the signal (see also figure 4.3) is:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j \tag{4.11}$$

where $\boldsymbol{h} = (h_{d\Delta t})_{d=0:D}$ is the HRF corresponding to a parcel $\mathcal{P}_\kappa$ and $\Delta t$ is the HRF sampling period. $\boldsymbol{X}^m = (x^m_{n-d\Delta t})_{n=1:N,d=0:D}$ is a binary matrix that encodes the stimulus for each condition $m$ and has dimension $N \times (D+1)$. $a_j^m$ is the response amplitude at voxel $j$ for the $m$th experimental condition. To account for spatial correlation, response amplitudes are assumed to follow spatial Gaussian mixture models [Vincent et al., 2010] governed by hidden Markov random fields (MRF) $\boldsymbol{q}$ that encode the voxel activation states. There is a MRF per experimental condition $\boldsymbol{q}^m$. The activation states are considered conditional to these MRF: $\boldsymbol{a}^m|\boldsymbol{q}^m$.$\boldsymbol{P}\boldsymbol{\ell}_j$ is a term that represents the drifts of the signal due to low frequency variations of the signal. $\boldsymbol{P} \in \mathbb{R}^N \times \mathbb{R}^F$ is an orthogonal function base. $\boldsymbol{b}_j \in \mathbb{R}^N$ is the noise vector for voxel $j$ and is assumed to follow a distribution $\boldsymbol{b}_j \sim \mathcal{N}(0, \boldsymbol{\Gamma}_j^{-1})$ where $\boldsymbol{\Gamma}_j = \boldsymbol{\Lambda}_j / v_{b_j}$. $\boldsymbol{\Lambda}_j$ is the identity matrix when i.i.d. Gaussian noise is considered, and a first order autoregressive model (AR) matrix when temporal correlation is considered (see [Makni et al., 2005] for more details). BOLD data is known to have correlated noise, although sometimes white noise is assumed in its analysis for simplicity.

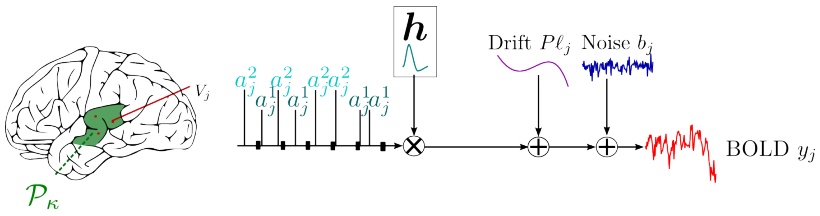



Figure 4.3: Parcel model of the BOLD signal. In each parcel, we consider homogeneous hemodynamic properties and we assume a single HRF $\boldsymbol{h}$. The amplitude of the HRF is modulated voxel-wise with the response levels $\boldsymbol{a}$. There are also drifts and noise.

Note here that the signal model is similar to the one used in the classical method GLM for BOLD: $\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{b}_j$. The main difference is that $\boldsymbol{X}$ in GLM contains the regressors constructed as the convolution of a canonical HRF and a vector encoding the stimulus for a given experimental condition. The matrix $\boldsymbol{X}^m$ in JDE is specific for the $m$th experimental condition and it encodes a convolution. The product $\boldsymbol{X}^m\boldsymbol{h}$ is equivalent to the convolution of the vector encoding the stimulus $\boldsymbol{x}^m$ for a given experimental condition $m$ and the estimated HRF: $\boldsymbol{x}^m * \boldsymbol{h}$. Therefore, it is equivalent to a GLM regressor considering an estimated HRF. The amplitudes $\boldsymbol{\beta}_j$ are independent across voxels in GLM, as opposed to $\boldsymbol{a}_j$'s in JDE, that consider a spatial correlation.

### 4.5.1 Bayesian modelling

JDE uses Bayesian modelling for the estimation of the different parameters. The joint probabilistic model can be decomposed as:

$$p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta}) = p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{a} \mid \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{h} \mid \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{\theta})$$

$\boldsymbol{\theta}$ being the parameters from the drifts and noise, and the ones introduced with the prior distributions considered. As shown in the joint model, $\boldsymbol{y}$ is assumed independent of $\boldsymbol{q}$ given $\boldsymbol{a}$ and $\boldsymbol{h}$, $\boldsymbol{a}$ is assumed to be independent of $\boldsymbol{h}$ conditionally to $\boldsymbol{q}$, and $\boldsymbol{h}$ is assumed to be independent of $\boldsymbol{q}$.

#### Likelihood

The likelihood reads:

$$p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}; \boldsymbol{\ell}, \boldsymbol{\Gamma}) = \prod_{j \in \mathcal{J}} p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}; \boldsymbol{\ell}_j, \boldsymbol{\Gamma}_j) \tag{4.12}$$

$$\text{with} \quad p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}; \boldsymbol{\ell}_j, \boldsymbol{\Gamma}_j) \sim \mathcal{N}\left(\sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j, \boldsymbol{\Gamma}_j^{-1}\right) \tag{4.13}$$

where $\boldsymbol{\Gamma}_j = \boldsymbol{\Lambda}_j / v_{b_j}$, and $\boldsymbol{\Lambda}_j$ is the identity matrix when i.i.d. Gaussian noise is considered, and a first order AR matrix[2] when autocorrelated noise is considered. $\boldsymbol{\Gamma}$ and $\boldsymbol{\ell}$ are unknown.

[2] first order AR precision matrix

$$\begin{pmatrix} 1 & -\rho_j & 0 & \dots & 0 \\ -\rho_j & 1+\rho_j^2 & -\rho_j & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ & & & 1+\rho_j^2 & -\rho_j \\ 0 & \dots & & -\rho_j & 1 \end{pmatrix}$$

#### Prior distributions

We assume prior distributions for the different variables.

- The prior distribution proposed for the HRF is a Gaussian with mean zero and covariance a matrix that introduces smoothness like matrix $\boldsymbol{R}$ in section 3.3: $p(\boldsymbol{h}; v_h) \sim \mathcal{N}(0, v_h \boldsymbol{R})$. A scale parameter $v_h$ controls how smooth is the response.

- Gaussian Mixture models are assumed for the response amplitudes, with assumed independence between the different experimental conditions. For each experimental condition, the mixture model segregates active from non-active voxels, activation being encoded in a different binary variable $\boldsymbol{q}$. In figure 4.4, $q_j^m = 1$ and $q_j^m = 0$ mean that voxel $j$ is active and not respectively, for experimental condition $m$. The prior probability distribution for the response amplitudes is:

  $$p(\boldsymbol{a} | \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v}) = \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{i=0}^{1} \mathcal{N}(a_j^m; \mu_i^m, v_i^m)^{\mathbb{1}(q_j^m = i)}$$

  Parameters $\boldsymbol{\mu}$ and $\boldsymbol{v}$ are unknown.

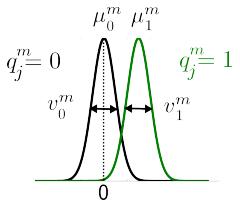



Figure 4.4: Prior distribution of the response amplitudes. It is a GMM with two Gaussians corresponding to not activated, with mean 0, and activated classes.

- Spatial correlation is directly incorporated in the probabilities of activation through a *hidden* Potts field on the binary variables $\boldsymbol{q}^m$ (see figure 4.5). The proportions of voxels for the different classes is not explicit. Independence between experimental conditions is assumed: $p(\boldsymbol{q}; \boldsymbol{\beta}) = \prod_{m=1}^{M} p(\boldsymbol{q}^m; \beta^m)$, where $\beta^m$ is the amount of spatial regularization for condition $m$. Here $p(\boldsymbol{q}^m; \beta^m)$ is a 2-class Potts (Ising)

model with interaction parameter $\beta^m$, that controls the amount of spatial regularization, and no external field.

$$p(\boldsymbol{q}_m;\beta_m) = Z(\beta^m)^{-1} \exp\left(\frac{1}{2}\beta^m \sum_{j\in J}\sum_{i\in\mathcal{N}(j)} \mathbb{1}(q_j^m = q_i^m)\right), \quad (4.14)$$

where $\mathbb{1}(K) = 1$ if $K$ is true, 0 otherwise. $Z(\beta^m)$ is the normalizing constant or the partition function of the Potts field. $\boldsymbol{\beta} = \{\beta^m, m = 1..M\}$ are the unknown parameters.

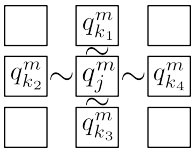



Figure 4.5: Prior distribution of the binary activation states: MRF introduces a dependency between neighbouring voxels.

- Drift coefficients $\boldsymbol{\ell}_j$ are assumed to be Gaussian a priori: $\boldsymbol{\ell}_j \sim \mathcal{N}(0, v_\ell I_F)$.

- For the rest of the parameters, we use non-informative Jeffreys priors for MCMC and conjugate priors for VEM. It will be further explained in chapters 5 and 6.

The prior parameters will be estimated in the model $\boldsymbol{\theta} = \{\boldsymbol{\ell}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{v}, v_h, \boldsymbol{\beta}, v_\ell\}$. The full joint model becomes:

$$p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta}) = p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}; \boldsymbol{\ell}, \boldsymbol{\Gamma})\ p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\ p(\boldsymbol{h}; v_h)\ p(\boldsymbol{q}; \boldsymbol{\beta}) p(\boldsymbol{\theta}).$$

The graphical hierarchical model can be found in figure 4.6.

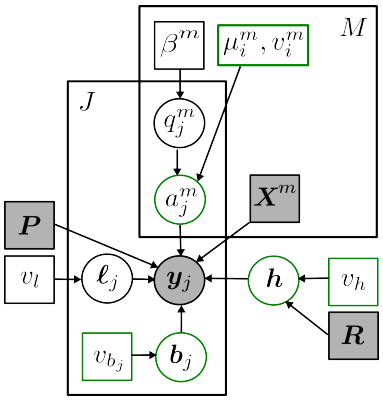



Figure 4.6: Graphical hierarchical model.

### 4.5.2 Estimation of the posterior probability distribution

The posterior probability distribution of this model is intractable due to the Markov Random Field introduced to consider spatial correlation. MCMC and VEM methods have been proposed to deal with this posterior density in the estimation of the parameters in JDE.

#### Markov Chain Monte Carlo solution for BOLD JDE

MCMC was proposed in [Makni et al., 2008, Vincent et al., 2010]. For each parameter, we sample from its full posterior conditional distribution:

$$\boldsymbol{h}^{(r)} \leftarrow p(\boldsymbol{h}|\boldsymbol{y}, \boldsymbol{a}^{(r-1)}; \boldsymbol{\theta}^{(r)}) \quad (4.15)$$
$$\boldsymbol{a}^{(r)} \leftarrow p(\boldsymbol{a}|\boldsymbol{y}, \boldsymbol{h}^{(r)}, \boldsymbol{q}^{(r-1)}; \boldsymbol{\theta}^{(r)}) \quad (4.16)$$
$$\boldsymbol{q}^{(r)} \leftarrow p(\boldsymbol{q}|\boldsymbol{y}, \boldsymbol{a}^{(r)}; \boldsymbol{\theta}^{(r)}) \quad (4.17)$$
$$\boldsymbol{\theta}^{(r+1)} \leftarrow p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{h}^{(r)}, \boldsymbol{a}^{(r)}, \boldsymbol{q}^{(r)}) \quad (4.18)$$

All parameters in $\boldsymbol{\theta}$ are sampled iteratively from their respective posterior densities. In this formalism, $\boldsymbol{\ell}_j$ is not treated as a component of $\boldsymbol{\theta}$ but either as a random variable or integrated out before sampling. Many samples are computed and the posterior means are found by averaging all the samples ($S$) after a *burn-in* period ($b$). For example, for $\boldsymbol{h}$

$$\hat{\boldsymbol{h}}^{PM} = \frac{1}{S-b+1}\sum_{r=b}^{S} \boldsymbol{h}^{(r)} \quad (4.19)$$

### Variational Expectation Maximization solution for BOLD JDE

A VEM solution was proposed in [Chaari et al., 2013] and the formulas of the **E-step** for this model read:

$$\tilde{p}_h^{(r)}(h) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r-1)}\tilde{p}_q^{(r-1)}}\left[\log p(\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{a}^{(r-1)}, \boldsymbol{q}^{(r-1)}; \boldsymbol{\theta}^{(r)})\right]\right) \quad (4.20)$$

$$\tilde{p}_a^{(r)}(a) \propto \exp\left(\mathrm{E}_{\tilde{p}_h^{(r)}\tilde{p}_q^{(r-1)}}\left[\log p(\boldsymbol{a} \mid \boldsymbol{y}, \boldsymbol{h}^{(r)}, \boldsymbol{q}^{(r-1)}; \boldsymbol{\theta}^{(r)})\right]\right) \quad (4.21)$$

$$\tilde{p}_q^{(r)}(q) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_h^{(r)}}\left[\log p(\boldsymbol{q} \mid \boldsymbol{y}, \boldsymbol{a}^{(r)}, \boldsymbol{h}^{(r)}; \boldsymbol{\theta}^{(r)})\right]\right) \quad (4.22)$$

The corresponding **M-step** is:

$$\boldsymbol{\theta}^{(r+1)} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\ \mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_h^{(r)}\tilde{p}_q^{(r)}}\left[\log p(\boldsymbol{y}, \boldsymbol{a}^{(r)}, \boldsymbol{h}^{(r)}, \boldsymbol{q}^{(r)}\ ; \boldsymbol{\theta}^{(r)})\right]. \quad (4.23)$$

For further detail, see the multiple-session model in appendix B. Both solutions are implemented in the python package PyHRF [3].

[3] pyhrf.org

### Outputs of JDE

Once inference performed over each parcel, we have an HRF estimated from each parcel of the brain, a brain map with the amplitudes of these HRFs per voxel and experimental condition, a brain image with the activation probabilities per experimental condition, and the rest of the parameters. Posterior probability maps can be afterwards obtained from the activation levels after the estimation of the posterior probability functions.

## 4.6 Posterior probability maps

Since the output of our Bayesian model are probability distributions, we can use this information to decide which voxels are activated.

As we saw in the previous chapter, statistical parametric maps (SPMs) are widely used in classical methods and they are easy to deal with because of their interpretability. However, derived p-values do not reflect the likelihood that an effect is present. Instead, they reflect the probability of observing the data in the absence of any effect [Penny and Friston, 2004]. If sufficiently small, it can be used to reject the null hypothesis that the effect is not present. This has several limitations according to [Penny and Friston, 2004]:

- The alternate hypothesis can not be rejected: the probability that an effect is exactly zero is zero.
- SPMs are very sensitive to overfitting: given enough scans or subjects one can always demonstrate a significant effect at every voxel if sufficient degrees of freedom are used to achieve a small variance.
- The need for a correction for multiple comparisons makes the inference of a brain region dependent on the rest of the regions.

This can be avoided by using a more Bayesian approach.

Posterior probability maps (PPM) were proposed in [Friston and Penny, 2003] to construct maps describing the probability that an activation exceeds some specified threshold in every voxel, given the data:

$$p(\boldsymbol{\gamma}^t \boldsymbol{\beta}_j > \delta | \boldsymbol{y}_j) > \alpha$$

where $\delta$ has to be chosen carefully, and $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}^t$ are the effects and the contrasts of interest, respectively (see section 3.2.2). In JDE, $\delta$ can be chosen as the point where the two Gaussian densities in the GMM encoding activation and non-activation in $\boldsymbol{a}$ intersect. Another possibility is choosing $\delta$ as the 5% of the parcel signal mean or of the brain signal mean.

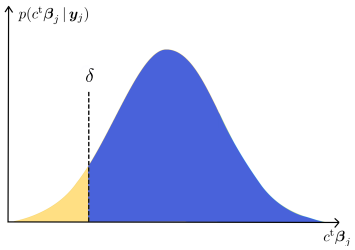



Figure 4.7: Posterior probability map.

As noted in [Friston et al., 2002a, Friston and Penny, 2003], thresholding a PPM could be seen as controlling the false discovery rate (FDR) or the percentage of false activations allowed, since it establishes an upper bound on the FDR [Benjamini and Hochberg, 1995]. SPM and PPM maps have been compared for fMRI in [Friston et al., 2002a].

PPMs are implemented in the SPM software[4] and derived by using empirical variational Bayes, a variational Expectation Maximization [Penny et al., 2003, Penny and Friston, 2004] approach in which all priors are estimated from the data.

[4] http://www.fil.ion.ucl.ac.uk/spm/

## 4.7 Parcellation of the brain

The JDE approach operates on a prior partitioning of the brain into functionally homogeneous parcels, where the hemodynamics is assumed constant. This parcellation can be either an atlas or a previously computed data-based parcellation of the brain. Several atlases exist that divide the brain in anatomical or functional regions:

*Automated Anatomical Labeling (AAL):* It is an anatomical atlas of the brain, containing 45 anatomical volumes of interest in each hemisphere. It was proposed in [Tzourio-Mazoyer et al., 2002] and the regions were drawn manually after the delineation of the sulci (cerebral cortex grooves) from a template. In the work, an automatic labelling of functional MRI activations was proposed. See figure 4.8.

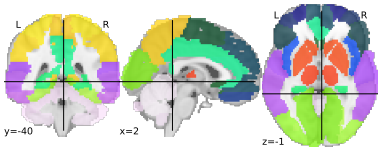



Figure 4.8: Structural atlas AAL (Automatic Atlas Labeling).

*Harvard-Oxford:* It can be found in FSL [Jenkinson et al., 2012]. It is a probabilistic atlas of human cortical brain areas (lateralized). It covers 48 cortical and 21 subcortical structural areas. It was derived from the structural images of 21 healthy male and 16 healthy female subjects of ages between 18 and 50. They were individually segmented using semi-automated tools. Segmentations were provided by the Harvard Center for Morphometric Analysis. See figure 4.9.

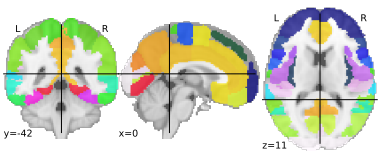



Figure 4.9: Harvard-Oxford functional atlas.

*Brodmann:* It contains 52 functional regions of the cerebral cortex. They were described by Korbinian Brodmann in 1909 by examining the different cellular morphology and organization of humans and monkeys brains. The english version of his book can be found in [Brodmann, 2007]. Posterior studies found a correlation with cortical function specialization.

*Neuromorphometrics:* It has been created by *Neuromorphometrics Inc.* from a database of 114 manually labeled brains of subjects ranging in age from 5 to 96. They have been labeled and double-checked by experts using two protocols: "general segmentation" defined by the MGH Center for Morphometric Analysis and parcellation of the cerebral cortex into regions defined by gyral and sulcal landmarks (cerebral cortex ridges and grooves, respectively). See figure 4.10.

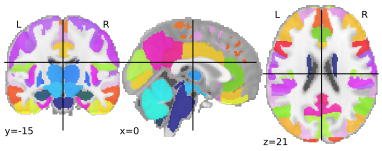



Figure 4.10: Neuromorphometrics functional atlas.

*Willard:* It is a functional atlas created from the findings in [Richiardi et al., 2015]. It provides a parcellation in 499 regions of interest, considering 13 well-known functional networks. From the resting-state functional MRI images of 14 subjects, 13 well-known functional networks were identified using Independent Component Analysis (ICA). The post-mortem cortex samples of 6 subjects from the Allen Institute for Brain Science human (AIBS) microarray dataset were then mapped to the functional regions. Validation was performed in 259 subjects from the IMAGEN database [Schumann et al., 2010]. This is a good parcellation for use in JDE because parcels are quite homogeneous and of approximately a hundred of voxels or two, when we consider a typical fMRI voxel resolution. See figure 4.11.

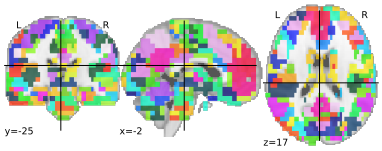



Figure 4.11: Willard functional atlas with a resolution $3 \times 3 \times 3$ mm$^3$.

[5] http://nilearn.github.io

Other atlases can be found in nilearn[5] [Abraham et al., 2014]: [Varoquaux et al., 2011, Power et al., 2011, Bellec et al., 2010, Craddock et al., 2012] between many.

A data-based parcellation could be also useful due to the variability of the hemodynamics in different brains but to ensure a good JDE performance the parcellation needs to be robust. A few attempts have been proposed to cope with this issue [G. Flandin et al., 2002, B. Thirion et al., 2006, T. Vincent et al., 2008, A.-L. Fouque et al., 2009, L. Chaari et al., 2012, Badillo et al., 2013a] but most of them are either too computationally demanding [A.-L. Fouque et al., 2009, L. Chaari et al., 2012, Badillo et al., 2013a] or do not account for hemodynamics variability [G. Flandin et al., 2002, B. Thirion et al., 2006, T. Vincent et al., 2008].

### 4.7.1 Hemodynamically informed parcellation of cerebral fMRI data

In [Frau-Pascual et al., 2014a], we attempt a fast hemodynamically based parcellation for use in daily applications prior to JDE inference. We propose a two-step approach consisting first of hemodynamics feature extraction, in which a general linear model (GLM) is used to discriminate hemodynamics information, followed by a parcellation of these features. The goal of the first step is finding features that are able to catch most of the hemodynamic information, without the need of perfectly estimating the HRF function. Afterwards, an agglomerative clustering algorithm based on Gaussian Mixtures is used to segment the features. The main contribution is the consideration of the detection-estimation effect within the parcellation step: there is a lack of hemodynamics information in the non- or slightly-activating voxels. The idea is then to enforce grouping these uncertain voxels with neighboring activating voxels. This is done through a spatial

constraint in the agglomerative step of the parcellation procedure. The uncertainty in a given voxel can be quantified by a statistics linked to its activation level, namely a p-value obtained in the GLM feature extraction step. This statistics is hence injected within the agglomeration criterion. For this reason, we call it Informed Gaussian Mixture based parcellation (IGMM). The approach has been tested on artificial fMRI data sets.

### Generation of artificial fMRI data sets

Let us define the set of all parcels as $\mathbb{P} = \{\mathcal{P}_1, ..., \mathcal{P}_\kappa, ..., \mathcal{P}_K\}$ where $\mathcal{P}_\kappa$ is the set of position indexes belonging to parcel $\kappa$ and $\mathcal{J}_\kappa$ denotes the set of positions in parcel $\kappa$. For this section, we introduce the subscript to indicate the parcel. Artificial BOLD data sets are generated using the regional BOLD model in (4.11). Recall that for a given voxel $j \in \mathcal{J}_\kappa$, and a given parcel $\kappa$:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j, \tag{4.24}$$

where $a_j^m$ is the response amplitude at voxel $j$ for a certain condition $m$, $\boldsymbol{X}^m = (x^m_{n-d\Delta t})_{n=1:N,d=0:D}$ is the binary matrix encoding the stimulus for each condition $m$, $\boldsymbol{h} = (h_{d\Delta t})_{d=0:D}$ is the HRF corresponding to parcel $\mathcal{P}_\kappa$, $\Delta t$ being the HRF sampling period, $\boldsymbol{P}$ the orthogonal function basis multiplied by the drift $\boldsymbol{\ell}_j$, and $\boldsymbol{b}_j$ the noise vector. Note that $\boldsymbol{y}_j = [y_j(t_1), ..., y_j(t_N)]^t$, where $t_n = n\ TR$ and $TR \gg \Delta t$. Typical values are $\Delta t = 0.6$ and $TR = 1.8$ s.

In this section, we considered artificial data at low SNR, with one experimental condition represented with a $20 \times 20$-voxel binary activation labels $\boldsymbol{q} = [q_1 ... q_J]$, and levels of activation $\boldsymbol{a} = [a_1 ... a_J]$, with $(a_j | q_j = 1) \sim \mathcal{N}(1.8, 0.25)$. We simulated a map of hemodynamics parcels, with different HRF shapes $\boldsymbol{h}$ (duration 25 s, $TR = 1$ s, $\Delta t = 0.5$ s) in each parcel $\mathcal{P}_\kappa$ (see Fig. 4.12), by using the combination of 3 Bezier's curves, each being controlled by 4 points, to describe the curves until the peak, from the peak to the undershoot, and from the undershoot to the end, given specific peak and undershoot widths. We considered a Discrete Cosine Transform for $\boldsymbol{P}$, a drift $\boldsymbol{\ell}_j \sim \mathcal{N}(0, 11\boldsymbol{I}_4)$, and white Gaussian noise with variance $v_{b_j} = 1.5$.

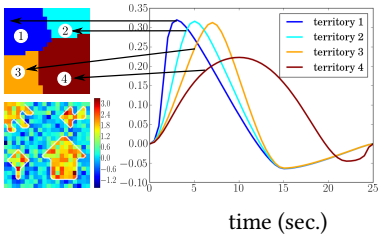



Figure 4.12: Artificial fMRI data sets. Top left: hemodynamic parcels. Bottom left: response levels. Right: HRFs associated with parcels.

### Feature extraction

As regards hemodynamic feature extraction, several approaches are available. Here, we only focus on GLM-based ones involving either canonical HRF and its derivative(s) [Friston et al., 1998]. We chose not to rely on Finite Impulse Response (FIR) models [Henson et al., 2001] since they may be too sensitive to noise. Although more flexible regularized FIR (RFIR) approaches such as [P. Ciuciu et al., 2003] are also theoretically able to recover any HRF shape accurately in contrast to canonical GLM, RFIR inference is pretty difficult when the SNR is too low since it proceeds voxelwise. Moreover, it is time consuming because it performs unsupervised estimation (cf [A.-L. Fouque et al., 2009]). Here, we are more interested in a quick feature extraction step that also allows us to disentangle true active voxels from non-active ones.

As regards canonical GLMs, our feature extraction step consists in fitting the following linear model: $\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{b}_j$, where $\boldsymbol{y}_j$ is the BOLD

signal, $\boldsymbol{X}$ the GLM design matrix, $\boldsymbol{\beta}_j$ the parameter estimates and $\boldsymbol{b}_j$ the noise at voxel $j$. Let us denote $\beta_{j,0}$, $\beta_{j,1}$ and $\beta_{j,2}$ the parameters associated with the regressors in $\boldsymbol{X}$, which derives from the convolution of the experimental paradigm with the canonical HRF $\boldsymbol{h}$, its temporal $\boldsymbol{h}'$ and dispersion derivatives $\boldsymbol{h}''$, respectively. We assume that $\beta_{j,0}$ rather contains information about the HRF magnitude, whereas $\beta_{j,1}$ and $\beta_{j,2}$ provide information about the HRF delay and dispersion respectively, and, hence, are useful to differentiate hemodynamic territories. Maximum likelihood (ML) inference enables to get the parameter estimates $\widehat{\boldsymbol{\beta}}_j$ in each voxel among which we only retain $\boldsymbol{\phi}_j = [\widehat{\beta}_{j,1}, \widehat{\beta}_{j,2}]^{\mathrm{t}}$, as input features to the parcellation method.

To quantify the activation level, we consider the p-value $p_{j,0}$ associated with testing $H_0 : \beta_{j,0} = 0$ and we use the notation $\alpha_j = 1 - p_{j,0} \in (0, 1)$ for these statistics in voxel $j$: The higher the $\alpha_j$ value, the larger our confidence in the presence of evoked activity in voxel $j$. Importantly, the statistics $\alpha_j$ does not enter in the parcellation along with the previously defined features $\boldsymbol{\phi}_j$, it is rather used as weights in the agglomeration criterion.

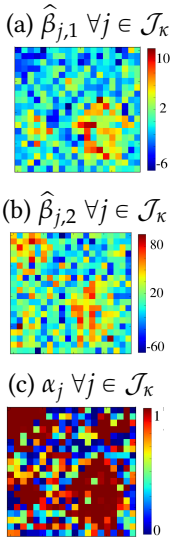



Figure 4.13: Hemodynamics features extracted $\boldsymbol{\phi}_j = [\widehat{\beta}_{j,1}, \widehat{\beta}_{j,2}]^{\mathrm{t}}\ \forall j \in \mathcal{J}_\kappa$, and activation levels $\alpha_j\ \forall j \in \mathcal{J}_\kappa$.

### Informed Gaussian mixture based parcellation

*Agglomerative clustering algorithms.* The model based interpretation of agglomerative clustering algorithms [Kamvar et al., 2002] makes the assumption that features have been generated by probability distributions that vary across parcels. In the context of model-based hard clustering, which aims at assigning classes to the input samples instead of weights, the goal is to maximize the classification likelihood with respect to both $\boldsymbol{\Theta}$ and $\boldsymbol{z}$ given a set of features $\boldsymbol{\Phi}$:

$$\mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{z} \,|\, \boldsymbol{\Phi}) = \prod_{j=1}^{J} p(\boldsymbol{\phi}_j | \boldsymbol{\theta}_{z_j}) \tag{4.25}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_\kappa\}_{\kappa=1:K}$ is the set of parcel-specific model parameters and $\boldsymbol{z} = \{z_j\}_{j=1:J}$ denotes the set of parcel labels associated with each voxel. If we assume that we can find the best possible parameters $\boldsymbol{\Theta}$, denoted by $\hat{\boldsymbol{\Theta}}$, then the problem amounts to optimizing the label likelihood function $\hat{\mathcal{L}}$:

$$\hat{\mathcal{L}}(\boldsymbol{z}|\boldsymbol{\Phi}) = \max_{\Theta} \mathcal{L}(\boldsymbol{\Theta}; \boldsymbol{z}|\boldsymbol{\Phi}) \tag{4.26}$$

In an agglomerative approach, this function is *approximately* optimized by successive merge operations, starting from an initial clustering guess or singleton clusters. Hence, at each step, i.e. when merging two clusters $\mathcal{P}_\kappa$ and $\mathcal{P}_\tau$ of the current parcellation $\boldsymbol{z}$ into the parcel $\mathcal{P}_{\kappa'} = \mathcal{P}_\kappa \cup \mathcal{P}_\tau$ of the resulting parcellation $\boldsymbol{z}'$, the relative increase of the log-likelihood has to be maximized:

$$\log \Delta\hat{\mathcal{L}}(\boldsymbol{z}, \boldsymbol{z}'|\boldsymbol{\Phi}) = \log \left( \frac{\hat{\mathcal{L}}(\boldsymbol{z}'|\boldsymbol{\Phi})}{\hat{\mathcal{L}}(\boldsymbol{z}|\boldsymbol{\Phi})} \right) \tag{4.27}$$

Note that the optimal model parameters $\hat{\boldsymbol{\Theta}}$ have to be obtained for each computation of the label likelihood function. A given merging step thus involves several likelihood maximization over parameters $\boldsymbol{\Theta}$.

*Gaussian-mixture model.* To account for the activation level $\alpha_j$ associated with each voxel $j$, we rely on an independent two-class Gaussian mixture in the agglomerative step. The rationale is that features $\boldsymbol{\phi}_j$ are distributed differently within a given parcel depending on the corresponding activation levels $\alpha_j$. Hence, the two-class mixture is expressed on every $\boldsymbol{\phi}_j$ in parcel $\kappa$ as a way to model parameter differences related to activation levels:

$$p(\boldsymbol{\phi}_j|\boldsymbol{\theta}_\kappa) = \sum_{i=0}^{1} \Pr(q_j = i) f(\boldsymbol{\phi}_j \,|\, q_j = i; \boldsymbol{\theta}_\kappa) = \sum_{i=0}^{1} \lambda_{\kappa,i} \mathcal{N}(\mu_{\kappa,i}, \Sigma_{\kappa,i}) \tag{4.28}$$

The latent variable $q_j$ encodes the activation state of voxel $j$ and $\Pr(q_j = 1) = \lambda_{\kappa,1}$ reflects the probability of activation. This latent variable can be directly linked to the activation statistics $\alpha_j$ obtained at the feature extraction step: $\hat{\lambda}_{\kappa,1} = \sum_{j\in\mathcal{J}_\kappa} \alpha_j / J_\kappa = 1 - \hat{\lambda}_{\kappa,0}$ where $\mathcal{J}_\kappa$ denotes the set of voxels in parcel $\kappa$ and $J_\kappa$ their cardinality. Then, straightforward calculations give the following ML estimators for the parcel-level mixture moments:

$$\hat{\mu}_{\kappa,0} = \frac{\sum_{j\in\mathcal{J}_\kappa}(1-\alpha_j)\,\boldsymbol{\phi}_j}{\sum_{j\in\mathcal{J}_\kappa}(1-\alpha_j)}, \tag{4.29}$$

$$\hat{\mu}_{\kappa,1} = \frac{\sum_{j\in\mathcal{J}_\kappa}\alpha_j\,\boldsymbol{\phi}_j}{\sum_{j\in\mathcal{J}_\kappa}\alpha_j}, \tag{4.30}$$

$$\hat{\Sigma}_{\kappa,0} = \frac{\sum_{j\in\mathcal{J}_\kappa}(1-\alpha_j)\,(\boldsymbol{\phi}_j - \hat{\mu}_{\kappa,0})(\boldsymbol{\phi}_j - \hat{\mu}_{\kappa,0})^{\mathrm{t}}}{\sum_{j\in\mathcal{J}_\kappa}(1-\alpha_j)}, \tag{4.31}$$

$$\hat{\Sigma}_{\kappa,1} = \frac{\sum_{j\in\mathcal{J}_\kappa}\alpha_j\,(\boldsymbol{\phi}_j - \hat{\mu}_{\kappa,1})(\boldsymbol{\phi}_j - \hat{\mu}_{\kappa,1})^{\mathrm{t}}}{\sum_{j\in\mathcal{J}_\kappa}\alpha_j}, \tag{4.32}$$

where $\hat{\mu}_{\kappa,i}$ and $\hat{\Sigma}_{\kappa,i}$, with $i = \{0,1\}$, define the empirical weighted mean and covariance of features in parcel $\kappa$. Note that we consider activation informed from previously estimated feature $\alpha_j$, or non-activation, informed from $(1-\alpha_j)$, for classes active ($i=1$) and non-active ($i=0$) respectively. Once the parameters $\boldsymbol{\Theta}$ have been estimated, the two parcels $\mathcal{P}_\kappa$ and $\mathcal{P}_\tau$ that are selected to be merged into $\mathcal{P}_{\kappa'} = \mathcal{P}_\kappa \cup \mathcal{P}_\tau$ are those which maximize:

$$\begin{aligned}\log \Delta\hat{\mathcal{L}}(\boldsymbol{z}, \boldsymbol{z}'|\boldsymbol{\Phi}) &= \sum_{j\in\mathcal{J}} \Big(\log p(\boldsymbol{\phi}_j|\hat{\boldsymbol{\theta}}_{z'_j}) - \log p(\boldsymbol{\phi}_j|\hat{\boldsymbol{\theta}}_{z_j})\Big) \\ &= \sum_{j\in\mathcal{J}_\kappa} \log\Big(\sum_{i=0}^{1} \hat{\lambda}_{\kappa,i}\mathcal{N}(\hat{\mu}_{\kappa,i}, \hat{\Sigma}_{\kappa,i})\Big) \\ &+ \sum_{j\in\mathcal{J}_\tau} \log\Big(\sum_{i=0}^{1} \hat{\lambda}_{\tau,i}\mathcal{N}(\hat{\mu}_{\tau,i}, \hat{\Sigma}_{\tau,i})\Big) \\ &- \sum_{j\in\mathcal{J}_{\kappa'}} \log\Big(\sum_{i=0}^{1} \hat{\lambda}_{\kappa',i}\mathcal{N}(\hat{\mu}_{\kappa',i}, \hat{\Sigma}_{\kappa',i})\Big)\end{aligned}$$

where the first two terms correspond to the parcels being merged $\mathcal{P}_\kappa$ and $\mathcal{P}_\tau$, and the third term to the final parcel $\mathcal{P}_{\kappa'}$.

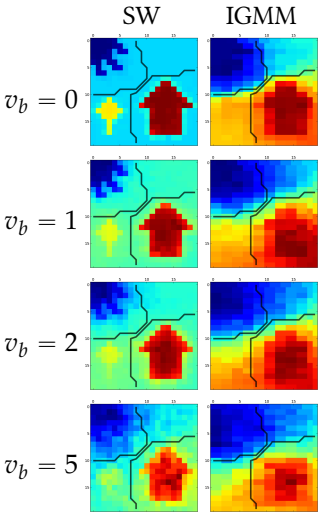



Figure 4.14: Parcellations obtained by Spatial Ward (SW) and IGMM methods, considering an averaged 100 iterations Monte Carlo experiment. From left to right: noise variance 0, 1, 2 and 5, respectively.

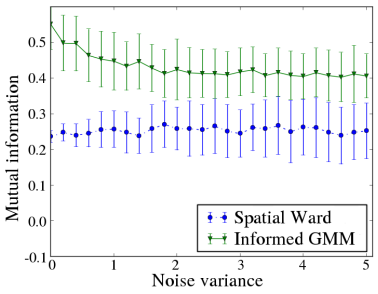



Figure 4.15: MI-based quantitative comparison of SW and IGMM parcellation results with the ground truth territories for noise variances ranging from 0.0 to 5.0.

### Parcellation results on artificial data

Using the generated artificial data, in figure 4.14 we compare the proposed Informed GMM (IGMM) method with the Ward [Ward, 1963] algorithm with connectivity constraints, that here we name Spatial Ward (SW). A Monte Carlo experiment is used, where the variability against several random data sets is assessed, to quantify the results of both parcellation methods. Fig. 4.14 shows results averaged across 100 runs and for different noise variance levels. SW makes a big parcel for all the non-active positions, while IGMM overcomes this issue and partitions the positions independently of the activation level.

For a quantitative evaluation, mutual information (MI) [Pedregosa et al., 2011] was used to compare our parcellation results with the ground truth territories. Fig. 4.15 shows the evolution of both parcellation techniques with respect to increasing noise variance. IGMM outperforms SW and has a decreasing mean value until $MI \approx 0.4$, whereas SW has an almost constant mean $MI \approx 0.25$. Note that both methods are sensitive to noise and have a high variance.

Since the motivation to create this parcellation method is its posterior use in the JDE approach, we study the impact of the parcellation in JDE. As input parcellations, we considered the hard clustering resulting from either the SW- or IGMM-based average parcels computed over the 100 individual results of our previous Monte Carlo experiment, i.e., the ones shown in Fig. 4.14 for $v_b = 1.5$.

Fig. 4.16 compares the averaged detection results over 100 JDE iterations. First, we can see a slightly lower mean detection MSE for SW-based hard parcellation. Also, a lower activation level can be observed for voxels considered in the wrong IGMM-based parcel. Fig. 4.17 shows the averaged estimated HRF profiles over 100 JDE iterations whether it is based on SW or IGMM parcellation methods, compared with the ground-truth HRFs in Fig. 4.12. Overall, HRFs are well recovered by both SW-based or IGMM-based JDE analyses. In region 2 (cyan), IGMM-based better fits the ground-truth since it mixes less voxels with different hemodynamics than SW which includes all non-activating voxels. In region 3 (yellow), SW seems to yield a MC-averaged HRF estimate slightly closer to the ground-truth than IGMM which produces a parcel that also spans region 4 (red). However, the MC variability is higher in this region than in the others for both parcellation methods as shown by the error bars.

Regarding to computation time, IGMM takes 130 times more than SW, but 1800 times less than the Consensus Clustering JDE (CC-JDE) parcellation method [Badillo et al., 2013a].

### Discussion on the hemodynamically informed parcellation

A hierarchical parcellation method that takes into account the activation levels in the parcellation process has been developed, so as to be consistent with JDE assumption that one parcel contains active and inactive positions, and find the underlying hemodynamic territories, independently of the activation level. In terms of computation load, it is not a demanding method. However, hierarchical agglomerative algorithms need the selection

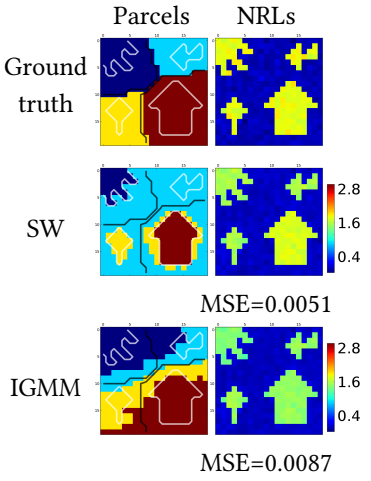



Figure 4.16: JDE detection results for SW and IGMM. From left to right: parcellation used, averaged estimated NRLs over 100 JDE iterations with detection MSE $= \frac{\|\boldsymbol{e}^d\|^2}{\|\boldsymbol{a}^{true}\|^2}$, being $\boldsymbol{e}^d = [e_1^d ... e_J^d]$, $\boldsymbol{a}^{true} = [a_1^{true} ... a_J^{true}]$, and $e_j^d = \hat{a}_j - a_j^{true}$.

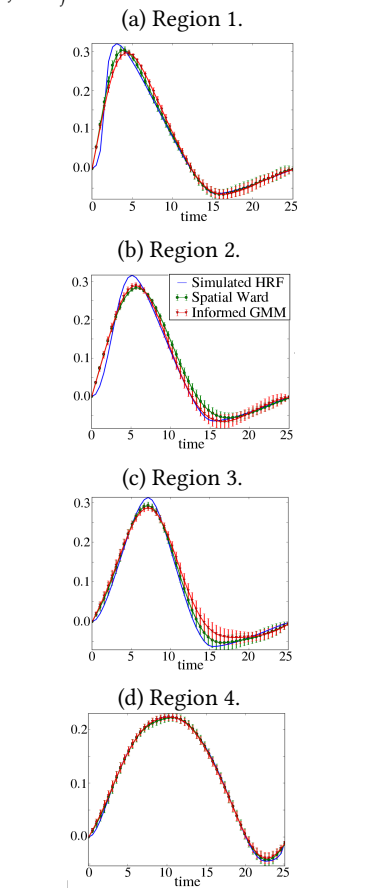



Figure 4.17: JDE HRF estimation averaged results over 100 JDE iterations (errorbars for standard deviation), for the different regions, labelled in Fig. 4.12.

in advance of the number of clusters, and this fact lead us to a definition problem. In the JDE framework, we want parcels big enough to be able to estimate accurately the HRF, although if parcels are too big, we will be losing information and averaging the hemodynamics. However, if we can have a good estimate of hemodynamics, we are interested in having as much as possible parcels to better recover the territories' singularities. The computation of the number of clusters in the context of JDE with an intrinsic parcellation computation, that has been called Joint Parcellation Detection Estimation (JPDE) [L. Chaari et al., 2012, Chaari et al., 2016], has been posteriorly addressed in [Albughdadi et al., 2014] and [Albughdadi et al., 2016a] for the subject level, and in [Albughdadi et al., 2016b] for the multiple-subject case.

We have quantitatively validated that the proposed IGMM approach enables a better recovery than a reference spatial Ward approach. Indeed, parcels obtained with IGMM are less influenced by highly activated positions and do not mix non-active positions altogether. JDE results are quite comparable in terms of detection and HRF estimation whether the input parcellation comes from the SW method or the proposed IGMM approach. Still, the proposed approach yields more reliable parcellations than SW and may be more adequate to treat real data sets.

Experiments in real data need to be done to properly validate the parcellation method. In the context of this thesis, no further work has been done in this topic. In the next chapters, an atlas has been used as JDE input parcellation for the sake of simplicity. First of all, results on simulated data show that JDE results do not vary dramatically when we use different parcellations. Also, using an atlas removes a layer of variability and complexity in the use of JDE, and the comparison across subjects and datasets becomes more direct and easier.

This work was presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) conference in 2014.

## 4.8 Multiple-session extension of the VEM JDE algorithm

The previously presented approach for BOLD analysis with a JDE approach considers a single session. Usually several sessions are recorded for one subject in order to be able to combine or compare them. Their combination is important to achieve a higher statistical power without making too long sessions that would be uncomfortable for the subject being scanned. The acquisition of several sessions can be used for validation too.

For a multiple-session approach, we need to make some other considerations in our model: $\boldsymbol{y} = \{\boldsymbol{y}_j^s, j \in \mathcal{J}, s \in \mathcal{S}\}$ where $\mathcal{J}$ is the set of voxels and $\mathcal{S}$ the set of sessions. In [Badillo et al., 2013b] a multiple-session extension of the JDE framework is proposed and it is solved using MCMC. In this extension, the response levels per condition $\boldsymbol{a}^m$ are considered to have a mean value $\bar{\boldsymbol{a}}^m$ that could be considered the true subject response level, and an inter-session variation that could be considered error measurement:

$$\boldsymbol{a}^{m,s} = \bar{\boldsymbol{a}}^m + \epsilon^s \quad \text{with} \quad \epsilon^s \sim \mathcal{N}(0, v_a \boldsymbol{I})$$

Therefore, we consider session specific response levels

$$\boldsymbol{a} = \left\{ \boldsymbol{a}_j^{m,s}, m = 1 : M, s \in \mathcal{S}, j \in \mathcal{J} \right\}.$$

The HRF $\boldsymbol{h}$ is considered unique through sessions to improve the estimation. The rest of the parameters are estimated separately per session.

In the context of this thesis, the VEM solution of this extension has been implemented to allow the fast joint analysis of multiple fMRI sessions. The implementation in VEM simplifies considerably the modelling with respect to the MCMC version. Experimental results showed that this approach is not very different from doing the posterior average of the independent single-session JDE results. However, analysing the data altogether means a higher computational load and a lower potential for parallelization. Moreover, if there is a very noisy session we might want to discard it and it will be easier to identify a noisy session by doing the independent single-session analysis. This concern lead us to the proposal of an heteroscedastic multiple-session extension of the VEM JDE approach. It consists in modelling a variable error instead of having it fixed. This amounts to considering

$$\boldsymbol{a}^{m,s} = \bar{\boldsymbol{a}}^m + \epsilon^s \quad \text{with} \quad \epsilon^s \sim \mathcal{N}(\mu_s, v_s \boldsymbol{I})$$

The response level mean remains independent conditionally upon $\boldsymbol{q}^m$:

$$\forall m, \;\; p(\bar{\boldsymbol{a}}^m \mid \boldsymbol{q}^m; \boldsymbol{\mu}^m, \boldsymbol{v}^m) = \prod_j p(\bar{\boldsymbol{a}}_j^m \mid q_j^m; \boldsymbol{\mu}^m, \boldsymbol{v}^m).$$

The goal is to be able to weight the importance of the different sessions so as to diminish the contribution of a potential noisy session to the final parameter estimates. Variance $v_s$ will be the responsible of this weighting and will modulate the contribution of $\mu^m$ and $\mu_s$ to the session-dependent estimate of the amplitude values.

For this extension, we propose a prior density for $\boldsymbol{a}$:

$$p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v}) = \prod_{m=1}^{M} \prod_{j \in \mathcal{J}} \prod_{s \in S} \prod_{i=0}^{1} \mathcal{N}(a_j^{m,s}; \mu_i^m + \mu_s, v_i^m + v_s)^{\mathbb{1}(q_j^m = i)}$$

where $\boldsymbol{\mu}^m$ and $\boldsymbol{v}^m$ are the GMM parameters. The rest of the prior distributions are the same as in the single session model already presented. The full joint model becomes:

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta}) &= p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}; \boldsymbol{\ell}, \boldsymbol{\Gamma}) \; p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v}) \; p(\boldsymbol{h}; v_h) \; p(\boldsymbol{q}; \beta) \; p(\boldsymbol{\theta}) \\ &= \prod_j \prod_s p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}; \boldsymbol{\ell}_j^s, \boldsymbol{\Gamma}_j^s) \\ &\quad \prod_m p(\boldsymbol{a}_j^{m,s} \mid \boldsymbol{q}_j^m; \mu^m + \mu_s, v^m + v_s) \\ &\quad p(\boldsymbol{q}_j; \beta) \; p(\boldsymbol{h}; v_h) \; p(\boldsymbol{\theta}) \end{aligned}$$

The complete model can be found in appendix B. Validation on simulated and real data needs to be done to assess the model expected behaviour.

## 4.9 Discussion

In this chapter, we have introduced Bayesian models and MCMC and VEM approaches. We have also presented the application of these methods for BOLD fMRI analysis. Two contributions have been presented:

- An hemodynamically informed parcellation of fMRI data.
- A heteroscedastic multiple-session extension of the VEM JDE algorithm.

In the next chapters, the methods presented are used in the analysis of functional ASL data.

**Outcome of this chapter:**

A. Frau-Pascual, T. Vincent, F. Forbes, and P. Ciuciu. *"Hemodynamically informed parcellation of cerebral fMRI data"*. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2079–2083. IEEE, 2014a.

# 5 Bayesian analysis of functional ASL data: MCMC solution

This chapter introduces a Bayesian model for the analysis of functional ASL (fASL). It is based on the previously presented Joint Detection Estimation (JDE) model for BOLD signal. As in the BOLD JDE, ASL JDE, first presented in [Vincent et al., 2013a], allows the estimation of a hemodynamic[1] response function (HRF) and a perfusion response function (PRF) per parcel along with the activation values of each voxel. Although the ASL JDE in [Vincent et al., 2013a] provides a good estimate of the HRF, the PRF estimation remains much more difficult because of the noisier nature of the perfusion component within the ASL signal. The whole perfusion component has a variation of around 1% [Golay et al., 2004, Liu and Brown, 2007] due to the fact that it is measured making the difference of control and tag images. The task-related perfusion can vary up to a 70%, which is a 1.7% variation with respect to the ASL signal. In this chapter, we propose to rely on physiological models to derive a linear operator that can be used as a tractable functional link between perfusion and hemodynamics components within the ASL signal and to exploit this link as prior knowledge for the accurate and robust recovery of the PRF shape in functional ASL data analysis. This way, we refine the separate estimation of the response functions performed in [Vincent et al., 2013b,a] by taking physiological information into consideration.

We first introduce the signal model for ASL, then the Bayesian model along with the prior distributions on the parameters, and finally we present the MCMC optimization strategy to estimate the unknown quantities in JDE. The outcome of this chapter was presented in October 2014 at the workshop in Bayesian and grAphical Models for Biomedical Imaging (BAMBI), a satellite event of the MICCAI conference.

[1] Note here that hemodynamic and perfusion response functions are both hemodynamic responses. In this work, this notation is used to be in line with the literature in BOLD fMRI. However, one can refer to the hemodynamic response as the BOLD response to remove any ambiguity.

**Outcome of this chapter:**

A. Frau-Pascual, T. Vincent, J. Sloboda, P. Ciuciu, and F. Forbes. *"Physiologically informed bayesian analysis of ASL fMRI data"*. In Bayesian and grAphical Models for Biomedical Imaging, pages 37–48. Springer, 2014.

## 5.1 The functional ASL signal model

The ASL Joint Detection Estimation (JDE) model was proposed in [Vincent et al., 2013b,a] and it assumes, as in BOLD JDE, the brain volume to be partitioned into a number of functionally homogeneous parcels, each of which gathers signals which share the same response shapes. In a given parcel $\mathcal{P}$, the generative model for ASL time series considering $M$ experimental conditions (see also figure 5.1) reads $\forall\, j \in \mathcal{P}$:

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \underbrace{a_j^m \boldsymbol{X}^m \boldsymbol{h}}_{(a)} + \underbrace{c_j^m \boldsymbol{W} \boldsymbol{X}^m \boldsymbol{g}}_{(b)} + \underbrace{\alpha_j \boldsymbol{w}}_{(c)} + \underbrace{\boldsymbol{P}\boldsymbol{\ell}_j}_{(d)} + \underbrace{\boldsymbol{b}_j}_{(e)} \tag{5.1}$$

ASL time series are measured at times $(t_n)_{n=1:N}$ where $N$ is the number of scans, $t_n = n\text{TR}$, and TR the repetition time. The signal is decomposed into task-related (a) hemodynamics and (b) perfusion components given by the first two terms respectively; (c) a perfusion baseline term $\alpha_j \boldsymbol{w}$ which completes the modelling of the perfusion component; (d) a drift component $\boldsymbol{P}\boldsymbol{\ell}_j$ already considered in the BOLD JDE [Vincent et al., 2010]; and (e) a noise term.

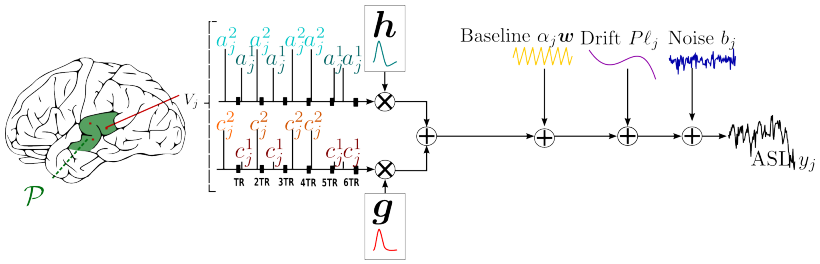



Figure 5.1: Forward model of JDE.

Acquiring ASL fMRI data consists in consecutively performing alternating measurements of control and magnetically tagged images. The hemodynamics component is present in both images, whereas the perfusion component comes from the subtraction of both control and tagged images. The hemodynamics component in ASL is noisier compared to a standard BOLD fMRI acquisition because the echo time (TE) used for ASL is not optimal for the BOLD contrast [Ogawa et al., 1993, Tjandra et al., 2005]. The control/tag effect is modelled with the use of matrix $\boldsymbol{W}$.

More specifically, we further describe each signal part below.

**(a) The hemodynamics component** represents the variation of the hemodynamics (or BOLD) signal when there is task-related activity. $\boldsymbol{h} \in \mathbb{R}^{D+1}$ represents the unknown HRF shape with size $D+1$ and assumed identical for all voxels in a parcel $\mathcal{P}$. The magnitude of activation or hemodynamics response levels are $\boldsymbol{a} = \left\{a_j^m, j \in \mathcal{P}, m = 1:M\right\}$.

**(b) The perfusion component** represents the variation of the perfusion from the baseline when there is task-related activity. $\boldsymbol{g} \in \mathbb{R}^{D+1}$ represents the unknown PRF shape, with size $D+1$ and constant within $\mathcal{P}$. The perfusion response levels are $\boldsymbol{c} = \left\{c_j^m, j \in \mathcal{P}, m = 1:M\right\}$. $\boldsymbol{W}$ models the control/tag effect in the perfusion component, further explained below.

**(a-b)** Considering $\Delta t < TR$ the sampling period of $\boldsymbol{h}$ and $\boldsymbol{g}$, whose temporal resolution is assumed to be the same, $\boldsymbol{X}^m = \{x^{n-d\Delta t}, n = 1:N, d =$

$0 : D\}$ is a binary matrix that encodes the lagged stimulus onset. $\boldsymbol{X}^m$ is the linear function that associates $\boldsymbol{h}$ to $\boldsymbol{x}^m * \boldsymbol{h}$. The hemodynamics (HRLs) and perfusion (PRLs) response levels (resp. $\boldsymbol{a}$ and $\boldsymbol{c}$) are assumed independent conditionally to a common binary hidden Markov random field $\boldsymbol{q}$ that introduces spatial correlation across neighbouring voxels, as in [Vincent et al., 2013b,a]. This Markov random field has been introduced in the previous chapter for the BOLD JDE.

**(c) The drift term** allows to account for a potential low frequency temporal scanner drift and any other nuisance effect (e.g. slow motion parameters). Matrix $\boldsymbol{P} = \left[\boldsymbol{p}_1, \ldots, \boldsymbol{p}_F\right]$ of size $N \times F$ with orthonormal columns (*i.e.*, $\boldsymbol{P}^{\mathrm{t}}\boldsymbol{P} = \boldsymbol{I}_F$). We denote by $\boldsymbol{\ell} = \left\{\boldsymbol{\ell}_j, j \in \mathcal{J}\right\}$ the set of low frequency drifts. Each $\boldsymbol{\ell}_j$ is a $L$-dimensional vector of coefficients to be estimated $\boldsymbol{\ell}_j = (\ell_{f,j}, f = 1 : F)^{\mathrm{t}}$.

**(b-d) The control/tag vector $\boldsymbol{w}$** ($N$-dimensional) encodes the difference in magnetization signs between control and tagged ASL volumes. $w_{t_n} = 1/2$ if $t_n$ is even (control volume) and $w_{t_n} = -1/2$ otherwise (tagged volume), and $\boldsymbol{W} = \mathrm{diag}(\boldsymbol{w})$ is the diagonal matrix with $\boldsymbol{w}$ as diagonal entries.

**(d) The perfusion baseline** is encoded by $\alpha_j$ at voxel $j$.

**(e) The noise term** encodes the noise in the signal. Here it is assumed i.i.d Gaussian. A straightforward extension to AR(1) noise process can be derived using [Vincent et al., 2010, Chaari et al., 2013].

The simultaneous estimation of the HRF and PRF response functions is difficult in this framework because the signal is noisy and because the alternation of control and tag images introduces variability between consecutive time points. See the results in [Vincent et al., 2013b,a]. The perfusion response function estimation is much more difficult because the perfusion effect is quite small compared to the rest of the signal and it comes from the subtraction of control and tag images, with the increase in temporal variability that comes with it. The use of prior knowledge on the PRF could appropriately inform the estimation of this function. For this reason, we derive a physiologically informed relationship between the perfusion and hemodynamic response functions from physiological models. We use the extended Balloon model previously introduced in chapter 2.

## 5.2 Physiological linear relationship between response functions

The extended Balloon model (see chapter 2) models the variations in flow inducing signal $\boldsymbol{\psi}$, local blood flow $\boldsymbol{f}_{in}$, local capillary volume $\boldsymbol{\nu}$, and deoxyhemoglobin concentration $\boldsymbol{\xi}$ as a response to neuronal activity $\boldsymbol{u}$. Recall that the interactions of these four state variables are described by the system of differential equations:

$$\begin{cases} \frac{d\boldsymbol{f}_{in}(t)}{dt} = \boldsymbol{\psi}(t) \\ \frac{d\boldsymbol{\psi}(t)}{dt} = \eta u(t) - \frac{\boldsymbol{\psi}(t)}{\tau_\psi} - \frac{\boldsymbol{f}_{in}(t)-1}{\tau_f} \\ \frac{d\boldsymbol{\xi}(t)}{dt} = \frac{1}{\tau_m}\left(\boldsymbol{f}_{in}(t)\frac{1-(1-E_0)^{1/\boldsymbol{f}_{in}(t)}}{E_0} - \boldsymbol{\xi}(t)\boldsymbol{\nu}(t)^{\frac{1}{\bar{w}}-1}\right) \\ \frac{d\boldsymbol{\nu}(t)}{dt} = \frac{1}{\tau_m}\left(\boldsymbol{f}_{in}(t) - \boldsymbol{\nu}(t)^{\frac{1}{\bar{w}}}\right) \end{cases} \tag{5.2}$$

with initial conditions $\boldsymbol{\psi}(0) = 0, \boldsymbol{f}_{in}(0) = \boldsymbol{\nu}(0) = \boldsymbol{\xi}(0) = 1$. The system depends on 6 physiological parameters that have already been described in section 2.5.1. The local blood flow $\boldsymbol{f}_{in}(t)$ is the normalized perfusion, with initial value 1. We can then define the perfusion response function (PRF), namely $\boldsymbol{g}(t)$, as the task-related perfusion variation from the baseline. Therefore, we consider $\boldsymbol{g}(t) = \boldsymbol{f}_{in}(t) - 1$ and $\boldsymbol{g}(0) = 0$. We can consider it as one of the state variables $\{\boldsymbol{\psi}, \boldsymbol{g}, 1-\boldsymbol{\nu}, 1-\boldsymbol{\xi}\}$.

The hemodynamic response function (HRF) is given by $\boldsymbol{h}(t)$ when $\boldsymbol{u}(t)$ is an impulse function, and [Buxton et al., 1998b] proposed a link of the BOLD response $\boldsymbol{h}(t)$ to the physiological quantities considering intra- and extra-vascular components [Buxton et al., 1998b, Obata et al., 2004]:

$$\boldsymbol{h}(t) = V_0 \left[ k_1(1-\boldsymbol{\xi}(t)) + k_2 \left(1 - \frac{\boldsymbol{\xi}(t)}{\boldsymbol{\nu}(t)}\right) + k_3(1-\boldsymbol{\nu}(t)) \right] \tag{5.3}$$

where $k_1$, $k_2$ and $k_3$ are scanner-dependent constants and $V_0$ is the resting blood volume fraction. In the following we will drop the time index $t$ and use variables in their discretized vector form.

From these equations, we derive an approximate relationship between the PRF ($\boldsymbol{g}$) and the HRF ($\boldsymbol{h}$). We can obtain it by linearizing the system of equations (see complete linearization in appendix C). Equation (5.3) can first be linearized into:

$$\boldsymbol{h} = V_0[(k_1+k_2)(1-\boldsymbol{\xi}) + (k_3-k_2)(1-\boldsymbol{\nu})] \,. \tag{5.4}$$

We then linearize the system (5.2) around the resting point $\{\boldsymbol{\psi}, \boldsymbol{g}, 1-\boldsymbol{\nu}, 1-\boldsymbol{\xi}\} = \{\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}\}$ as in [Khalidov et al., 2011]. From this linearization, denoting by $\mathcal{D}$ the first order differential operator and $\boldsymbol{I}$ the identity matrix, we obtain:

$$\begin{cases} \mathcal{D}\{\boldsymbol{g}\} = -\boldsymbol{\psi} \\ \left(\mathcal{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)\{1-\boldsymbol{\nu}\} = -\frac{1}{\tau_m}\boldsymbol{g} \\ \left(\mathcal{D} + \frac{\boldsymbol{I}}{\tau_m}\right)\{1-\boldsymbol{\xi}\} = -\left(\gamma\boldsymbol{I} - \frac{1-\tilde{w}}{\tilde{w}\tau_m^2}\left(\mathcal{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right)\boldsymbol{g} \end{cases} \tag{5.5}$$

where $\gamma = \frac{1}{\tau_m}\left(1 + \frac{(1-E_0)\ln(1-E_0)}{E_0}\right)$. It follows a linear link between $\boldsymbol{h}$ and $\boldsymbol{g}$ that we write as $\boldsymbol{g} = \boldsymbol{\Omega}\boldsymbol{h}$ where:

$$\boldsymbol{\Omega} = V_0^{-1}\left(-(k_1+k_2)\gamma\boldsymbol{B} + (k_1+k_2)\frac{1-\tilde{w}}{\tilde{w}\tau_m^2}\boldsymbol{B}\boldsymbol{A} - \frac{k_3-k_2}{\tau_m}\boldsymbol{A}\right)^{-1} \tag{5.6}$$

$$\text{with } \boldsymbol{A} = \left(\mathcal{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1} \text{ and } \boldsymbol{B} = \left(\mathcal{D} + \frac{\boldsymbol{I}}{\tau_m}\right)^{-1} \tag{5.7}$$

The entire derivation of $\boldsymbol{\Omega}$ can be found in appendix C.

Using values of physiological constants as proposed in [Friston et al., 2000], Fig. 5.2 shows the HRF and PRF results that we get ($\boldsymbol{h}_{lin}$, $\boldsymbol{g}_{lin}$) by applying the linear operator to physiologically generated PRF ($\boldsymbol{g}_{physio}$) or HRF ($\boldsymbol{h}_{physio}$): $\boldsymbol{h}_{lin} = \boldsymbol{\Omega}^{-1}\boldsymbol{g}_{physio}$ or $\boldsymbol{g}_{lin} = \boldsymbol{\Omega}\boldsymbol{h}_{physio}$ compared to these physiologically generated $\boldsymbol{h}_{physio}$ and $\boldsymbol{g}_{physio}$ functions, computed by using the physiological model differential equations. The error of this linearization is $\frac{\|\boldsymbol{h}_{physio}-\boldsymbol{h}_{lin}\|}{\sqrt{dt}} = 0.087$ and $\frac{\|\boldsymbol{g}_{physio}-\boldsymbol{g}_{lin}\|}{\sqrt{dt}} = 0.099$ for a $dt = 0.5$. There is a time-to-peak (TTP) difference of 0.4 s in the case of

the PRFs and 0.5 s in the case of the HRFs. Note that, although TTP values are not exact, the linear operator maintains the shape of the functions and satisfyingly captures the main features of the two responses. We considered a finer temporal resolution than TR for $\mathbf{\Omega}$.

This linear operator gives us a new tractable tool for analyzing the ASL signal. Although this link is subject to caution due to its linear nature, as it is only valid around its development point and incurs approximation errors, we can nevertheless consider it in a prior distribution of the perfusion response function.

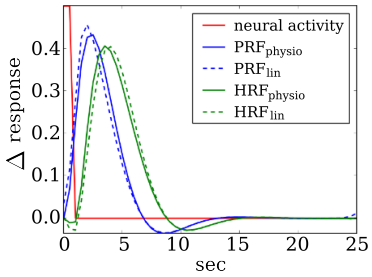



Figure 5.2: Physiological responses generated with the physiological model, using parameters proposed in [Friston et al., 2000]: neural activity $\boldsymbol{\psi}$, physiological ($\boldsymbol{h}_{physio}$ or $\text{HRF}_{physio}$) and linearized ($\boldsymbol{h}_{lin}$ or $\text{HRF}_{lin}$) HRFs, physiological ($\boldsymbol{g}_{physio}$ or $\text{PRF}_{physio}$) and linearized ($\boldsymbol{g}_{lin}$ or $\text{PRF}_{lin}$) PRFs.

## 5.3 Bayesian modelling

The joint probabilistic distribution for the model in equation 5.1 can be decomposed as follows:

$$p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta}) = p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{a} \mid \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{c} \mid \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta}) \cdot \\ \cdot\ p(\boldsymbol{h} \mid \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{g} \mid \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{\theta})$$

with $\boldsymbol{\theta}$ being the parameters that need to be estimated: perfusion, drift and noise parameters, as well as the parameters included with the prior distributions that are presented below. Here $\boldsymbol{y}$ is assumed independent of $\boldsymbol{q}$ given $\boldsymbol{a}$, $\boldsymbol{h}$, $\boldsymbol{c}$, $\boldsymbol{g}$. $\boldsymbol{a}$ and $\boldsymbol{c}$ are assumed independent of $\boldsymbol{h}$ and $\boldsymbol{g}$, conditionally to $\boldsymbol{q}$. $\boldsymbol{h}$ and $\boldsymbol{g}$ are assumed independent of $\boldsymbol{q}$ because HRF and PRF do not depend on the activation states of the voxels, just their amplitudes. This is expressed by the factorization of the probabilities in the above formula.

### 5.3.1 Likelihood

According to this model, the likelihood reads:

$$p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\alpha}; \boldsymbol{\ell}, \boldsymbol{\Gamma}) = \prod_{j \in \mathcal{J}} p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}; \alpha_j, \boldsymbol{\ell}_j, \boldsymbol{\Gamma}_j) \tag{5.8}$$

$$\text{with} \quad p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}; \alpha_j, \boldsymbol{\ell}_j, \boldsymbol{\Gamma}_j) \sim \mathcal{N}\left(\bar{\boldsymbol{y}}_j, \boldsymbol{\Gamma}_j^{-1}\right) \tag{5.9}$$

$$\text{and} \quad \bar{\boldsymbol{y}}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h} + c_j^m \boldsymbol{W} \boldsymbol{X}^m \boldsymbol{g} + \boldsymbol{w} \alpha_j + \boldsymbol{P} \boldsymbol{\ell}_j \tag{5.10}$$

### 5.3.2 Prior distribution of the response functions

Following [Vincent et al., 2010], the HRF and PRF shapes are assumed to follow a prior multivariate Gaussian distribution whose covariance matrices $\boldsymbol{\Sigma}_h$ and $\boldsymbol{\Sigma}_g$ impose temporal smoothness as in section 3.3.2. They embody a constraint on the second order derivative so as to penalise high variations between neighbouring points and have a smooth temporal shape. As this modelling remains non-parametric, any shape is allowed and it is worth noting that the two HRF and PRF shapes may differ. In [Vincent et al., 2013a], the priors used for HRF/PRF are

$$\boldsymbol{h} \sim \mathcal{N}(0, v_h \boldsymbol{\Sigma}_h) \tag{5.11}$$

$$\boldsymbol{g} \sim \mathcal{N}(0, v_g \boldsymbol{\Sigma}_g) \tag{5.12}$$

where $v_h$ and $v_g$ are unknown variances, and $\boldsymbol{\Sigma}_h = \boldsymbol{\Sigma}_g = \boldsymbol{R} = \left(\frac{\boldsymbol{D}_2^t \boldsymbol{D}_2}{(\Delta t)^4}\right)^{-1}$. See section 3.3.2 for the definition of $\boldsymbol{D}_2$[2] and the derivation of $\boldsymbol{R}$. Priors on $v_h$ and $v_g$ can also be introduced. In MCMC, we use non-informative Jeffreys priors: $p(v_h) = v_h^{-1/2}$, $p(v_g) = v_g^{-1/2}$. In VEM, since $v_h$ and $v_g$ are positive, suitable priors can be exponential distributions with means $-\lambda_h$ and $-\lambda_g$ respectively as in [Chaari et al., 2013]: $p(v_h; \lambda_h) = \lambda_h \exp(-\lambda_h)$ and $p(v_g; \lambda_g) = \lambda_g \exp(-\lambda_g)$.

We can also enforce a prior mean shape by using a different mean prior: a physiological shape in HRF and PRF, or just in the PRF, or the use of a physiological link between HRF and PRF. In the latter case, a prior on the PRF would regularize the estimation by using the already estimated HRF to construct a candidate for the PRF. Part of this work consisted in investigating the use of different priors on the response functions, and different ways to introduce them. We have focused on the derived physiological link $\boldsymbol{\Omega}$ between both responses (see Appendix C for its derivation) as prior knowledge in $\boldsymbol{g}$, to compensate for the lower SNR of the perfusion component of the signal. Constraints on the shape of the responses (e.g. norm constraint or positivity) have also been considered.

[2] Recall that

$$\boldsymbol{D}_2 = \begin{pmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 1 & -2 & 1 \\ 0 & \dots & 0 & 1 & -2 \end{pmatrix}$$

### Stochastic Ω constraint

The first prior investigated encodes a hierarchical relationship between $\boldsymbol{g}$ and $\boldsymbol{h}$. We use the equality $\boldsymbol{g} = \boldsymbol{\Omega h}$ for the prior mean of $\boldsymbol{g}$:

$$\boldsymbol{h} \sim \mathcal{N}(0, v_h \boldsymbol{\Sigma}_h) \tag{5.13}$$

$$\boldsymbol{g}|\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{\Omega h}, v_g \boldsymbol{\Sigma}_g) \tag{5.14}$$

For $\boldsymbol{h}$, we use the covariance matrix $\boldsymbol{\Sigma}_h = \boldsymbol{R}$. $\boldsymbol{\Sigma}_g$ can add smoothness if we use the prior $\boldsymbol{\Sigma}_g = \boldsymbol{R}$ (regularized physiological prior) or not if we use the identity matrix $\boldsymbol{\Sigma}_g = \boldsymbol{I}_{D+1}$ (not regularized physiological prior).

If we consider a joint prior with a single variance parameter, the relation of $\boldsymbol{h}$ and $\boldsymbol{g}$ becomes:

$$\boldsymbol{h} \sim \mathcal{N}(0, v_{gh} \boldsymbol{\Sigma}_h) \tag{5.15}$$

$$\boldsymbol{g}|\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{\Omega h}, v_{gh} \boldsymbol{\Sigma}_g) \tag{5.16}$$

A joint variance $v_{gh}$ amounts to setting $v_h = v_g$, decreasing the number of parameters to estimate or set. This can be potentially interesting when the values have to be set because they are not well estimated. We can find the joint multivariate normal distribution of $(\boldsymbol{h}, \boldsymbol{g})$ from the relationship $p(\boldsymbol{h})p(\boldsymbol{g}|\boldsymbol{h})$ as

$$\begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{g} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, v_{gh} \begin{bmatrix} \boldsymbol{\Sigma}_h & \boldsymbol{\Sigma}_h \boldsymbol{\Omega}^T \\ \boldsymbol{\Omega \Sigma}_h & \boldsymbol{\Sigma}_g + \boldsymbol{\Omega \Sigma}_h \boldsymbol{\Omega}^T \end{bmatrix} \right) \tag{5.17}$$

When we include $\boldsymbol{\Omega}$, we are considering some variability around the prior mean $\boldsymbol{\Omega h}$. Therefore, any approximation error coming from the derivation of $\boldsymbol{\Omega}$ can be compensated for. However, imposing a relationship between $\boldsymbol{h}$ and $\boldsymbol{g}$ also affects the computation of $\boldsymbol{h}$. See for example in appendix D that the expression (D.17) in the computation of $p(\boldsymbol{h}|\boldsymbol{y}, ..., \boldsymbol{\theta})$ contains a term with $\boldsymbol{g}$.

### Deterministic $\Omega$ constraint

We also investigate the simplest option: including directly the relationship $\boldsymbol{g} = \boldsymbol{\Omega h}$. $\boldsymbol{g}$ is directly obtained from $\boldsymbol{h}$. Here the same "classical" regularized prior is used for the HRF. The priors read:

$$\boldsymbol{h} \sim \mathcal{N}(0, v_h \boldsymbol{\Sigma}_h) \tag{5.18}$$

$$\boldsymbol{g} = \boldsymbol{\Omega h} \tag{5.19}$$

where: $\boldsymbol{\Sigma}_h = \boldsymbol{R}$. We reformulate the forward model to inject the relation (5.19). $\forall\, j \in \mathcal{J}$,

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \left[ c_j^m \boldsymbol{W} \boldsymbol{X}^m \boldsymbol{\Omega} + a_j^m \boldsymbol{X}^m \right] \boldsymbol{h} + \boldsymbol{P} \boldsymbol{\ell}_j + \alpha_j \boldsymbol{w} + \boldsymbol{b}_j \tag{5.20}$$

### Hierarchical model

Here we consider a hierarchical model in which both response functions are conditional to a variable $\boldsymbol{h}_t$.

$$\boldsymbol{h}_t \sim \mathcal{N}(\boldsymbol{h}_{can}, v_{h_t} \boldsymbol{\Sigma}_{h_t}) \tag{5.21}$$

where $\boldsymbol{\Sigma}_{h_t} = \boldsymbol{R}$ and $v_{h_t}$ contain the regularising terms to obtain a smooth $\boldsymbol{h}_t$. $\boldsymbol{h}_t$ is in some sense the "true" HRF, and we have access to the noisy version of the true HRF:

$$\boldsymbol{h} | \boldsymbol{h}_t \sim \mathcal{N}(\boldsymbol{h}_t, v_h \boldsymbol{\Sigma}_h) \tag{5.22}$$

$$\boldsymbol{g} | \boldsymbol{h}_t \sim \mathcal{N}(\boldsymbol{\Omega h}_t, v_g \boldsymbol{\Sigma}_g) \tag{5.23}$$

where $v_h$ is small (good precision on $\boldsymbol{h}$) and $v_g$ is large (less precision on $\boldsymbol{g}$). We use $\boldsymbol{\Sigma}_h = \boldsymbol{I}_{D+1}$ and $\boldsymbol{\Sigma}_g = \boldsymbol{I}_{D+1}$ because we are conditional to the "true" HRF, which will be smooth.

The random variables $\boldsymbol{h}$ and $\boldsymbol{g}$ are independent but conditional to $\boldsymbol{h}_t$, so that when you estimate $\boldsymbol{h}$ in the MCMC algorithm the lower precision on $\boldsymbol{g}$ should not influence $\boldsymbol{h}$ too much. We hope that the effect of $\boldsymbol{g}$ on $\boldsymbol{h}_t$ is lower by the fact that $v_g$ is large because using a large $v_g$ is equivalent to saying that the observation is not as reliable as with a small $v_h$.

### Balloon model prior

The Balloon model can be used to generate an $\boldsymbol{h}_{balloon}$ and $\boldsymbol{g}_{balloon}$ that can be used as prior mean in the response functions.

$$p(\boldsymbol{h}; v_h) \sim \mathcal{N}(\boldsymbol{h}_{balloon}, v_h \boldsymbol{\Sigma}_h)$$

$$p(\boldsymbol{g}; v_g) \sim \mathcal{N}(\boldsymbol{g}_{balloon}, v_g \boldsymbol{\Sigma}_g)$$

Usually we use the smooth prior covariance $\boldsymbol{\Sigma}_h = \boldsymbol{\Sigma}_g = \boldsymbol{R}$ already explained. This prior is not used in this chapter, but will be used in chapter 7.

## 5.3.3 Prior distributions for the rest of parameters

Here we present the rest of the prior distributions used for each one of the parameters of the model.

### Response Levels

For a given experimental condition $m$, the hemodynamic (HRLs) and perfusion (PRLs) response levels are assumed to follow spatial Gaussian mixture models with different means and variances but governed by the same hidden Markov random field encoding voxels activation states. Indeed, the actual hemodynamics and perfusion response levels are likely to differ as they do not reflect the same physiological measure: CBF versus a combination of CBV, CBF and $CMRO_2$. However, the activation states are assumed to be the same for both response levels. The aim is to model a coupling between the perfusion and hemodynamics components, allowing the extraction of the perfusion component by taking advantage of the higher SNR in the hemodynamics component.

Mixture models segregate active voxels from non-active ones. Let $q_j^m$ be the allocation variable that states whether voxel $j$ is active ($q_j^m = 1$) or non-active ($q_j^m = 0$) in response to the $m$th experimental condition. Spatial correlation is directly incorporated in the probabilities of activation through a *hidden Potts* field on the allocation variables $\boldsymbol{q}^m$. Here, the proportion of voxels for the different classes is not explicit.

We assume that different types of stimulus (experimental conditions) induce statistically independent response levels. HRLs and PRLs are a priori spatially independent conditionally on $\boldsymbol{q}$. Let $\mathbb{1}(K) = 1$ if $K$ is true, 0 otherwise.

$$
\begin{aligned}
p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v}) &= \prod_{m=1}^{M} \prod_{j \in \mathcal{J}} \prod_{i \in \{0,1\}} \mathcal{N}(a_j^m; \mu_i^m, v_i^m)^{\mathbb{1}(q_j^m = i)} \\
&\propto \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{i \in \{0,1\}} \left( (v_i^m)^{-\frac{1}{2}} \exp\left( \frac{-(a_j^m - \mu_i)^2}{2v_i} \right) \right)^{\mathbb{1}(q_j^m = i)}
\end{aligned}
$$

where we used the notations $\boldsymbol{\mu}^m$ and $\boldsymbol{v}^m$ for $\boldsymbol{\mu}^m = \left[\mu_0^m, \mu_1^m\right]^{\mathrm{t}}$ and $\boldsymbol{v}^m = \left[v_0^m, v_1^m\right]^{\mathrm{t}}$. The unknown parameters are the Gaussian parameters $\boldsymbol{\mu}$ and $\boldsymbol{v}$. The same way, for the PRLs the conditional prior pdf:

$$
\begin{aligned}
p(\boldsymbol{c} \,|\, \boldsymbol{q}; \boldsymbol{\eta}, \boldsymbol{\rho}) &= \prod_{m=1}^{M} \prod_{j \in \mathcal{J}} \prod_{i \in \{0,1\}} \mathcal{N}(c_j^m; \eta_i^m, \rho_i^m)^{\mathbb{1}(q_j^m = i)} \\
&\propto \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{i \in \{0,1\}} \left( (\rho_i^m)^{-\frac{1}{2}} \exp\left( \frac{-(c_j^m - \eta_i)^2}{2\rho_i} \right) \right)^{\mathbb{1}(q_j^m = i)}
\end{aligned}
$$

where we used the notations $\eta_m$ and $\rho^m$ for $\eta_m = \left[\eta_0^m, \eta_1^m\right]^{\mathrm{t}}$ and $\rho^m = \left[\rho_0^m, \rho_1^m\right]^{\mathrm{t}}$. The unknown parameters are the Gaussian parameters $\boldsymbol{\eta}$ and $\boldsymbol{\rho}$.

### Activation states (labels)

As proposed in [Vincent et al., 2010], a Potts field is expressed to model spatial correlation in a parcel. We assume prior independence between experimental conditions regarding the hidden Potts fields:

$$
p(\boldsymbol{q}; \boldsymbol{\beta}) = \prod_{m=1}^{M} p(\boldsymbol{q}^m; \beta^m)
$$

where $\boldsymbol{\beta}$ is the amount of spatial regularization. Here $p(\boldsymbol{q}^m; \beta^m)$ is a 2-class Potts model with interaction parameter $\beta^m$ and no external field.

$$p(\boldsymbol{q}^m; \beta^m) = Z(\beta^m)^{-1} \exp\left(\frac{1}{2}\beta^m \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{N}(j)} \mathbb{1}(q_j^m = q_k^m)\right) \tag{5.24}$$

The Potts field is defined on the neighborhood graph of each parcel. $Z(\beta^m)$ is the normalizing constant or the partition function of the Potts field. It is independent of a specific value of $\boldsymbol{q}$, as it is the marginalisation over all possible values of $\boldsymbol{q}$. $\boldsymbol{\beta} = \{\beta^m, m = 1 : M\}$ are unknown parameters. A prior on $\boldsymbol{\beta}$ can also be added in the VEM case: $p(\beta^m; \lambda_\beta) = \lambda_\beta \exp(-\lambda_\beta)$.

### Perfusion Baseline

We assume a Gaussian prior for the perfusion baseline: $\alpha_j \sim \mathcal{N}(0, v_\alpha)$

$$p(\alpha|v_\alpha) \propto v_\alpha^{-\frac{J}{2}} \exp\left(\sum_{j=1}^{J} \frac{-\alpha_j^2}{2v_\alpha}\right) \tag{5.25}$$

### Drift Coefficients

We assume a Gaussian prior for the drift coefficients: $\boldsymbol{\ell}_j \sim \mathcal{N}(0, v_\ell I_F)$

$$p(\boldsymbol{\ell}|v_\ell) \propto v_\ell^{-JF/2} \exp\left(\sum_{j=1}^{J} -\frac{1}{2}\boldsymbol{\ell}_j^t (v_\ell I_F)^{-1} \boldsymbol{\ell}_j\right)$$

### Hyper-parameters

Non-informative Jeffreys Priors[3] are adopted for the hyper parameters:

$$\tilde{\boldsymbol{\theta}} = \{v_h, v_g, v_\alpha, v_\ell, v_b, \boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\eta}, \boldsymbol{\rho}\}$$

$$\begin{aligned}
\text{Thus } \; & p(v_h) = v_h^{-1/2} \quad p(v_g) = v_g^{-1/2} \quad p(v_{b_j}) = v_{b_j}^{-1/2} \\
& p(v_\ell) = v_\ell^{-1/2} \quad p(v_\alpha) = v_\alpha^{-1/2} \quad p(\boldsymbol{\mu}_i) = 1 \\
& p(\boldsymbol{v}_i) = \boldsymbol{v}_i^{-1/2} \quad p(\boldsymbol{\eta}_i) = 1 \quad p(\boldsymbol{\rho}_i) = \boldsymbol{\rho}_i^{-1/2}
\end{aligned}$$

All the prior parameters will be estimated in the model

$$\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\eta}, \boldsymbol{\rho}, v_h, v_g, \boldsymbol{\beta}\}.$$

Considering all these parameters, the full joint model becomes:

$$\begin{aligned}
p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta}) \; = \; & p(\boldsymbol{y} \,|\, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\ell}, \boldsymbol{\Gamma}) \cdot \\
& \cdot \; p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v}) \; p(\boldsymbol{c} \,|\, \boldsymbol{q}; \boldsymbol{\eta}, \boldsymbol{\rho}) \; p(\boldsymbol{q}; \boldsymbol{\beta}) \cdot \\
& \cdot \; p(\boldsymbol{h}; v_h) \; p(\boldsymbol{g}; v_g) \; p(\boldsymbol{\theta})
\end{aligned}$$

[3] Jeffrey's prior is a non-informative prior distribution defined using Fisher information $\mathcal{I}_f(\theta)$:

$$\begin{aligned}
p(\theta) &\propto \sqrt{\mathcal{I}_f(\theta)} \\
&= \sqrt{-\mathrm{E}_{p_x}\left[\frac{d^2 \log p(X|\theta)}{d\theta^2}\right]}
\end{aligned}$$

## 5.3.4 Posterior distribution

The posterior distribution can be computed from the prior densities and the likelihood of the model:

$$\begin{aligned}
p(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\ell}, \boldsymbol{\alpha}, v_b \,|\, \boldsymbol{y}) \propto \; & p(\boldsymbol{y} \,|\, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\ell}, \boldsymbol{\Gamma}) \cdot \\
& \cdot \; p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v}) \; p(\boldsymbol{c} \,|\, \boldsymbol{q}; \boldsymbol{\eta}, \boldsymbol{\rho}) \; p(\boldsymbol{q}; \boldsymbol{\beta}) \cdot \\
& \cdot \; p(\boldsymbol{h}; v_h) \; p(\boldsymbol{g}; v_g) \;\; p(\boldsymbol{\theta})
\end{aligned}$$

Using the expressions of the likelihood and prior distributions, we get a full posterior distribution that can be found in appendix D. The JDE ASL hierarchical graph model can be found in figure 5.3.

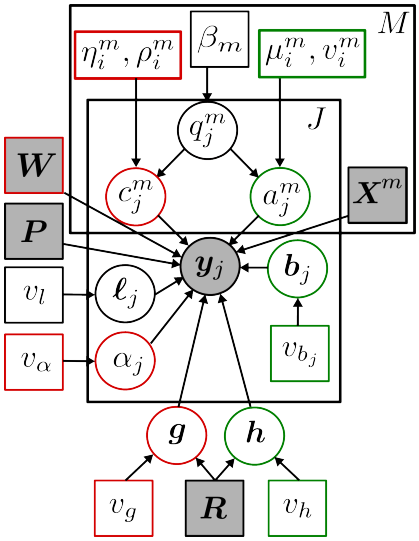



Figure 5.3: Hierarchical graph model for JDE ASL.

## 5.4 Estimation of the posterior probability distribution with MCMC

From the posterior conditional distributions of each variable $x \in \mathcal{X}$, with $\mathcal{X} = \{\boldsymbol{a}, \boldsymbol{q}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{\theta}\}$ samples are generated using a Gibbs sampling scheme and posterior mean (PM) estimates are computed from the samples according to: $\hat{x}^{\text{PM}} = \sum_{k=K_0}^{K_1} x^{(k)}/K, \quad \forall x \in \mathcal{X}$ where $K = K_1 - K_0 + 1$ and $K_0$ stands for the length of the burn-in period. The posterior conditional distributions $p(x^k \mid (\mathcal{X}\backslash x)^{k-1}, \boldsymbol{y})$ are computed at each iteration $k$ of the sampling scheme, considering the rest of the variables $\mathcal{X}\backslash x$ updated in last iteration $k-1$. The samples are drawn from the conditional posterior distributions. For the hyperparameter conditional posterior, let $\widetilde{\boldsymbol{\theta}}$ be the set of all hyperparameters except the one in question.

**Algorithm 1** Gibbs sampling for ASL JDE

1: **initialize**
2: **repeat**
3: **Sampling from posterior conditional distributions:**

$$
\begin{aligned}
(a_j^m)^{(k)} &\leftarrow p(a_j^m \mid q_j^m = i, \ldots) & & \\
(c_j^m)^{(k)} &\leftarrow p(c_j^m \mid q_j^m = i, \ldots) & (q_j^m)^{(k)} &\leftarrow p(q_j^m \mid \boldsymbol{y}, \ldots, \boldsymbol{\theta}) \\
\boldsymbol{h}^{(k)} &\leftarrow p(\boldsymbol{h} \mid \boldsymbol{y}, \ldots, \boldsymbol{\theta}) & \boldsymbol{g}^{(k)} &\leftarrow p(\boldsymbol{g} \mid \boldsymbol{y}, \ldots, \boldsymbol{\theta}) \\
\alpha_j^{(k)} &\leftarrow p(\alpha_j \mid \boldsymbol{y}_j, \ldots, \boldsymbol{\theta}) & \boldsymbol{\ell}_j^{(k)} &\leftarrow p(\boldsymbol{\ell}_j \mid \boldsymbol{y}_j, \ldots, \boldsymbol{\theta}) \\
\mu_i^{m(k)} &\leftarrow p(\mu_i^m \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) & v_i^{m(k)} &\leftarrow p(v_i^m \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) \\
\eta_i^{m(k)} &\leftarrow p(\eta_i \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) & \rho_i^{m(k)} &\leftarrow p(\rho_i \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) \\
v_h^{(k)} &\leftarrow p(v_h \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) & v_g^{(k)} &\leftarrow p(v_g \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) \\
v_\alpha^{(k)} &\leftarrow p(v_\alpha \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) & v_\ell^{(k)} &\leftarrow p(v_\ell \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) \\
v_{b_j}^{(k)} &\leftarrow p(v_{b_j} \mid \boldsymbol{y}, \ldots, \widetilde{\boldsymbol{\theta}}) & &
\end{aligned}
$$

4: **until** $k = N_{iterations}$
5: **Compute posterior means**

$$\hat{x}^{\text{PM}} = \frac{1}{K} \sum_{k=K_0}^{K_1} x^{(k)}, \quad \forall x \in \mathcal{X}$$

6: **RETURN** posterior mean estimates $\hat{x}^{\text{PM}} \quad \forall x \in \mathcal{X}$

The posterior conditional distributions and the whole model can be found in appendix D.

## 5.5 A 2-step MCMC inference procedure

When using the stochastic physiological prior, we are trying to influence the shape of the PRF using the HRF as prior information, but this also leads

to an influence of the PRF on the HRF (see expression (D.17) in appendix D). As a mid step in the analysis of this model, a 2-step procedure was developed to deal with the two components, hemodynamics and perfusion, separately. We first identify hemodynamics properties ($\hat{\boldsymbol{h}}$, $\hat{a}_j^m$), and then use the linear operator $\boldsymbol{\Omega}$ and the previously estimated hemodynamic properties to recover the perfusion component ($\hat{\boldsymbol{g}}$, $\hat{c}_j^m$). This way, we avoid any contaminating effect of $\boldsymbol{g}$ on the estimation of $\boldsymbol{h}$. Each step is based on a Gibbs sampling procedure as in [Vincent et al., 2013b,a].

### Hemodynamics estimation step $\mathcal{M}_1$

In a first step $\mathcal{M}_1$, our goal is to extract the hemodynamic components and the drift term from the ASL data. In the JDE framework (5.1), it amounts to initially considering the perfusion component as a generalized perfusion term, including perfusion baseline and event-related perfusion response. By grouping the perfusion terms involving $\boldsymbol{W}$, the generative model (5.1) for ASL time series can be equivalently written, as

$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{W}\left(\sum_{m=1}^{M} c_j^m \boldsymbol{X}^m \boldsymbol{g} + \alpha_j \mathbf{1}\right) + \boldsymbol{b}_j \tag{5.26}$$

where we consider $\alpha_j \boldsymbol{w} = \boldsymbol{W}\alpha_j \mathbf{1}$. Note that the hemodynamics components HRF $\boldsymbol{h}$ and the drift term $\boldsymbol{\ell}_j$ can be estimated first, by segregating them from a general perfusion term and a noise term. However, the perfusion component is considered in the residuals, in order to properly estimate its different contributions in a second step $\mathcal{M}_2$.

Given the estimated $\widehat{\boldsymbol{h}}^{\mathcal{M}_1}$, $\widehat{\boldsymbol{\ell}}^{\mathcal{M}_1}$ and $\widehat{\boldsymbol{a}}^{\mathcal{M}_1}$, we then compute residuals $\boldsymbol{r}^{\mathcal{M}_1}$ containing the remaining perfusion component:

$$\boldsymbol{r}_j^{\mathcal{M}_1} = \boldsymbol{y}_j - \sum_{m=1}^{M} \widehat{a}_j^{m,\mathcal{M}_1} \boldsymbol{X}^m \widehat{\boldsymbol{h}}^{\mathcal{M}_1} - \boldsymbol{P}\widehat{\boldsymbol{\ell}}_j^{\mathcal{M}_1} \tag{5.27}$$

### Perfusion response estimation step $\mathcal{M}_2$

From the residuals of the first step $\boldsymbol{r}^{\mathcal{M}_1}$, we estimate the perfusion component. The remaining signal is, according to (5.1), $\forall j = 1 : J$,

$$\boldsymbol{y}_j^{\mathcal{M}_2} = \sum_{m=1}^{M} c_j^m \boldsymbol{W}\boldsymbol{X}^m \boldsymbol{g} + \alpha_j \boldsymbol{w} + \boldsymbol{b}_j = \boldsymbol{r}_j^{\mathcal{M}_1} \tag{5.28}$$

In this step, we introduce a prior on $\boldsymbol{g}$, to account for the already described physiological relationship $\boldsymbol{g} = \boldsymbol{\Omega}\boldsymbol{h}$:

$$\boldsymbol{g}|\widehat{\boldsymbol{h}}^{\mathcal{M}_1} \sim \mathcal{N}(\boldsymbol{\Omega}\widehat{\boldsymbol{h}}^{\mathcal{M}_1}, v_{\boldsymbol{g}}\boldsymbol{\Sigma}_g), \text{ with } \boldsymbol{\Sigma}_g = \boldsymbol{I}_{D+1}\ . \tag{5.29}$$

The significance of the 2-step approach is to first preprocess the data to subtract the hemodynamics component within the ASL signal, as well as the drift effect, and to focus in a second step on the analysis of the smaller perfusion effect. In [Mumford et al., 2006], differencing methods were used to subtract components with no interest in the perfusion analysis and directly analyse the perfusion effect in the time series. In contrast to

these methods, we expect to disentangle perfusion from hemodynamics components by identifying all the components contained in the signal, and to recover them more accurately.

## 5.6 Comparison of the models on simulated data

The generative model for ASL time series in (5.1) has been used to generate artificial ASL data. A low SNR has been considered, with $TR = 1$ s, mean $ISI = 5.03$ s, duration 25 s, $N = 325$ scans and two experimental conditions ($M = 2$) represented with $20 \times 20$-voxel binary activation label maps corresponding to HRL and PRL maps shown in Fig. 5.4. For both conditions: $(a_j^m | q_j = 1) \sim \mathcal{N}(2.2, 0.3)$ and $(c_j^m | q_j = 1) \sim \mathcal{N}(0.48, 0.1)$. Parameters were chosen to simulate a typical low SNR ASL scenario, in which the perfusion component is much lower than the hemodynamics component. A drift $\boldsymbol{\ell}_j \sim \mathcal{N}(0, 10\boldsymbol{I}_4)$ and noise variance $v_{b_j} = 7$ were used. HRF and PRF shapes were simulated with the physiological model, using the physiological parameters used in [Friston et al., 2000].

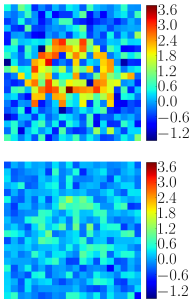



Figure 5.4: HRL and PRL ground truth.

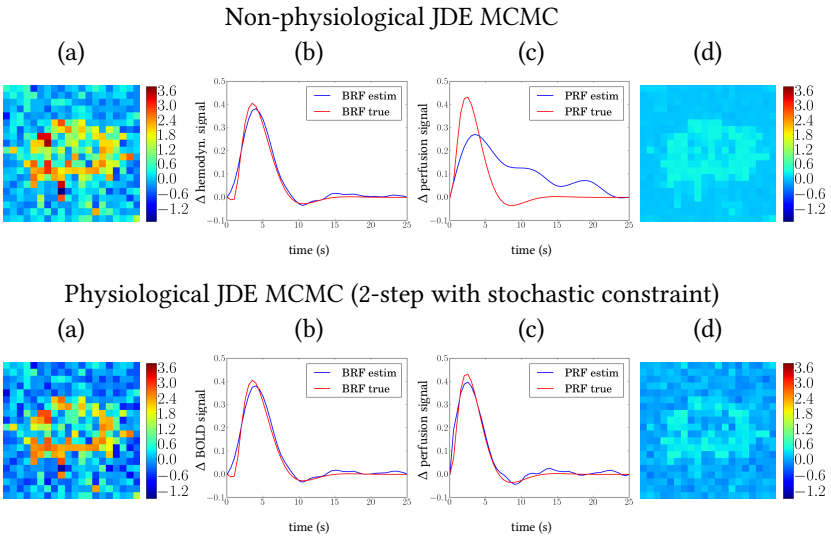



Figure 5.5: Results on artificial data. **Top row**: non-physiological version. **Bottom row**: physiological 2-step version (stochastic constraint in 2-step). **(a,d)**: estimated HRL and PRL effect size maps respectively. The ground-truth maps for the HRL and PRL are depicted in Fig.5.4. **(b,c)**: HRF and PRF estimates, respectively, with their ground truth.

In a low SNR context, the PRF estimate retrieved by the former approach developed in [Vincent et al., 2013b,a] is not physiologically relevant as shown in Fig. 5.5[(c), Top]. In the case of a physiologically informed 2-step Bayesian approach proposed here delivers a PRF estimate very close to the simulated ground truth (see Fig. 5.5[(c), Bottom] The same holds for the HRF, too.

In Fig. 5.6, the robustness of both approaches with respect to the noise variance is studied, in terms of HRF and PRF recovery. The relative root-mean-square-error (rRMSE) is computed for the PRF and HRF estimates, i.e. $\text{rRMSE}_{\boldsymbol{\phi}} = \|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}^{(true)}\| / \|\boldsymbol{\phi}^{(true)}\|$ where $\boldsymbol{\phi} \in \{\boldsymbol{h}, \boldsymbol{g}\}$. We observed that, maintaining a good performance in the HRF estimation, we achieved a much better recovery of the PRF for noise variances larger than $v_{b_j} = 1$, with respect to the non-physiological prior. Therefore, with the introduction of the physiological link between HRF and PRF, we have improved the PRF estimation.

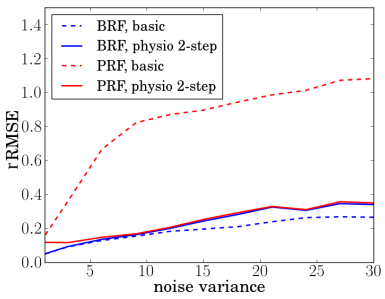



Figure 5.6: Relative RMSE for the HRF and PRF and the two JDE versions, wrt noise variance $v_{b_j}$ ranging from 0.5 to 30.

## 5.7 Validation of the physiologically informed MCMC approach in real data

After testing the performance with simulated data, we test on real data from the AINSI initiative[4]. Real ASL data were recorded during an experiment designed to map auditory and visual brain functions, which consisted of $N = 291$ scans lasting $TR = 3$ s, with $TE = 18$ ms, FoV 192 mm, each yielding a 3-D volume composed of $64 \times 64 \times 22$ voxels (resolution of $3 \times 3 \times 3.5$ mm$^3$). The tagging scheme used was PICORE Q2T, with $TI_1 = 700$ ms, $TI_2 = 1700$ ms. The paradigm was a fast event-related design (mean $ISI = 5.1$ s) comprising 60 auditory and visual stimuli in total. Two regions of interest in the right temporal lobe for the auditory cortex, and left occipital lobe for the visual cortex, were defined manually.

[4] http://thalie.ujf-grenoble.fr/ainsi

Fig. 5.7(b-c) depicts the response estimates superimposed to the canonical shape which is in accordance with the HRF estimates for both methods. We consider here an auditory region where the canonical version has been fitted. Accordingly, the HRL maps (Fig. 5.7(a)) also look alike for both methods. However, PRF estimates significantly differ and the effect of the physiologically-inspired regularization yields a more plausible PRF shape for the 2-step approach compared with the non-physiological JDE version. Results on PRL maps (Fig. 5.7(d)) confirm the improved sensitivity of detection for the proposed approach. In the same way, in the visual cortex, Fig. 5.7(f-g) shows the HRF and PRF estimates, giving a more plausible PRF shape for the 2-step approach, too. For the detection results (Fig. 5.7(h)), the 2-step approach seems also to provide a much better sensitivity of detection.

## 5.8 Discussion

Starting from non-linear systems of differential equations induced by physiological models of the neuro-vascular coupling, we derived a tractable linear operator linking the perfusion and hemodynamics responses. This operator showed good approximation performance and we demonstrated its ability to capture both realistic perfusion and hemodynamics components.

This derived linear operator was easily incorporated in a JDE framework at no additional cost. Different models have been considered when including this link: stochastic constraint, deterministic constraint and a hierarchical formulation.

- The deterministic formulation considers $\boldsymbol{g}$ fixed to $\boldsymbol{\Omega h}$ during Gibbs sampling. It enables the recovery of the ground-truth HRF and PRF at low SNR. However, this approach does not allow any flexibility in the PRF.

- The hierarchical formulation considers a "true" HRF $\boldsymbol{h}_t$, and the HRF $\boldsymbol{h}$ and PRF $\boldsymbol{g}$ noisy versions that are conditional to the "true" HRF, with a certain precision. We enforce a high precision on the HRF to be close to the "true" HRF, and a low precision on the PRF to account for possible variability around the prior mean. This model definition is proper theoretically. However, the addition of other parameters to

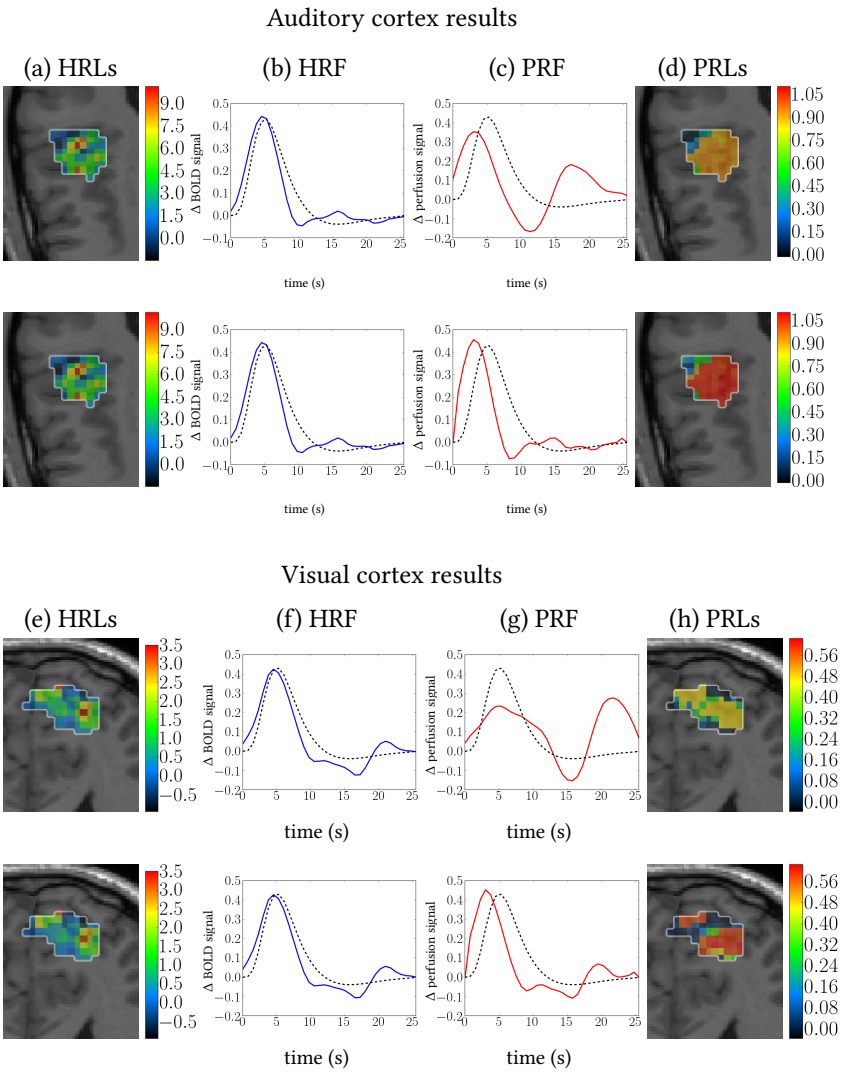



Figure 5.7: Comparison of the two JDE versions on real data in the auditory and visual cortices. **(top row in auditory and visual cortex results)**: non-physiological version. **(bottom row in auditory and visual cortex results)**: physiological 2-step version (stochastic constraint in 2-step). **(a,e)** and **(d,h)**: estimated HRL and PRL effect size maps, respectively. **(b,f)** and **(c,g)**: HRF and PRF estimates, respectively. The canonical HRF is depicted as a black dashed line, while PRF and HRF estimated are depicted in solid red and blue lines, respectively.

estimate (the "true" HRF and its variance) adds more complexity to the model.

- The stochastic constraint seemed to be the best one to be used in this framework, allowing to couple both responses with flexibility. It has been implemented in a 1-step and in a 2-step procedure.

Results in this chapter correspond to the stochastic constraint in a 2-step procedure, but in the next chapter results are shown for the stochastic constraint in 1-step case. With the introduction of this prior, we achieve significant improvement in PRF estimation, especially in critical low SNR situations. In this chapter, this has been shown in both simulated and real data.

# 6 Bayesian analysis of functional ASL data: VEM solution

In the previous chapter, a Bayesian model for the analysis of functional Arterial Spin Labeling (fASL) data has been proposed based on [Vincent et al., 2013b], using a Markov Chain Monte Carlo optimization strategy. One of the novelties of the previous chapter, presented in [Frau-Pascual et al., 2014a], is the introduction of prior knowledge through the relationship between perfusion and hemodynamic responses derived from physiological models. This relationship allows us to inform the perfusion response function (PRF) estimation from the hemodynamic response function (HRF), as the hemodynamic component has a higher SNR than the perfusion one due to the acquisition procedure.

In this chapter, following the spirit of [Chaari et al., 2013], we provide an alternative solution based on the variational expectation-maximization (VEM) algorithm. As explained in chapter 4, this solution is less computationally expensive than MCMC. The VEM algorithm has been explained in detail in chapter 4. This framework is also more convenient to deal with constraints, as normalization or positivity. In this chapter, we focus in the use of the physiological operator $\mathbf{\Omega}$ as a stochastic constraint. First, we explain the VEM for JDE ASL. Then, JDE-VEM and JDE-MCMC are compared to point out the differences on simulated and real data. The outcome of this chapter was presented at the International Workshop on Pattern Recognition in NeuroImaging (PRNI) and at Medical Image Computing and Computer-Assisted Intervention (MICCAI) conference in 2015, and can be read in [Frau-Pascual et al., 2015b,c].

**Outcome of this chapter:**

A. Frau-Pascual, F. Forbes, and P. Ciuciu. *"Variational physiologically informed solution to hemodynamic and perfusion response estimation from ASL fMRI data"*. In Pattern Recognition in NeuroImaging (PRNI), 2015 International Workshop on, pages 57–60. IEEE, 2015a.

A. Frau-Pascual, F. Forbes, and P. Ciuciu. *"Comparison of stochastic and variational solutions to ASL fMRI data analysis"*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, pages 85–92. Springer, 2015b.

## 6.1 Variational expectation maximization for JDE ASL

Here, a Variational Expectation-Maximization (VEM) algorithm is proposed to deal with the intractable posterior of our model $p(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}|\boldsymbol{y})$, when estimating the variables and parameters of the model. VEM algorithm approximates the posterior with a function $\tilde{p}$, and minimizes the Kullback Leibler divergence between the approximation $\tilde{p}$ and the actual posterior: $\mathrm{KL}(\tilde{p}, p(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}|\boldsymbol{y}))$.

The missing variables in our model are the hemodynamic response levels $\boldsymbol{a} \in \mathcal{A}$, the HRF $\boldsymbol{h} \in \mathcal{H}$, the perfusion response levels $\boldsymbol{c} \in \mathcal{C}$, the PRF $\boldsymbol{g} \in \mathcal{G}$, and the activation states $\boldsymbol{q} \in \mathcal{Q}$. Let $\mathcal{D}$ be the set of all probability distributions on $\mathcal{A} \times \mathcal{H} \times \mathcal{C} \times \mathcal{G} \times \mathcal{Q}$. One way of dealing with several missing variables is to restrict the class of approximate posteriors with the mean field approximation: we choose our approximate posterior distribution over a restricted class of probability distributions, $\tilde{\mathcal{D}}$, that can factorize as

$$\tilde{p}(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}) = \tilde{p}_a(\boldsymbol{a})\ \tilde{p}_h(\boldsymbol{h})\ \tilde{p}_c(\boldsymbol{c})\ \tilde{p}_g(\boldsymbol{g})\ \tilde{p}_q(\boldsymbol{q})$$

where $\tilde{p}_a \in \mathcal{D}_A$, $\tilde{p}_h \in \mathcal{D}_H$, $\tilde{p}_c \in \mathcal{D}_C$, $\tilde{p}_g \in \mathcal{D}_G$ and $\tilde{p}_q \in \mathcal{D}_Q$ are sets of probability distributions on $\mathcal{A}, \mathcal{H}, \mathcal{C}, \mathcal{G}, \mathcal{Q}$ respectively.

As we have seen in previous chapters, EM can be viewed [Neal and Hinton, 1998] as an alternating maximization procedure of the free energy function $\mathcal{F}$. For any $\tilde{p} \in \mathcal{D}$,

$$\mathcal{F}(\tilde{p}, \boldsymbol{\theta}) = \mathrm{E}_{\tilde{p}}\big[\log p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}\ ;\ \boldsymbol{\theta})\big] + I[\tilde{p}] \tag{6.1}$$

where $I[\tilde{p}] = -\mathrm{E}_{\tilde{p}}\big[\log \tilde{p}(\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q})\big]$ is the entropy of $\tilde{p}$, and $\mathrm{E}_{\tilde{p}}\big[\big]$ denotes the expectation with respect to $\tilde{p}$. The first term thus becomes a type of reconstruction measure: It quantifies to what extent probable events for $\tilde{p}$ are also probable for $p$. The second term quantifies the uncertainty in $\tilde{p}$. Denoting current parameter values by $\boldsymbol{\theta}^{(r)}$, the alternating procedure, as in section 4.4, proceeds as follows:

$$\textbf{E-step: } \tilde{p}^{(r)} = \underset{\tilde{p} \in \mathcal{D}}{\arg\max}\ \mathcal{F}(\tilde{p}, \boldsymbol{\theta}^{(r)}) \tag{6.2}$$

$$\textbf{M-step: } \boldsymbol{\theta}^{(r+1)} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\max}\ \mathcal{F}(\tilde{p}^{(r)}, \boldsymbol{\theta}) \tag{6.3}$$

When applying the mean field approximation, the E-step can be further decomposed into five stages updating the different variables in turn. At iteration $(r)$, with current estimates denoted by $\tilde{p}_a^{(r-1)}$, $\tilde{p}_h^{(r-1)}$, $\tilde{p}_c^{(r-1)}$, $\tilde{p}_g^{(r-1)}$, $\tilde{p}_q^{(r-1)}$ and $\boldsymbol{\theta}^{(r)}$, the updating formulae are of the form:

**E-A-step:**

$$\tilde{p}_a^{(r)} = \underset{\tilde{p}_a \in \mathcal{D}_A}{\arg\max}\, \mathcal{F}\left(\tilde{p}_a\ \tilde{p}_h^{(r-1)}\ \tilde{p}_c^{(r-1)}\ \tilde{p}_g^{(r-1)}\ \tilde{p}_q^{(r-1)}; \theta^{(r)}\right) \tag{6.4}$$

**E-C-step:**

$$\tilde{p}_c^{(r)} = \underset{\tilde{p}_c \in \mathcal{D}_C}{\arg\max}\, \mathcal{F}\left(\tilde{p}_a^{(r)}\ \tilde{p}_h^{(r-1)}\ \tilde{p}_c\ \tilde{p}_g^{(r-1)}\ \tilde{p}_q^{(r-1)}; \theta^{(r)}\right) \tag{6.5}$$

**E-Q-step:**

$$\tilde{p}_q^{(r)} = \underset{\tilde{p}_q \in \mathcal{D}_Q}{\arg\max}\, \mathcal{F}\left(\tilde{p}_a^{(r)}\ \tilde{p}_h^{(r-1)}\ \tilde{p}_c^{(r)}\ \tilde{p}_g^{(r-1)}\ \tilde{p}_q; \theta^{(r)}\right) \tag{6.6}$$

**E-H-step:**

$$\tilde{p}_h^{(r)} = \arg\max_{\tilde{p}_h \in \mathcal{D}_H} \mathcal{F}\left(\tilde{p}_a^{(r)}\ \tilde{p}_h\ \tilde{p}_c^{(r)}\ \tilde{p}_g^{(r-1)}\ \tilde{p}_q^{(r)}; \theta^{(r)}\right) \tag{6.7}$$

**E-G-step:**

$$\tilde{p}_g^{(r)} = \arg\max_{\tilde{p}_g \in \mathcal{D}_G} \mathcal{F}\left(\tilde{p}_a^{(r)}\ \tilde{p}_h^{(r)}\ \tilde{p}_c^{(r)}\ \tilde{p}_g\ \tilde{p}_q^{(r)}; \theta^{(r)}\right) \tag{6.8}$$

Hereafter, for the ease of presentation, the $(r)$ and $(r-1)$ superscripts are omitted. From the Kullback-Leibler divergence properties (see the derivation in section 4.4), we can derive as in Eq. (4.8) the solutions of Eq. (6.6) for each step as:

$$\textbf{E-A:}\ \tilde{p}_a(a) \propto \exp\left(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_c\tilde{p}_q}\left[\log p(\boldsymbol{a} \mid \boldsymbol{y}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\right]\right) \tag{6.9}$$

$$\textbf{E-C:}\ \widetilde{p}_c(c) \propto \exp\left(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_a\tilde{p}_q}\left[\log p(\boldsymbol{c} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\right]\right) \tag{6.10}$$

$$\textbf{E-Q:}\ \tilde{p}_q(q) \propto \exp\left(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_a\widetilde{p}_c}\left[\log p(\boldsymbol{q} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\theta})\right]\right) \tag{6.11}$$

$$\textbf{E-H:}\ \tilde{p}_h(h) \propto \exp\left(\mathrm{E}_{\tilde{p}_g\tilde{p}_a\tilde{p}_c\tilde{p}_q}\left[\log p(\boldsymbol{a} \mid \boldsymbol{y}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\right]\right) \tag{6.12}$$

$$\textbf{E-G:}\ \widetilde{p}_g(g) \propto \exp\left(\mathrm{E}_{\tilde{p}_h\tilde{p}_a\tilde{p}_c\tilde{p}_q}\left[\log p(\boldsymbol{c} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\right]\right) \tag{6.13}$$

Note that the dependence between random variables, as in MCMC, is translated into dependence between statistical moments in VEM. This can be seen in the formulas derived in the appendices D and E: for the computation of $\widetilde{p}_h$ for example, the MCMC expression (D.16) becomes in VEM (E.17), and (D.17) in MCMC becomes (E.18) in VEM.

The corresponding **M-step**, where parameters are estimated, is given by:

$$\begin{aligned}\textbf{M:}\ \boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Big[ & \mathrm{E}_{\tilde{p}_a\tilde{p}_c}\left[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{v}_b)\right] \\ & + \mathrm{E}_{\tilde{p}_h}\left[\log p(\boldsymbol{h}; v_h)\right] + \mathrm{E}_{\tilde{p}_g}\left[\log p(\boldsymbol{g}; v_g)\right] \\ & + \mathrm{E}_{\tilde{p}_a\tilde{p}_q}\left[\log p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\right] \\ & + \mathrm{E}_{\tilde{p}_c\tilde{p}_q}\left[\log p(\boldsymbol{c} \mid \boldsymbol{q}; \boldsymbol{\eta}, \boldsymbol{\rho})\right] \\ & + \mathrm{E}_{\tilde{p}_q}\left[\log p(\boldsymbol{q}; \boldsymbol{\beta})\right] + \log p(\boldsymbol{\theta}) \Big]\end{aligned} \tag{6.14}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{v}_b, \boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\eta}, \boldsymbol{\rho}, v_h, v_g, \boldsymbol{\beta}\}$. Hyperparameters were considered for the parameters $v_h$, $v_g$ and $\boldsymbol{\beta}$, and therefore we include them in the M-step computation. Given the separability of the prior probability density functions, the M-step can be divided into separate M-steps, as in [Chaari et al., 2013]. The stopping criterion of the algorithm will be based on the convergence of the free energy:

$$\begin{aligned}\mathcal{F}(\tilde{p}, \boldsymbol{\theta}) = & \mathrm{E}_{\tilde{p}}\big[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\ p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\theta}_{\boldsymbol{a}})\ p(\boldsymbol{c} \mid \boldsymbol{q}; \boldsymbol{\theta}_{\boldsymbol{c}}) \\ & p(\boldsymbol{q}; \boldsymbol{\beta}) p(\boldsymbol{\beta}; \lambda_\beta) p(\boldsymbol{h}; v_h) p(v_h; \lambda_h)\ p(\boldsymbol{g} \mid \boldsymbol{h}; v_g) p(v_g; \lambda_g)\big] \\ & + \mathcal{I}(\tilde{p}_a) + \mathcal{I}(\tilde{p}_h) + \mathcal{I}(\tilde{p}_c) + \mathcal{I}(\tilde{p}_g) + \mathcal{I}(\tilde{p}_q)\end{aligned} \tag{6.15}$$

Algorithm 2 shows the alternation between E and M steps. Details of the model and the approximate posterior distributions can be found in appendix E. Free energy computation can be found in appendix F.

**Algorithm 2** VEM algorithm for JDE in a given parcel.

1: **initialize**
2: **repeat**
3: **Expectation step:** Compute probability distributions

$$\begin{aligned}
\widetilde{p}_{a_j}(\boldsymbol{a}_j) &= \mathcal{N}(\boldsymbol{a}_j; \tilde{\boldsymbol{m}}_{a_j}, \tilde{\boldsymbol{\Sigma}}_{a_j}) \\
\widetilde{p}_{c_j}(\boldsymbol{c}_j) &= \mathcal{N}(\boldsymbol{c}_j; \tilde{\boldsymbol{m}}_{c_j}, \tilde{\boldsymbol{\Sigma}}_{c_j}) \\
\widetilde{p}_{q^m}(\boldsymbol{q}^m) &= \prod_{j \in \mathcal{J}} \tilde{p}_{q_j^m}(q_j^m) \\
\widetilde{p}_h(\boldsymbol{h}) &= \mathcal{N}(\tilde{\boldsymbol{m}}_h, \tilde{\boldsymbol{\Sigma}}_h) \\
\widetilde{p}_g(\boldsymbol{g}) &= \mathcal{N}(\tilde{\boldsymbol{m}}_g, \tilde{\boldsymbol{\Sigma}}_g)
\end{aligned}$$

4: **Maximization step:** Compute parameters

$$\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\eta}, \boldsymbol{\rho}, v_h, v_g, \beta, \boldsymbol{\ell}_j, \alpha_j, v_{b_j}\}$$

5: **Compute free energy** $\mathcal{F}(\tilde{p}, \boldsymbol{\theta})$
6: **until** free energy $\Delta\mathcal{F}(\tilde{p}, \boldsymbol{\theta}) < 10^{-5}$ or n_iterations$> 100$
7: **RETURN** estimates $\tilde{\boldsymbol{m}}_h, \tilde{\boldsymbol{\Sigma}}_h, \tilde{\boldsymbol{m}}_g, \tilde{\boldsymbol{\Sigma}}_g, \tilde{\boldsymbol{m}}_a, \tilde{\boldsymbol{\Sigma}}_a, \tilde{\boldsymbol{m}}_c, \tilde{\boldsymbol{\Sigma}}_c, \tilde{p}_{q_j^m}, \boldsymbol{\theta}$

It is worth mentioning that, in this approach, the hemodynamic and perfusion response levels $\boldsymbol{a}$ and $\boldsymbol{c}$, and the activation states $\boldsymbol{q}$ are hidden variables because their dimensionality increases with the dimensionality of the data. The HRF $\boldsymbol{h}$ and PRF $\boldsymbol{g}$ are actually random parameters because they do not change dimensionality with the data. However, since they have priors and we estimate a posterior density, they are in the E step as it is the case for parameters in Variational Bayes EM (VBEM). Although we consider the rest of the parameters in the M step, some of them have priors too, as it is the case of the spatial regularization parameter $\boldsymbol{\beta}$ or the HRF variance $v_h$, when we use hyperpriors. These parameters could actually be in the E step because we compute a posterior density, but we then keep the MAP estimate. For this reason, we have these parameters in the M step.

### 6.1.1 Normalization constraint

One of the issues of JDE is that the model is bilinear. We estimate the HRF $\boldsymbol{h}$ and the amplitudes $\boldsymbol{a}$, but they are being multiplied, so there is a scaling indetermination. If we do not take care of it, numerical issues can emerge when computing the parameters. For this reason, we usually fix the $\ell_2$-norm of the HRF to 1, and this way we get all the information of the scale in the response levels $\boldsymbol{a}$. The normalization of $\boldsymbol{h}$ is not easy to deal with in MCMC since we need to use truncated Gaussians: $\boldsymbol{h} \sim \mathcal{N}_{[-1,+1]}(0, v_h\Sigma_h)$ and $\boldsymbol{g} \sim \mathcal{N}_{[0,+\infty]}(0, v_g\Sigma_g)$. The implementation is not straightforward.

In VEM, the inclusion of normalization or positivity constraints becomes easier. We can consider normalization constraints in the priors on $\boldsymbol{h}$ and $\boldsymbol{g}$ so that our solution has unit $\ell_2$-norm. For example, to constrain $\boldsymbol{h}$:

$$p(\boldsymbol{h}) \propto \mathcal{N}(0, v_h\boldsymbol{R}) \quad \text{if } \boldsymbol{h} \in S^2(\mathbb{R}^{D+1}) \tag{6.16}$$

$$= 0 \quad \text{otherwise} \tag{6.17}$$

where $S^2(\mathbb{R}^{D+1})$ is the $L_2$ unit ball of $\mathbb{R}^{D+1}$. This induces a quadratic constraint. Note that the normalizing constant is not necessarily easy to compute but we won't need it. The same can be done for $\boldsymbol{g}$. We could also impose positivity very easily: $\boldsymbol{g} \geqslant \mathbf{0}$.

For this, we modify the sought variational approximation to $\tilde{p} = \tilde{p}_a \ \delta_{\tilde{\boldsymbol{h}}} \ \tilde{p}_c \ \delta_{\tilde{\boldsymbol{g}}} \ \tilde{p}_q$, where the probabilities on $\boldsymbol{h}$ and $\boldsymbol{g}$ are replaced by Dirac distributions. This reduces the search to pointwise estimates $\tilde{\boldsymbol{h}}$ and $\tilde{\boldsymbol{g}}$. The E-H and E-G steps in Eqs. (6.7)-(6.8) then yield maximization problems which are easily constrained to account for normalization:

$$\textbf{E-H:}\ \tilde{\boldsymbol{h}} = \underset{\boldsymbol{h} \text{ s.t. } \boldsymbol{h}^t\boldsymbol{h}=1}{\arg\max}\ \mathrm{E}_{\tilde{p}_a\tilde{p}_c\tilde{p}_q}\big[\log p(\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \tilde{\boldsymbol{g}}, \boldsymbol{q}; \boldsymbol{\theta})\big] \tag{6.18}$$

$$\textbf{E-G:}\ \tilde{\boldsymbol{g}} = \underset{\boldsymbol{g} \text{ s.t. } \boldsymbol{g}^t\boldsymbol{g}=1}{\arg\max}\ \mathrm{E}_{\tilde{p}_a\tilde{p}_c\tilde{p}_q}\big[\log p(\boldsymbol{g} \mid \boldsymbol{y}, \boldsymbol{a}, \tilde{\boldsymbol{h}}, \boldsymbol{c}, \boldsymbol{q}; \boldsymbol{\theta})\big] \tag{6.19}$$

Considering a normalized $\boldsymbol{h}$ and $\boldsymbol{g}$ amounts to minimizing a quadratic function under a quadratic constraint, namely $\|\boldsymbol{h}\|_2 = 1$ and $\|\boldsymbol{g}\|_2 = 1$ respectively. The other E-steps can be derived from standard expressions replacing expectations over $\boldsymbol{h}$ and $\boldsymbol{g}$ by $\tilde{\boldsymbol{h}}$ and $\tilde{\boldsymbol{g}}$, e.g.:

$$\textbf{E-A:}\ \tilde{p}_a(\boldsymbol{a}) \propto \exp\Big(\mathrm{E}_{\tilde{p}_c\tilde{p}_q}\big[\log p(\boldsymbol{a} \mid \boldsymbol{y}, \tilde{\boldsymbol{h}}, \boldsymbol{c}, \tilde{\boldsymbol{g}}, \boldsymbol{q}; \boldsymbol{\theta})\big]\Big) \tag{6.20}$$

$$\textbf{E-C:}\ \tilde{p}_c(\boldsymbol{c}) \propto \exp\Big(\mathrm{E}_{\tilde{p}_a\tilde{p}_q}\big[\log p(\boldsymbol{c} \mid \boldsymbol{y}, \boldsymbol{a}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}}, \boldsymbol{q}; \boldsymbol{\theta})\big]\Big) \tag{6.21}$$

$$\textbf{E-Q:}\ \tilde{p}_q(\boldsymbol{q}) \propto \exp\big(\mathrm{E}_{\tilde{p}_a\tilde{p}_c}\big[\log p(\boldsymbol{q} \mid \boldsymbol{y}, \boldsymbol{a}, \tilde{\boldsymbol{h}}, \boldsymbol{c}, \tilde{\boldsymbol{g}}; \boldsymbol{\theta})\big]\big)\,, \tag{6.22}$$

The corresponding **M-step** is given by:

$$\begin{aligned}\textbf{M:}\ \boldsymbol{\theta} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\Big[&\mathrm{E}_{\tilde{p}_a\tilde{p}_c}\big[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \tilde{\boldsymbol{h}}, \boldsymbol{c}, \tilde{\boldsymbol{g}}; \boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{v}_b)\big] \\ &+ \log p(\tilde{\boldsymbol{h}}; v_h) + \log p(\tilde{\boldsymbol{g}}; v_g) \\ &+ \mathrm{E}_{\tilde{p}_a\tilde{p}_q}\big[\log p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big] \\ &+ \mathrm{E}_{\tilde{p}_c\tilde{p}_q}\big[\log p(\boldsymbol{c} \mid \boldsymbol{q}; \boldsymbol{\eta}, \boldsymbol{\rho})\big] \\ &+ \mathrm{E}_{\tilde{p}_q}\big[\log p(\boldsymbol{q}; \boldsymbol{\beta})\big] + \log p(\boldsymbol{\theta})\Big]\end{aligned}$$

where $\boldsymbol{\theta} = \big\{\boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{v}_b, \boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\eta}, \boldsymbol{\rho}, v_h, v_g, \boldsymbol{\beta}\big\}$. For the constrained case, VEM becomes algorithm 3.

**Algorithm 3** VEM algorithm for JDE in a given parcel, using constraints in the norm of the response functions.

1: **initialize**
2: **repeat**
3: **Expectation step:** Compute probability distributions and point wise

estimates of response functions

$$\tilde{p}_{a_j}(\boldsymbol{a}_j) = \mathcal{N}(\boldsymbol{a}_j; \tilde{\boldsymbol{m}}_{a_j}, \tilde{\boldsymbol{\Sigma}}_{a_j}) \tag{6.23}$$

$$\tilde{p}_{c_j}(\boldsymbol{c}_j) = \mathcal{N}(\boldsymbol{c}_j; \tilde{\boldsymbol{m}}_{c_j}, \tilde{\boldsymbol{\Sigma}}_{c_j}) \tag{6.24}$$

$$\tilde{p}_{q^m}(\boldsymbol{q}^m) = \prod_{j \in \mathcal{J}} \tilde{p}_{q_j^m}(q_j^m) \tag{6.25}$$

$$\tilde{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \left( \log \mathcal{N}(\tilde{\boldsymbol{m}}_h, \tilde{\boldsymbol{\Sigma}}_h) + \lambda_{c1}\, \boldsymbol{h}^T \boldsymbol{h} \right) \tag{6.26}$$

$$\tilde{\boldsymbol{g}} = \arg\min_{\boldsymbol{g}} \left( \log \mathcal{N}(\tilde{\boldsymbol{m}}_g, \tilde{\boldsymbol{\Sigma}}_g) + \lambda_{c2}\, \boldsymbol{g}^T \boldsymbol{g} \right) \tag{6.27}$$

4: **Maximization step:** Compute parameters

$$\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{v}, \boldsymbol{\eta}, \boldsymbol{\rho}, v_h, v_g, \beta, \boldsymbol{\ell}_j, \alpha_j, v_{b_j}\}$$

5: **Compute free energy** $\mathcal{F}(\tilde{p}, \boldsymbol{\theta})$

6: **until** free energy $\Delta\mathcal{F}(\tilde{p}, \boldsymbol{\theta}) < 10^{-5}$ or n_iterations$> 100$

7: **RETURN** estimates $\tilde{\boldsymbol{h}}, \tilde{\boldsymbol{g}}, \tilde{\boldsymbol{m}}_a, \tilde{\boldsymbol{\Sigma}}_a, \tilde{\boldsymbol{m}}_c, \tilde{\boldsymbol{\Sigma}}_c, \tilde{p}_{q_j^m}, \boldsymbol{\theta}$

---

In the free energy computation, we need to introduce Lagrange multipliers $\lambda_{c1}$ and $\lambda_{c2}$ of the minimization under a quadratic constraint for the computation of $\tilde{\boldsymbol{h}}$ and $\tilde{\boldsymbol{g}}$. If we constrain both response functions:

$$\begin{aligned}
\mathcal{F}(\tilde{p}, \boldsymbol{\theta}) = {} & \mathrm{E}_{\tilde{p}_a \tilde{p}_c}\big[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\big] + \log p(\boldsymbol{h}; v_h) \\
& + \log p(\boldsymbol{g} \mid \boldsymbol{h}; v_g) + \mathrm{E}_{\tilde{p}_a \tilde{p}_q}\big[\log p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big] \\
& + \mathrm{E}_{\tilde{p}_c \tilde{p}_q}\big[\log p(\boldsymbol{c} \mid \boldsymbol{q}; \boldsymbol{\eta}, \boldsymbol{\rho})\big] + \mathrm{E}_{\tilde{p}_q}\big[\log p(\boldsymbol{q}; \boldsymbol{\beta})\big] + p(\boldsymbol{\beta}; \lambda_\beta) \\
& + p(v_h; \lambda_h) + p(v_g; \lambda_g) + \mathcal{I}(\tilde{p}_a) + \mathcal{I}(\tilde{p}_c) + \mathcal{I}(\tilde{p}_q) \\
& + \lambda_{c1}(\|\boldsymbol{h}\| - 1) + \lambda_{c2}(\|\boldsymbol{g}\| - 1)
\end{aligned} \tag{6.28}$$

The transformation $\boldsymbol{g} = \boldsymbol{\Omega}\boldsymbol{h}$ does not give an $\ell_2-$norm$= 1$ $\boldsymbol{g}$ when $\boldsymbol{h}$ has $\ell_2-$norm$= 1$. For this reason, if we constraint $\boldsymbol{h}$ and $\boldsymbol{g}$ at the same time, we need to do it through the relationship $\boldsymbol{g} = \frac{\boldsymbol{\Omega}\boldsymbol{h}}{\|\boldsymbol{\Omega}\boldsymbol{h}\|_2}$. Otherwise, we can just apply the normalization to $\boldsymbol{h}$, and $\boldsymbol{g}$ will be scaled accordingly due to $\boldsymbol{\Omega}$. For further details of the model and the approximate posterior distributions see appendix E, and for details about free energy computation see appendix F.

## 6.2 Comparison of stochastic and variational solutions

In this chapter and the previous one, we have presented the MCMC and VEM solutions to the ASL JDE approach. In theory, MCMC solution gives asymptotically the correct estimation, whereas VEM solution can fall into local minima and calls for attention to the parameter initialization.

The implementation of these two solutions can be found in PyHRF[1]. In this section, we assess the performance of both solutions by comparing them, as already done for BOLD in [Chaari et al., 2013]. Note that we consider the MCMC version with a stochastic $\boldsymbol{\Omega}$ constraint computed in one step, and the VEM with $\ell_2-$norm$= 1$ constraint.

[1] https://github.com/pyhrf/pyhrf

Different data sets have been analysed to compare the performance of the VEM and MCMC approaches: first, artificial data synthesized with the generative model (5.1); and second, real data acquired on several subjects from the AINSI initiative[2].

[2] http://thalie.ujf-grenoble.fr/ainsi

### 6.2.1 Artificial data

To assess the correct estimation of the parameters in JDE ASL, we simulate some artificial data in order to have a ground truth with which to compare our results. We try to emulate real ASL data by using low temporal resolution (high $TR$) and a very low SNR. We use the experimental design of the real data that we are going to analyse next: the AINSI dataset.

$N = 292$ artificial ASL images (i.e. 146 control/tag pairs) were simulated using a realistic low SNR according to the observation model in Eq. (5.1). Different levels of SNR were used, in order to show the performance of the method depending on the noise level. To emulate the slow sampling rate of ASL images, Eq. (5.1) was synthesized at $\Delta t = 0.5$ s and then down-sampled to a certain repetition time (TR), which means that the temporal resolution of rows (TR) and columns ($\Delta t$) of $\boldsymbol{X}^m$ is different. Here, we used a fast event-related paradigm comprising two conditions ($M = 2$), with mean ISI = 5 s. A TR = 3 s is considered as a realistic ASL experiment, compared to the TR = 1 s that could be used for a realistic experiment when using BOLD signal.

In the experiments, $\boldsymbol{h}$ and $\boldsymbol{g}$ are generated as depicted in Fig.6.1(a)-(b) by dashed lines. $\boldsymbol{P}$ is a polynomial basis of order $F = 4$. Drift coefficients and noise realizations were drawn according to $\boldsymbol{\ell}_j \sim \mathcal{N}(0, 10\ \boldsymbol{I}_F)$ and $\boldsymbol{b}_j \sim \mathcal{N}(0, 2\ \boldsymbol{I}_N)$, respectively. HRLs were sampled from $(a_j^m | q_j^m = 1) \sim \mathcal{N}(2.2, 0.3)$ (for active voxels) and from $(a_j^m | q_j^m = 0) \sim \mathcal{N}(0, 0.3)$ (for non-active voxels). PRLs were generated with a lower contrast than HRLs: $(c_j^m | q_j^m = 1) \sim \mathcal{N}(1.6, 0.3)$ and $(c_j^m | q_j^m = 0) \sim \mathcal{N}(0, 0.3)$. PRLs and HRLs were chosen so as to make this synthetic setting realistic: PRLs lower than HRLs, and active/non-active voxels distribution means close. Activation states (assignment variables $\boldsymbol{q}$) were set by a hand-drawn map.

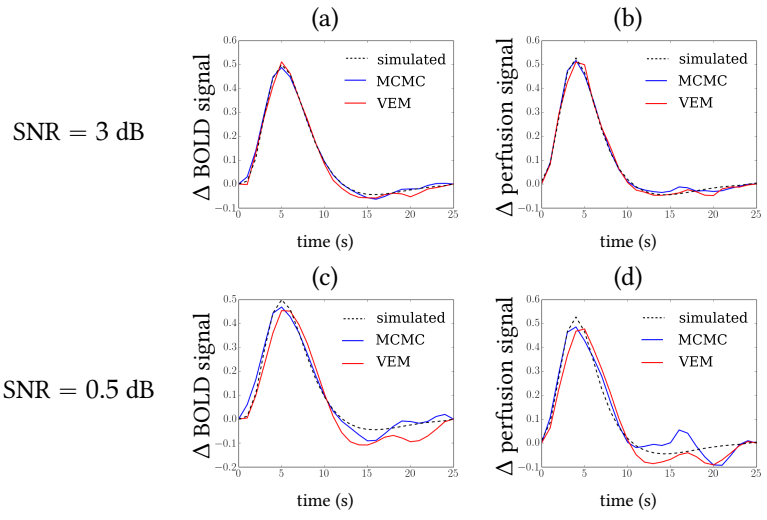



Figure 6.1: Artificial data with 2 noise settings: (a, b) SNR = 3 dB, (c, d) SNR = 0.5 dB. Ground-truth response curves (black dashed lines) and estimated hemodynamic (a, c) and perfusion (b, d) response functions with **MCMC** and **VEM**.

Fig. 6.1(a-d) shows the HRF and PRF estimates obtained for two different noise levels. Both response functions were similarly well recovered with MCMC and VEM at 3dB SNR with a degradation at lower SNR (i.e. 0.5 dB). In the latter case, MCMC recovers the time-to-peak slightly better. The labels (active/non-active) in Fig. 6.2 are well recovered with both MCMC and VEM at the higher SNR. In the high-noise regime both solutions fail to

recover the ground-truth label maps. VEM labels maps are more *contrasted* than with MCMC which is likely to better estimate variability. Fig. 6.3 shows the root mean squared errors (RMSE) for a range of SNR levels. Response functions are well recovered with small RMSE in all cases (Fig. 6.3(a)) but with better estimations with MCMC. In contrast, response levels are better recovered with VEM (Fig. 6.3(b)). This is consistent with previous comparisons between VEM and MCMC on BOLD signals [Chaari et al., 2013].

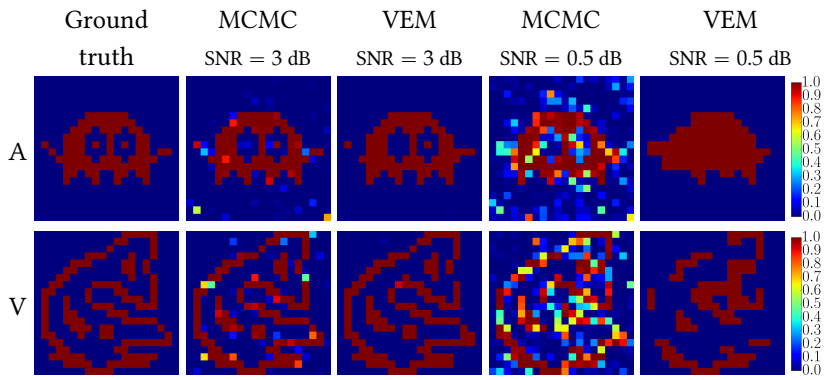



Figure 6.2: Results on artificial data for labels $\boldsymbol{q}$. The probability to be activated is shown for each voxel, for 2 experimental conditions, namely auditory (A) and visual (V) stimuli. The ground truth as well as the MCMC and VEM activation probability estimates are shown in two different SNR scenarios.

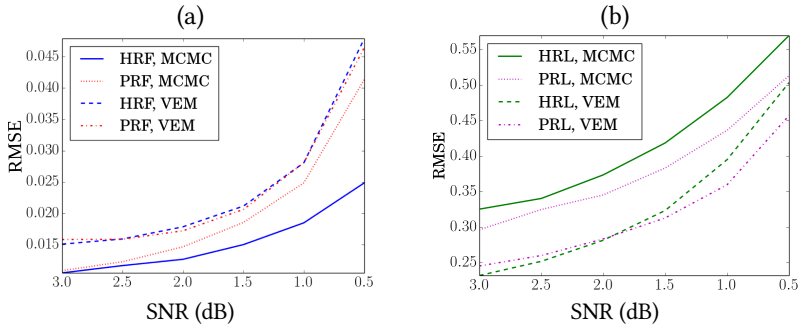



Figure 6.3: RMSE comparison between MCMC and VEM approaches. (a) Response functions HRF and PRF. (b) Mean over conditions of the RMSE of the response levels HRL and PRL.

### 6.2.2 Real data

The real ASL data used in this section is the AINSI initiative dataset, already used in section 5.7. Recall that the experiment was designed to map auditory and visual brain functions with a fast event-related paradigm comprising 60 auditory and visual stimuli in total, randomly distributed according to a mean inter-stimulus interval of 5.1 s. It consists of $N = 291$ scans with TR $= 3$ s, TE $= 18$ ms, FoV 192 mm, with spatial resolution $3 \times 3 \times 7$ mm$^3$. The tagging scheme used was PICORE Q2T, with $(\text{TI}_1, \text{TI}_2) = (700, 1700)$ ms.

In Fig. 6.4, the MCMC and VEM results are shown in the left and right visual and auditory cortices. The HRL maps in Fig. 6.4 are very similar for the two approaches and for auditory (A) and visual (V) conditions in contrast to the larger variability reported in the PRL maps owing to the lower effect size. Interestingly, the PRL maps yielded by the two algorithms

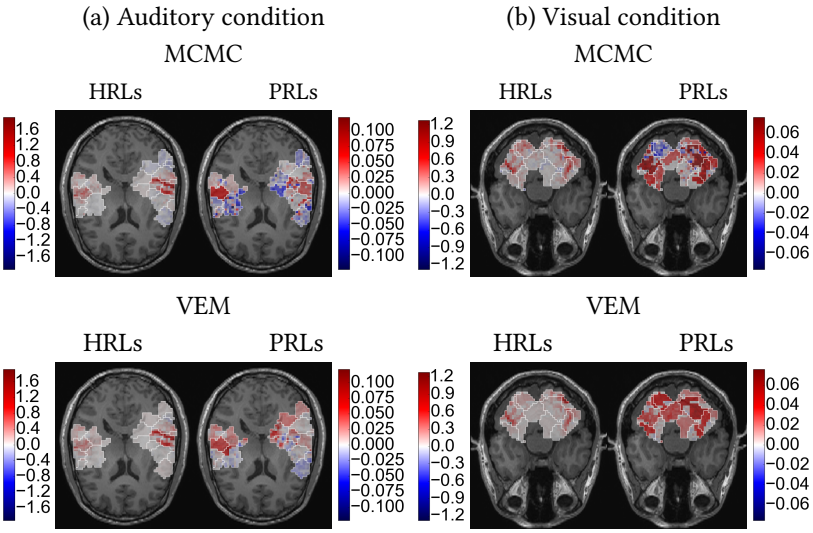



Figure 6.4: Results on real fASL data for a single subject of the AINSI database for both conditions: (a) Auditory and (b) Visual. The first line contains the results for the approach MCMC, and the second line for the VEM approach. For each one, HRL are on the left and PRL on the right, with its respective scales.

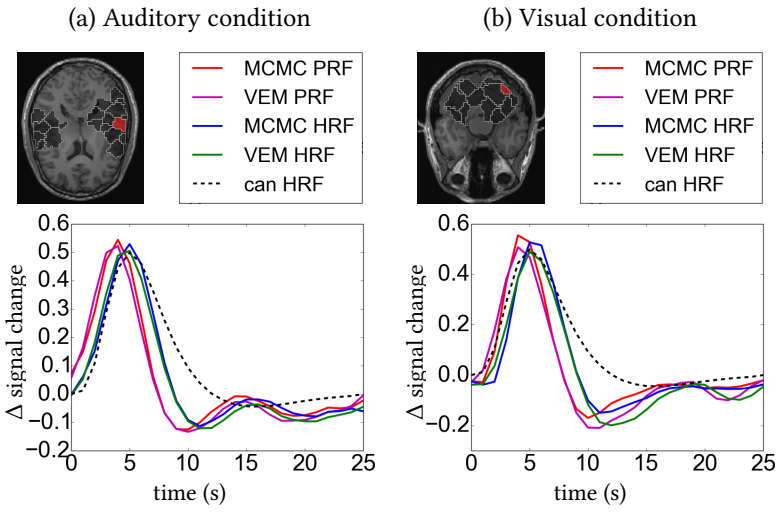



Figure 6.5: Results on real fASL data for a single subject of the AINSI database for both conditions: (a) Auditory and (b) Visual. The brain image shows the region of interest (ROI) where the response functions are estimated. As indicated in the legend, the curves represent the **PRF** and **HRF** for the MCMC approach and the **PRF** and **HRF** for the VEM approach. As a reference, we depicted the canonical HRF with a black dashed line.

are consistent for the V condition in contrast to what we observed for the A condition.

The regions of interest (ROI) in Fig. 6.5 correspond to the parcels with stronger mean HRL and PRL values for each condition respectively. The HRF and PRF estimates in these ROIs have plausible shapes and both approaches recover similar profiles. For both conditions, the PRF peaks before the HRF, as enforced by the physiological prior.

Regarding computational times, a substantial decrease was observed for VEM as compared to the MCMC solution, which is consistent with previous reports [Chaari et al., 2013]. In artificial data, a parcel containing 400 voxels takes approximately 270 s in MCMC, when 1500 iterations are done, and approximately 22 s in VEM, when 15 iterations are done. In real data, a parcel of 214 voxels takes 300 s in MCMC when 3000 iterations are done and 50 s in VEM when 30 iterations are done. The ratios are 12 and 6, respectively. However, some parts of the MCMC implementation have been optimized in C code. Time spent loading data is included in these timings.

## 6.3 Discussion

In this chapter a variational Expectation-Maximization algorithm has been proposed to address the issue of jointly detecting activity and estimating hemodynamic and perfusion responses from functional ASL data. Compared to MCMC, VEM delivers estimations in analytic form for each latent variable. Although the VEM setting remains an approximation, it facilitates the inclusion of additional information such as constraints. In particular, we considered a physiologically informed link between normalized hemodynamic and perfusion responses so as to compensate the low signal-to-noise ratio of the perfusion component. We can also easily introduce constraints in the norm of the responses or positivity. This is not the case for MCMC. Results in this chapter demonstrate a good performance of VEM when compared to MCMC at a significantly lower computation time. This suggests VEM as a fast and valid alternative for functional ASL data analysis.

# 7 Physiological models and physiological parameters

In the previous chapters, physiological models have been used as *a priori* knowledge in the estimation of the parameters of the Joint Detection Estimation model. In chapter 5, we have derived a physiological link from the so called extended Balloon model, previously introduced in chapter 2. This link has been plugged into the JDE framework to inform and strengthen the estimation of the parameters of the perfusion component, in chapter 5 for the MCMC approach and in chapter 6 for the VEM approach. The physiological parameters being used for the link operator $\mathbf{\Omega}$ were found in the literature: see e.g. [Khalidov et al., 2011, Friston et al., 2000], which we explored in chapters 5 and 6. Different settings have also been proposed in the literature for the coupling between blood flow and hemodynamic response, and a comparison can be found in [Stephan et al., 2007].

In this chapter, we present the different settings of the extended Balloon and hemodynamic models, we check their impact when being used in JDE as part of the physiological link, and we compare the convergence of the JDE algorithm when using them. The method used in this chapter is the MCMC solution with the physiological link injected as a stochastic constraint prior. The conclusions of this chapter were presented in the IEEE International Symposium on Biomedical Imaging (ISBI) conference in 2015 [Frau-Pascual et al., 2015a].

**Outcome of this chapter:**

A. Frau-Pascual, F. Forbes, and P. Ciuciu. *"Physiological models comparison for the analysis of ASL fMRI data"*. In Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, pages 1348–1351. IEEE, 2015.

## 7.1 Physiological models for fMRI

As introduced in the previous chapters, in the past decade physiological models have been described to explain the changes caused by neural activity. In [Buxton et al., 1998b, Friston et al., 2000, Buxton et al., 2004] different models have been introduced: neural coupling, which maps neural activity to ensuing cerebral blood flow (CBF); the *Balloon model*, which relates CBF to volume and deoxyhemoglobin changes, and the hemodynamic model, also referred to as BM (for BOLD model) in [Stephan et al., 2007], which relates these parameters to the blood-oxygen-level-dependent (BOLD) effect. These models thus provide a description of the physiological process underlying hemodynamic activity, from neural activation to the hemodynamics or BOLD effect measurement. In this work, we call the extended Balloon model to the combination of the Balloon and hemodynamic models. Different parameter settings have been proposed in the Balloon model formulation and several versions of the hemodynamic model have been presented in [Stephan et al., 2007]. The different behaviours induced by the various parameter sets provide flexibility to model physiological responses but also introduce more complexity.

We have previously explained that Arterial Spin Labelling (ASL) signal embodies a hemodynamic or BOLD component and a perfusion component. Recall that the ASL signal is a time series of successive alternate *control/tag* images, with inverted magnetization in the *tag* image. A hemodynamic or BOLD effect can be found in both *control/tag* images, while a perfusion effect can be captured from the *control-tag* difference. Their typical shapes are respectively described by the hemodynamic response function (HRF), and the perfusion response function (PRF). As we have seen in chapter 5, these two response functions can be estimated using the probabilistic joint detection-estimation (JDE) formalism [Vincent et al., 2013a,b], although the PRF estimation remains difficult because of the noisier nature of the perfusion component within the ASL signal [Golay et al., 2004]. For this reason, in chapter 5 we used a physiological link between PRF and HRF shapes as a prior knowledge in the JDE framework [Vincent et al., 2013a,b]. However, the physiological parameters we chose and the hemodynamic model we used were not completely in accordance with the analysis performed in [Stephan et al., 2007], where the performance of the different models was compared. In this chapter we replicate the analysis of [Stephan et al., 2007] but for ASL data and to identify which model outperforms the other for informing the link between perfusion and hemodynamics (PRF/HRF link) in the JDE analysis of ASL data. Results on real data could give us a hint on the best set of parameters to use in the Balloon and hemodynamic models.

## 7.2 A physiological link between perfusion and hemodynamics

An approximate relationship between the perfusion and hemodynamic response functions has been derived from physiological models in chapter 5. In this section, we recall the work presented in chapter 5, in which the

extended Balloon model and the hemodynamic model were used to describe a link between perfusion and hemodynamic response functions. Following the same reasoning as in [Stephan et al., 2007], we further analyse the different models to recover a more accurate perfusion/hemodynamics link, with the correct set of parameters.

The extended Balloon model (see chapter 2) describes the changes in blood flow $\boldsymbol{f}_{in}$, blood volume $\boldsymbol{\nu}$ and oxygen concentration $\boldsymbol{\xi}$ when a hemodynamic response $\boldsymbol{h}$ is ensuing neural activation. Recall:

$$\begin{cases} \frac{d\boldsymbol{f}_{in}(t)}{dt} = \boldsymbol{\psi}(t) \\ \frac{d\boldsymbol{\psi}(t)}{dt} = \eta u(t) - \frac{\boldsymbol{\psi}(t)}{\tau_\psi} - \frac{\boldsymbol{f}_{in}(t)-1}{\tau_f} \\ \frac{d\boldsymbol{\xi}(t)}{dt} = \frac{1}{\tau_m}\left(\boldsymbol{f}_{in}(t)\frac{1-(1-E_0)^{1/\boldsymbol{f}_{in}(t)}}{E_0} - \boldsymbol{\xi}(t)\boldsymbol{\nu}(t)^{\frac{1}{\bar{w}}-1}\right) \\ \frac{d\boldsymbol{\nu}(t)}{dt} = \frac{1}{\tau_m}\left(\boldsymbol{f}_{in}(t) - \boldsymbol{\nu}(t)^{\frac{1}{\bar{w}}}\right) \end{cases} \tag{7.1}$$

where $\eta$ is the neuronal efficacy weighting term; $\tau_\psi$, $\tau_f$ and $\tau_m$ are time constants respectively for signal decay/elimination, auto-regulatory feedback from blood flow and mean transit time; $\bar{w}$ reflects the ability of a vein to eject blood; $E_0$ is the oxygen extraction fraction. See chapter 2 for more details.

Next, a *hemodynamic model* links these variables to the BOLD effect. Taken together, these equations allow the precise modeling of the coupling between the cerebral blood flow and hemodynamic response. However, several competing versions of the hemodynamic model and different physiological parameters values have been described in the literature.

## 7.3 The variants of the hemodynamic model

[Buxton et al., 2004] proposed the following expression to link the hemodynamic response (HRF)[1] $\boldsymbol{h}(t)$ to physiological quantities considering the intra-vascular and extra-vascular components [Buxton et al., 1998b, Obata et al., 2004]:

$$\boldsymbol{h}(t) = V_0\left[k_1(1-\boldsymbol{\xi}(t)) + k_2\left(1-\frac{\boldsymbol{\xi}(t)}{\boldsymbol{\nu}(t)}\right) + k_3(1-\boldsymbol{\nu}(t))\right] \tag{7.2}$$

where $k_1$, $k_2$ and $k_3$ are scanner-dependent constants and $V_0$ is the resting blood volume fraction. This equation can be linearized (see the details in appendix C) into:

$$\boldsymbol{h}(t) = V_0[(k_1+k_2)(1-\boldsymbol{\xi}(t)) + (k_3-k_2)(1-\boldsymbol{\nu}(t))]\,. \tag{7.3}$$

As synthesized in [Stephan et al., 2007], where the hemodynamic model is referred to as BM, different expressions were proposed for $k_1$, $k_2$ and $k_3$: the classical ones (classical BM) in [Buxton et al., 2004], and their revised (revised BM) version in [Obata et al., 2004]. Hereafter, we will use the same notation as Stephan et al [Stephan et al., 2007]: CBM and RBM stand for models using the *classical* and *revised* expressions, respectively, and subscripts "L" and "N" for the nonlinear (Eq. (7.2)) and linear (Eq. (7.3)) expressions:

[1] To clarify, the hemodynamic response corresponds to the impulse response, namely the HRF, only when a single stimulus or neural event is considered as input.

| CBM | RBM | |
|---|---|---|
| $k_1 = (1 - V_0)4.3\vartheta_0 E_0 TE$ | $k_1 = 4.3\vartheta_0 E_0 TE$ | (7.4) |
| $k_2 = 2E_0$ | $k_2 = \epsilon r_0 E_0 TE$ | (7.5) |
| $k_3 = 1 - \epsilon$ | $k_3 = 1 - \epsilon$ | (7.6) |

where $\vartheta_0$ is the frequency offset at the outer surface of the magnetized vessel for fully deoxygenated blood, $r_0$ the slope of the relation between intravascular relaxation rate and oxygen saturation, and $\epsilon$ the ratio of intra- and extravascular signal.

In the end, we have different combinations: classical linear BM ($CBM_L$), revised linear BM ($RBM_L$), classical nonlinear BM ($CBM_N$) and revised nonlinear BM ($RBM_N$). Different values have been proposed [Friston et al., 2000], [Khalidov et al., 2011] (see Tab. 7.1) for the physiological and BM parameters, and we consider some of them in this work. According to [Behzadi and Liu, 2005], we also considered at 3T: $r_0 = 100s^{-1}$ and $\vartheta_0 = 80.6s^{-1}$. For the $\epsilon$ parameter, the values given by [Stephan et al., 2007] have been used: 0.4, 1 and 1.43.

| | $\eta$ | $\tau_\psi$ | $\tau_f$ | $\tau_m$ | $\tilde{w}$ | $E_0$ | $V_0$ |
|---|---|---|---|---|---|---|---|
| [Friston et al., 2000] | 0.5 | 1.25 | 2.5 | 1 | 0.2 | 0.8 | 0.02 |
| [Khalidov et al., 2011] | 0.54 | 1.54 | 2.46 | 0.98 | 0.33 | 0.34 | 0.01 |

Table 7.1: Physiological and BM parameters used in [Friston et al., 2000] and [Khalidov et al., 2011].

Other parameters were tested in the context of this thesis, but we do not use them because they were either too similar to the ones we use here and the impact can be quantified with the present ones, or they gave non plausible shapes for human fMRI data.

### 7.3.1 Physiological linear relationship between response functions

Akin to [Frau-Pascual et al., 2014b] and here explained in chapter 5, starting from the system of differential equations, we derive an approximate relationship between the PRF (named $\boldsymbol{g}$) and the HRF (named $\boldsymbol{h}$). Both PRF and HRF are percent signal changes. By linearising the system of differential equations around the resting point as explained in chapter 5, and considering the hemodynamic model equations (linear (7.3)/nonlinear (7.2) forms), we find a linear relationship between $\boldsymbol{h}$ and $\boldsymbol{g}$, $\boldsymbol{g} = \boldsymbol{\Omega h}$ which reads:

$$\boldsymbol{\Omega} = V_0^{-1} \left((k_1 + k_2)\boldsymbol{B} + (k_3 - k_2)\boldsymbol{A}\right)^{-1} \tag{7.7}$$

when Eq. (7.3) holds and

$$\boldsymbol{\Omega} = V_0^{-1} \left(k_1 \boldsymbol{B} + k_2(\boldsymbol{B} - \boldsymbol{A})(\boldsymbol{I} - \boldsymbol{A})^{-1} + k_3 \boldsymbol{A}\right)^{-1} \tag{7.8}$$

with

$$\begin{cases} \boldsymbol{A} &= -\frac{1}{\tau_m}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1} \\ \boldsymbol{B} &= -\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tau_m}\right)^{-1}\left(\gamma \boldsymbol{I} - \frac{1-\tilde{w}}{\tilde{w}\tau_m^2}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right). \end{cases}$$

when Eq. (7.2) holds instead. Here $\gamma = \frac{1}{\tau_m}\left(1 + \frac{(1-E_0)\ln(1-E_0)}{E_0}\right)$.

Hence, we have different $\mathbf{\Omega}$ matrices depending on the Balloon model parameters (see Tab. 7.1), the classical or revised expression for $k_1$, $k_2$ and $k_3$ (see Eqs. (7.4)-(7.6)), directly impacted by $\epsilon$ parameter, and the model (Eqs. (7.7)-(7.8)). Changing $\mathbf{\Omega}$ might therefore affect the PRF and HRF estimation results from ASL data. Identifying the best combination of these ingredients is the issue we address here.

## 7.4 Perfusion/hemodynamics link analysis on simulated data

The matrix $\mathbf{\Omega}$ varies depending on the model under consideration: $\text{CBM}_L$, $\text{RBM}_L$, $\text{CBM}_N$ and $\text{RBM}_N$ with different possible values for $\epsilon$, and on which set of parameters we use (see [Friston et al., 2000] or [Khalidov et al., 2011]). Here, we simulate a PRF by applying $\boldsymbol{g}_m = \mathbf{\Omega}_m \boldsymbol{h}_{can}$ to the canonical HRF shape ($\boldsymbol{h}_{can}$) using different $\mathbf{\Omega}_m$ ($m$ coding the model that we use) to find out which factors have a stronger impact on $\mathbf{\Omega}$. We choose as canonical HRF the one presented in [Glover, 1999], and we use it as a reference for the PRF.

Fig. 7.1(a) illustrates the cases for which we observed the largest difference between the simulated PRF shapes when applying $\boldsymbol{g}_m = \mathbf{\Omega}_m \boldsymbol{h}$ when $\mathbf{\Omega}_m$ is defined either by (7.7) or (7.8) using the parameters proposed in [Khalidov et al., 2011]. We also found variability between the generated PRFs associated with different $\epsilon$ values. Fig. 7.1(b) shows this fact for the $CBM_L$ model.

One interesting question to answer is whether there is a significant impact of the different hemodynamic models and of the different sets of parameters in the generation of the PRF using $\boldsymbol{g} = \mathbf{\Omega}\boldsymbol{h}$. To draw significant conclusions about the statistical significance of the measured differences between PRF and canonical HRF, for each $\epsilon$ value we performed a 2 way-ANOVA including the model type ($CBM_L$, $RBM_L$, $CBM_N$ and $RBM_N$) and the setting of parameters (see Tab. 7.1) as the two factors of interest. We entered the squared differences between the canonical HRF and the different PRFs as observations in our analysis. For $\epsilon = 0.4$ only, we identified a significant difference between the sets of parameters, but none between classical and revised or linear and nonlinear BM models at a 0.01 level of significance (F-test: $F = 38.98$, p-val $< 10^{-4}$). This result is also confirmed by the discrepancy depicted in Fig. 7.1(b) between the blue curve and the other traces. To sum up, the setting of physiological parameters and $\epsilon$ can impact the quality of the link between perfusion and hemodynamic response functions, whereas choosing a particular BM model has a limited influence.

In what follows, we address the same concern on real ASL data acquired along an fMRI experiment.

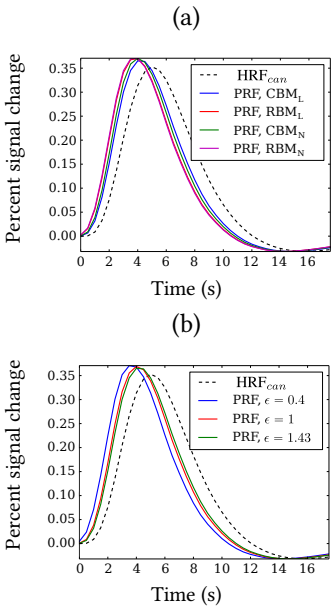



Figure 7.1: PRFs resulting from $\mathbf{\Omega}_m \boldsymbol{h}_{can}$ when using model $m$, for parameters in [Khalidov et al., 2011] (a), and for the $CBM_L$ model (b).

## 7.5 Impact of the physiological parameters in the HRF and PRF shapes

Every parameter has a biological interpretation, and therefore have to be kept within a range of biological validity. A variation in the physiological

parameters creates different HRF and PRF shapes. As we have already seen in chapter 2 and we can see in figure 7.2, each parameter affects differently to the shapes:

- $\eta$: Neuronal efficacy models BOLD response amplitude without changing the shape. There is a point of amplitude saturation ( cannot be higher). Typical values are $0.54 \pm 0.633$ owing to mean $\pm$ standard deviation.
- $\tau_\psi$: Signal decay or elimination. If it decreases it removes the final undershoot and if it increases oscillations appear after the undershoot. Typical values are $0.65 \pm 0.862$.
- $\tau_f$: Auto-regulation feedback from blood flow. This represents the time that takes the blood flow to decrease after the peak (post-stimulus undershoot). Increasing it makes BOLD smoother and softens the undershoot. Typical values are $0.41 \pm 0.654$.
- $\tau_m$: Transit time. If it increases it slows down BOLD signal dynamics, which dilates the HRF. Typical values are $0.98 \pm 1.015$.
- $\tilde{w}$: The Windkessel parameter influences the amplitude of the response. It represents the degree of nonlinearity of the system. Typical values are $0.33 \pm 0.596$.
- $E_0$: Oxygen extraction changes the BOLD response: dip and amplitude. Typical values are $0.34 \pm 0.659$, although it remains positive. Therefore, values have to be kept from 0 to 1.
- $V_0$: Resting venous cerebral blood volume fraction. Typical values are $0.02 \pm 0.152$.

The typical range values are proposed in [Mesejo et al., 2016].

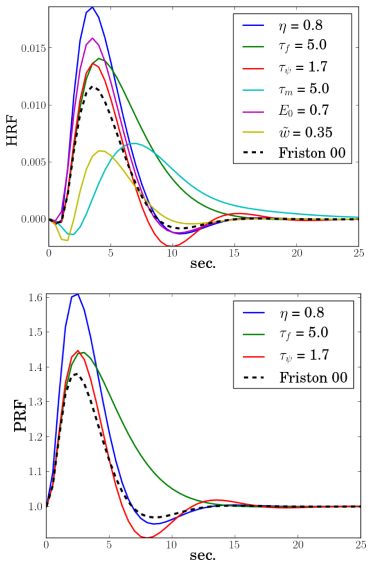



Figure 7.2: Effect of the physiological parameters on the BRF and PRF shapes. The parameters values proposed in [Friston et al., 2000] are used except for one parameter whose identity and value is modified as indicated in the plot.

## 7.6 Impact of the extended Balloon model settings on real data

Here, we performed ASL JDE analyses on ASL fMRI data, considering the different models and parameter settings for $\boldsymbol{\Omega}$ explored in section 7.3. The analysis was performed on 8 individuals, although the results are shown for 1 subject only. The same conclusions hold for the other subjects.

The ASL data used has already been described in chapters 5 and 6. Recall that the experiment (fast event-related paradigm with mean $ISI = 5.1$ s) was designed to map primary auditory and visual cortices (auditory and visual stimuli), with 291 scans, repetition time $TR = 3$ s, echo time $TE = 18$ ms, FoV 192 mm, dimensions $64 \times 64 \times 22$ voxels (resolution of $3 \times 3 \times 3.5$ mm$^3$). The tagging scheme used was PICORE Q2T, with $TI_1 = 700$ ms, $TI_2 = 1700$ ms. Two regions of interest in the right temporal and left occipital lobes were defined manually as parcels of interest to study the evoked response in the auditory and visual cortices, respectively.

Fig. 7.3 shows the perfusion and hemodynamic response functions estimated using different $\boldsymbol{\Omega}$ matrices for $p(\boldsymbol{g} \mid \boldsymbol{h})$ in the ASL JDE inference, after 3000 MCMC iterations. Here, the two parameter sets in [Friston et al.,

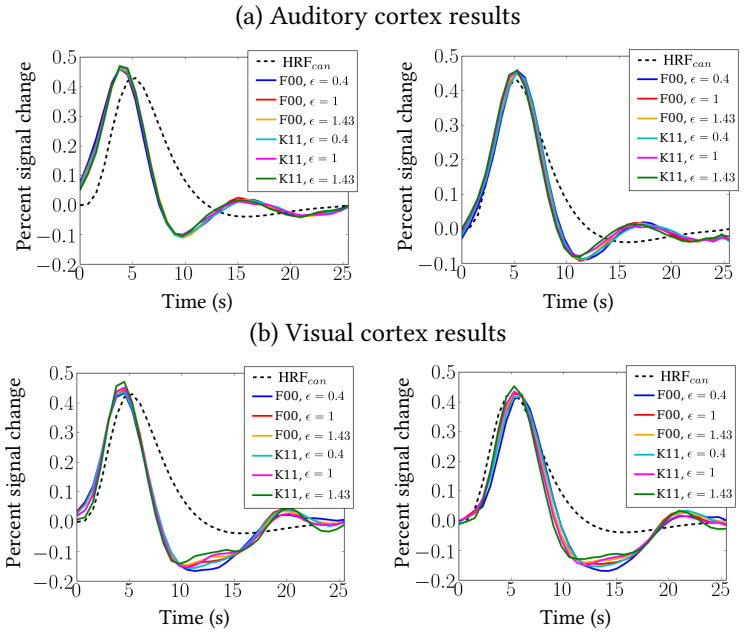



Figure 7.3: PRF (left) and HRF (right) estimates for model $RBM_N$ with parameters in [Friston et al., 2000] (F00 label) and [Khalidov et al., 2011] (K11 label), considering different $\epsilon$ values, estimated in auditory **(a)** and visual **(b)** cortices.

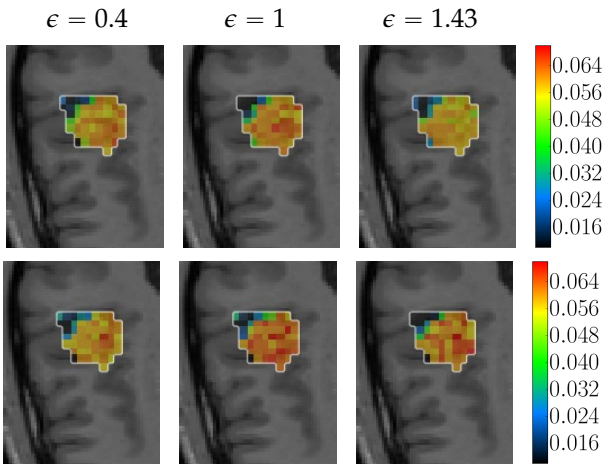



Figure 7.4: Auditory cortex PRLs for model $RBM_N$ with parameters: [Friston et al., 2000] (top), [Khalidov et al., 2011] (bottom), considering $\epsilon$ (left to right) 0.4, 1 and 1.43.

2000, Khalidov et al., 2011] were tested and $\mathbf{\Omega}$ was computed using the abovementioned models and $\epsilon$ values. Fig. 7.3 also depicts the canonical HRF (dashed line), which is in accordance with the HRF estimates for both methods. We observed very similar shapes, as well as similar perfusion response levels in Fig. 7.4 for auditory cortex using $RBM_N$. A variation in PRF could impact the PRLs retrieved.

Fig. 7.5 shows the convergence of the relative reconstruction error over MCMC iterations for the different parameter settings. The relative reconstruction error measures how good is the fitting with respect to the signal measured: $e_{rec} = \frac{\|\boldsymbol{y}_{measured} - \boldsymbol{y}_{estimated}\|^2}{\|\boldsymbol{y}_{measured}\|^2}$. The lower the reconstruction error the better the fitting. Hemodynamic models (BM) are not all shown as they have similar convergence speed. Interestingly, we observed a stronger variability between the two parameter sets as compared to changing the $\epsilon$ value. Important results are the lower relative reconstruction error of the parameters proposed in [Khalidov et al., 2011] from the first iteration, and the better performance in both sets of parameters, [Khalidov et al., 2011] and [Friston et al., 2000], for $\epsilon = 1.43$. For this reason we can consider the

combination [$\epsilon = 1.43$ and parameters as in [Khalidov et al., 2011]] as the one performing best and offering promising perspectives. However, after 3000 iterations, the algorithm converges to good parameter estimates in all cases.

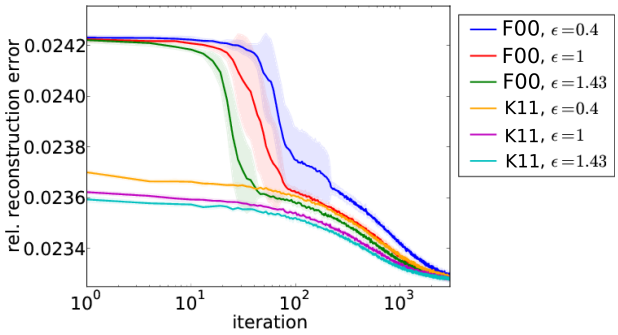



Figure 7.5: Convergence of the averaged relative reconstruction error over 10 runs for the auditory cortex and model $RBM_N$. Standard deviations are shown with shaded colors. F00 and K11 labels correspond to parameters proposed in [Friston et al., 2000] and [Khalidov et al., 2011], respectively.

## 7.7 Discussion

A physiological link has been described, combining the Balloon and hemodynamic models, to achieve a better estimation of parameters in an ASL JDE framework. Different versions of the hemodynamic model have been described in the literature, and different parameter settings for the Balloon model have also been proposed. In this paper, we considered them all together to assess their impact in the context of ASL data analysis. On simulated data the selection of physiological parameters used in the Balloon model as well as the setting of $\epsilon$, were more critical than that of the hemodynamic model itself. On real ASL data we confirmed this finding with a faster convergence in the joint estimation of perfusion and hemodynamic components of the signal in the auditory and visual regions.

## 7.8 Perspectives

This work is an attempt to assess the impact of changing models and parameters in the estimation of JDE quantities. One interesting observation is that although the results of JDE are not very different for different parameter settings, the convergence is much faster if the correct model is used. If we assume that the good model and parameters fit better the JDE estimation and a faster convergence is achieved, then this can be interpreted as though these parameters are the most convenient for this specific data. For this reason, we consider as a possible future work the inclusion of a step to estimate the physiological parameters inside the JDE framework. With a good estimation of the physiological parameters, we make also the JDE run faster and be more accurate.

### 7.8.1 Estimation of the physiological parameters from the HRF and PRF

The estimation of the physiological parameters in the Balloon model from BOLD fMRI has already been studied in [Mesejo et al., 2015, 2016]. The extended Balloon model used in [Mesejo et al., 2016] includes two non-linear equations to explain excitatory and inhibitory neuronal activity,

besides the previously presented equations, that add other parameters to the model. In total, 15 physiological parameters are estimated and, since identifiability issues can arise when estimating them, statistical prior knowledge about the parameters is used in the estimation to constrain the possible set of solutions. For this, the target function to minimize when estimating the parameters is derived with Bayesian modelling and includes *a priori* knowledge. Within this framework local optima are also avoided. The approach proposed in [Mesejo et al., 2016] to minimize this function is based on metaheuristics (MH): an Evolutionary Computation global search method called Differential Evolution (DE). In [Mesejo et al., 2016], this method provides stable and realistic estimates of the physiological parameters in rat fMRI data, compared to the *de facto* standard Expectation Maximization Gauss-Newton (EM/GN) approach, that is implemented in SPM[2].

[2] http://www.fil.ion.ucl.ac.uk/spm/

### 7.8.2 Adaptative prior for JDE ASL

In previous chapters, we have introduced the Gaussian priors used for $\boldsymbol{h}$ and $\boldsymbol{g}$, considering a certain mean and introducing smoothness with the covariance matrix. The mean can be the HRF and PRF generated with the Balloon model:

$$\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{h}_{balloon}, \Sigma_h) \tag{7.9}$$

$$\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{g}_{balloon}, \Sigma_g) \tag{7.10}$$

The physiological parameters can be either extracted from the literature as in previous chapters or estimated as in [Mesejo et al., 2016].

We propose to estimate the physiological parameters of the Balloon model from the shapes of the HRF and PRF estimated using JDE. Then, this model with these physiological parameters can be used as prior information in JDE for its estimation, since the estimated set of physiological parameters will give physiologically plausible HRF/PRF shapes, and will therefore "denoise" the JDE result. However, this circularity could drive our estimation to unphysiological shapes, or make it more computationally expensive with no improvement in the estimation. For this reason, the estimation of the physiological parameters would be done when the algorithm is already close to convergence, as a refinement step. This way, the estimated parameters would give us the closest physiologically plausible response shape to the estimated one. This amounts to "projecting" the result onto the physiologically plausible space of HRF/PRFs, discarding noise and artefact effects that give non physiologically plausible response shapes. The prior can adapt during some iterations of the JDE algorithm to let the rest of the parameters converge.

In any case, estimating the physiological parameters is interesting in itself, since it describes the physiological changes undergone during brain activity in a given subject and region.

This chapter opens a line of research for the analysis of physiological parameters and its use in the estimation of response functions from fMRI data in the context of the Joint Detection Estimation framework, since it was shown that it affects the convergence of the algorithm.

# 8 Methodology comparison in the analysis of HEROES dataset

In this chapter, we analyse a dataset with both BOLD and ASL fMRI data. We use the proposed JDE method presented in chapter 6, and compare it to a classical general linear model (GLM). First, we present the dataset and the fMRI setting, then the results with both methods for BOLD and functional ASL (fASL). We also present the measured basal cerebral blood flow (CBF) maps, for subject and group levels, and compare it to the perfusion baseline estimated from the functional data.

## 8.1 Data: design and acquisition

The HEROES dataset has been acquired at Neurospin[1], CEA Saclay, in the context of the HEROES initiative[2]. It contains anatomical T1-weighted images, blood-oxygen-level-dependent (BOLD) data, basal Arterial Spin Labeling (ASL), functional ASL data, and other images acquired for the quantification of cerebral blood flow (CBF). The dataset contains data of 13 subjects, 7 men and 6 women, of age between 20 and 29. They all are right-handed and they were asked not to drink coffee before the acquisition. Two subjects were discarded for this analysis because of acquisition differences.

[1] http://www-centre-saclay.cea.fr/fr/NeuroSpin

[2] HEROES stands for "HEmodynamics-infoRmed atlas of brain functiOnal and vascular territoriES from multimodal MR images". Based on the information obtained from the combination of different MR modalities for a group of healthy subjects, the goal was to build an atlas of functional and vascular territories with homogeneous hemodynamic properties to address clinical questions.

[3] Magnetization prepared rapid gradient-echo

### 8.1.1 Image acquisition

After a localizer scan (13 s) [Pinel et al., 2007], several types of data were acquired during one acquisition:

- Anatomical T1-weighted image acquired with a MP-RAGE[3] acquisition sequence, with resolution $1 \times 1 \times 1$ mm$^3$ comprising 160 sagittal slices. TR = 2.3 s, TE =2.98 ms, field of view FoV = 256 mm and phase FoV = 93.8%.
- BOLD data was acquired with a gradient echo EPI acquisition sequence. The images contain 165 scans with TR= 2.5 s, TE= 30 ms and resolution $3 \times 3 \times 3$ mm$^3$. This results in 42 slices that were acquired interleaved. The field of view used is FoV = 192 mm and phase FoV = 100%.
- Functional ASL data was acquired using pulsed ASL and a Q2TIPS PICORE scheme [Luh et al., 1999] as in figure 8.1. Two in-plane presaturation pulses before the inversion pulse improve the cancellation of signal from static tissues between tag and control states. The gradient in gray is alternately applied to tag and control states. Periodic saturation

pulses are applied from $TI_1$ to $TI_{1S}$ to improve saturation of the remaining tagged blood in the region being tagged [Luh et al., 1999]. EPI acquisition is used to acquire the images at time $TI_2$. The time series contain 165 scans acquired with TR=2.5 s and TE=11 ms. The first scan in the time series corresponds to an $M_0$ image. It has a spatial resolution of $3 \times 3 \times 7.5$ mm$^3$, the slice thickness being larger than that of BOLD images. Each volume contain 14 slices acquired sequentially (down-up). Slices have actually a 6 mm width and there is a 1.5 mm space between slices. The field of view used is FoV = 192 mm and phase FoV = 100%. The output of the scanner contains three images: raw ASL, perfusion weighted image, and relative CBF.

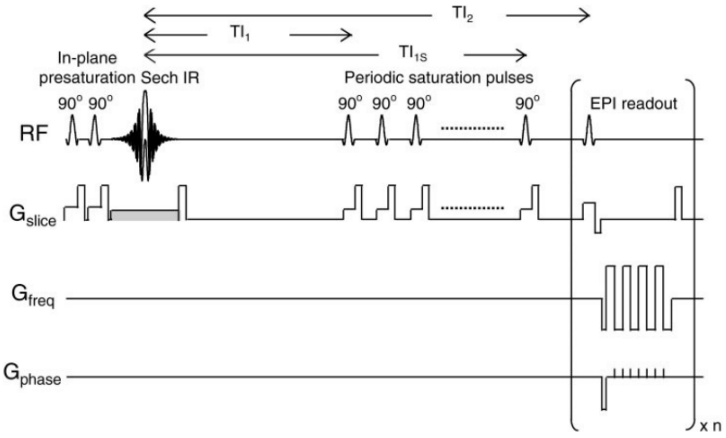



Figure 8.1: Functional ASL acquisition scheme used in the HEROES dataset. Acknowledged from [Luh et al., 1999].

- Perfusion baseline ASL contains 151 scans with TR = 2.5 s and resolution $3 \times 3 \times 7.5$ mm$^3$ as in fASL. The first scan in the time series corresponds to an $M_0$ image. The output of the scanner contains three images: basal raw ASL, basal perfusion weighted image, and basal relative CBF.
- B1 Mapping image. The resolution is $4 \times 4 \times 4$ mm$^3$ with 36 slices, and two images are acquired for magnitude and phase, respectively. This image determines the flip angle distribution to be used in the T1 mapping. The field of view used is FoV = 256 mm and phase FoV = 100%. TR = 14 ms and 4 TE are used: 3.061 ms, 3.061 ms, 4.5 ms, 7 ms. Images were acquired sequentially (down-up).
- T1 PSSFP[4] with angles 20° and 5°. They have resolution $1 \times 1 \times 1$ mm$^3$ which results in 144 slices acquired interleaved. Other parameters used are TR = 14 ms, TE = 3.06 ms, FoV = 256 mm. These images are acquired to be used for the T1 mapping.

[4] partially Spoiled Steady State Free Precession

The total acquisition time is 1 h 10 min. The functional BOLD and ASL sessions were collected with two different versions (conditions ordering changed randomly between the two) of the same experimental paradigm.

### 8.1.2 Data preprocessing and postprocessing

BOLD and ASL fMRI data have been preprocessed following the steps in section 3.4 using the python toolbox *process-asl*[5]. It has been developped by

[5] https://github.com/process-asl

S. Bougacha under the supervision of P. Ciuciu, and reproduces the pipeline developped by J. Warnking at the Grenoble Institute of Neuroscience (GIN). The toolbox uses *nipype*[6] to create a preprocessing pipeline using SPM[7].

A CBF quantification module has been added to this toolbox for pre- and post-processing of ASL fMRI data, following the recommendations in [Alsop et al., 2015]. CBF quantification has been coded by S. Bougacha under the supervision of P. Ciuciu, and in collaboration with A. Vignaud at Neurospin (CEA Saclay) and J. Warnking at GIN.

[6] http://nipy.org/nipype/0.10.0/

[7] Statistical Parametric mapping www.fil.ion.ucl.ac.uk/spm/

### 8.1.3 Experimental design

We chose an experimental paradigm with a block design to increase the statistical power of the effects, since the ASL signal strength is quite low (around 1% perfusion baseline variation) [Golay et al., 2004]. However, we used mini-blocks to be able to study response dynamics. With TR 2.5 s, each session takes 400 s and has 16 blocks with 4 blocks per condition. Each block lasts 15 s and it is followed by 10 s of rest so the signal has time to go back to baseline, completing 25 s cycles. The paradigm consists of visual, motor and auditory tasks. When a block starts, a blinking checkerboard is shown, and therefore the visual task is performed (see figure 8.2). After a certain delay of 2.5 s or 5 s a blinking half-circle blue or red is displayed in the left or right hemifield respectively and at the same time "beep" sounds are delivered at the same pace/frequency as the one of the flashing half-circle in the ipsilateral ear (see figure 8.2). This delay allows to avoid habituation to the stimuli. The subject is instructed to push a button at the frequency of the "beep". This auditory/motor mini-block lasts 10 s, so if the delay is 2.5 s (as in the first block of figure 8.2) we have also 2.5 s at the end of only visual task. Otherwise, we have 5 s at the beginning of the block of only visual task and then the three tasks simultaneous until the end of the block (as in the second block of figure 8.2).

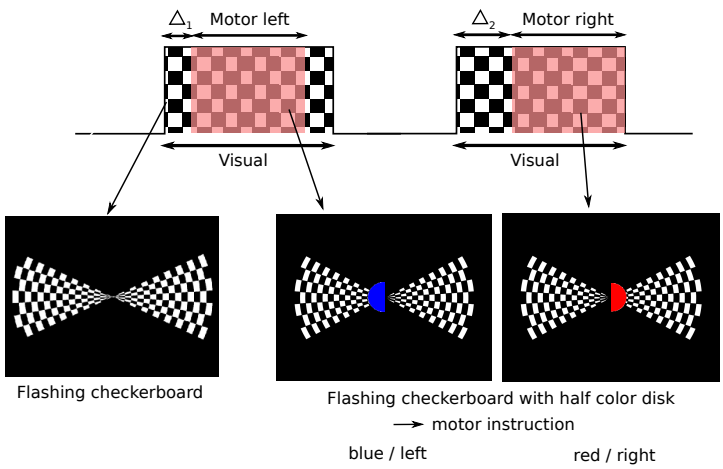



Figure 8.2: After a certain delay $\Delta_i$ a half-circle is shown blinking and at the same time a "beep" sounds. For the left condition, the half-circle is blue in the left side, and the beep sounds in the left ear. For the right condition, the half-circle is red in the right side, and the beep sounds in the right ear. The subjects start pushing the button at the frequency of the "beep" while the half-circle is shown.

## 8.2 BOLD data analysis

BOLD fMRI data is used as the reference functional MRI technique here. The design matrix has been constructed by considering 3 conditions: left motor-auditory condition, right motor-auditory condition and visual condition. Each block lasts 15 s, and it is followed by 10 s of rest. We have 4 blocks per condition, which makes 16 blocks in total. In figure 8.3 polynomial drifts were used.

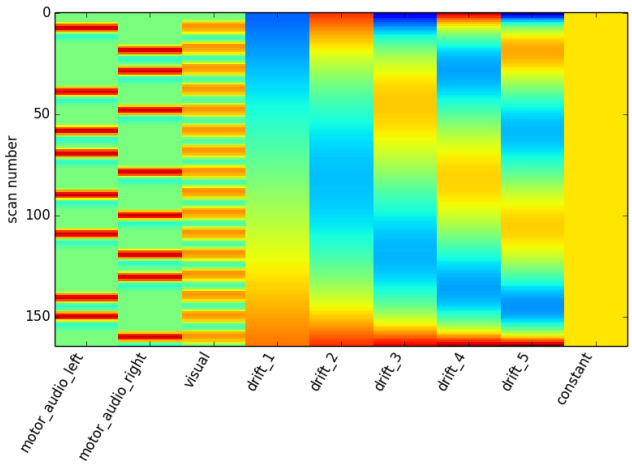



Figure 8.3: Design matrix used in the General Linear Model. The 3 task related regressors, 5 polynomial drifts and a constant regressor to capture the mean.

In this section, BOLD data has been analysed using a general linear model (GLM)[8], the joint detection-estimation (JDE) framework using the canonical HRF, and the JDE framework estimating an HRF per parcel. For GLM, we need data that has been previously smoothed in order to account for spatial correlation and reduce noise. For the HEROES dataset, with resolution $3 \times 3 \times 3$ mm$^3$, we have applied a smoothing of kernel width 5 mm full width at half maximum (FWHM). For JDE, we do not need smoothed data because the spatial correlation is modelled with a Markov Random Field (MRF), as explained in chapter 5. We use the Willard parcellation (see section 4.7). By using the JDE with fixed canonical HRF, we show an intermediate step between GLM and JDE. We can therefore see the impact of the estimation of the HRF by comparing JDE estimating the HRF and with fixed HRF shape. We also see the impact of the parcel based approach with the MRF modelling using non-smoothed data, by comparing GLM and JDE with fixed canonical HRF. The impact of the smoothing is discussed in subsection 8.3.3.

[8] Note here that GLM can be given some flexibility by adding regressors for capturing delay and dispersion, but that for simplicity we have compared to the simplest version of it. The addition of these regressors would need the use of F-tests instead of t-tests in the statistical analysis.

We have analysed[9] the 2 sessions of all 13 subjects and in the following we show subject level results for a session of one subject, and group level results for all 13 subjects and both sessions. The subject illustrated in the subject-level results has been chosen to be representative of the group. It has a high correlation score of subject-to-group activation maps compared to the rest of the subjects in BOLD and ASL results, for all the methods presented (see figures 8.10 and 8.18). The correlation score is better explained in section 8.2.2.

[9] This analysis was done in a joint work with Thomas Perret.

For the evaluation of activation detection, contrast maps are shown. In the case of GLM, we show contrast $\gamma^t \beta_j$ z-score maps and the impact of a

correction for multiple comparisons in these tests. We also show $-\log_{10}$(p-vals), in order to focus on the order of magnitude of the p-value. For JDE, as explained in chapter 4, we cannot use the same maps. We show the contrast effect maps $\gamma^t \boldsymbol{a}_j$ and the posterior probability maps (PPM) per experimental condition, that can be computed as

$$p(a_j^m > \delta | \boldsymbol{y}_j) > \alpha \tag{8.1}$$

Note that we have two thresholds to set. We set $\delta$ to get a posterior probability distribution, and $\alpha$ is the threshold that we set to see a certain level of significance, as we do for the p-values. See section 4.6 for more details on PPMs. In BOLD JDE, we chose a threshold $\delta$ for each experimental condition $m$ as the intersection of the two Gaussian densities of the Gaussian Mixture Model (GMM) that represent active and non-active voxels:

$$\frac{(\delta - \mu_1^m)^2}{v_1^m} - \frac{(\delta - \mu_0^m)^2}{v_0^m} = \log\left(\frac{v_1^m}{v_0^m}\right) \tag{8.2}$$

$\mu_i^m$ and $v_i^m$ being the parameters of the GMM in $a_j^m$ corresponding to active ($i = 0$) and non-active ($i = 1$) voxels for experimental condition $m$. Figure 8.4 shows the Gaussian densities of JDE together with the histogram of the amplitude levels $\boldsymbol{a}$ corresponding to a parcel. We use logarithmic scale: $-\log_{10}(1-\text{ppm})$.

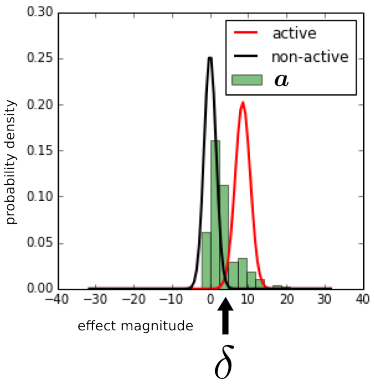



Figure 8.4: Histogram of response levels corresponding to the voxels in one parcel (around 200 voxels) in green, and overlapping the Gaussian densities corresponding to **active** and **non-active** voxels. The threshold $\delta$ chosen is the intersection between these densities for each experimental condition.

In order to be able to see the differences between methods, we have chosen representative slices for each region of interest. We show an axial slice containing the occipital visual cortex for the activation related to the visual task and a coronal slice for the activation related to the motor task. A part of the auditory cortex can also be seen in this coronal view. See for example figure 8.5.

Results on HRF estimation are also illustrated for some of the parcels of interest. HRFs are expected to be similar between different sessions of the same subject. Note that HRFs are usually well estimated in the parcels where some activation is found, but not in the others. The time step used in the HRFs is equal to the TR because our paradigm has a block design and the beginning of the blocks is not jittered (cycles task-rest are 25 s long). Therefore, the HRF points being measured are always the same: multiples of the TR. This implies a better measure of the estimated points, but a low confidence outside these measurement points. For this reason, the temporal resolution chosen for the estimated HRF is low. Note that we could use a higher temporal resolution because the prior on the HRF enforces a smooth interpolation. This is particularly useful when the measurements are jittered and we have some knowledge about the temporal points between TR points (e.g. fast event related designs). In our case, a higher resolution would just increase the degrees of freedom and the computational load of the estimation.

### 8.2.1 Single subject

The subject chosen is representative of the group under study, as we show in the next section (see figure 8.10). Results for GLM in figure 8.5 show big activated clusters in the visual, motor and auditory cortices with a high significance, as expected.

Figures 8.6 and 8.7 show contrast effect maps for the three contrasts described before and the PPM maps of the three experimental conditions. For the PPM maps, we use as threshold the intersection of the two Gaussian densities of the GMM that are imposed in the response levels and that correspond to active and non-active classes in JDE. Activations for JDE are much less spread than in GLM, probably because we are using non-smooth data. Note here that there is not much visual difference between JDE estimating the HRF or using the canonical HRF.

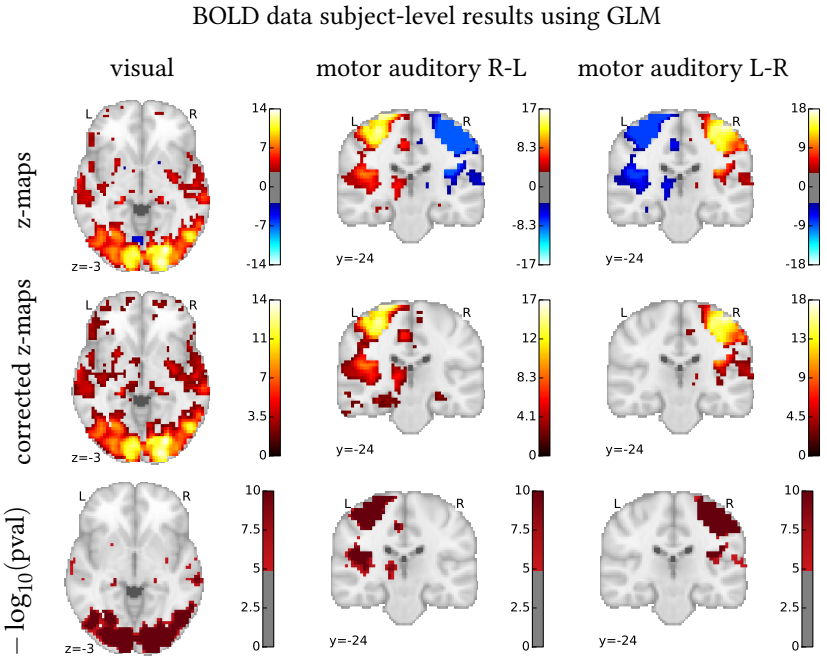



Figure 8.5: Subject level z-maps of BOLD data computed using GLM. From top to bottom: z-maps, z-maps corrected for multiple comparisons using a FDR of 0.05, and $-\log_{10}(\text{pval})$ of the FDR-corrected z-maps. They are thresholded to show values from $10^{-5}$ to $10^{-10}$. Note that the FDR-corrected z-maps and the p-value maps are one-sided.

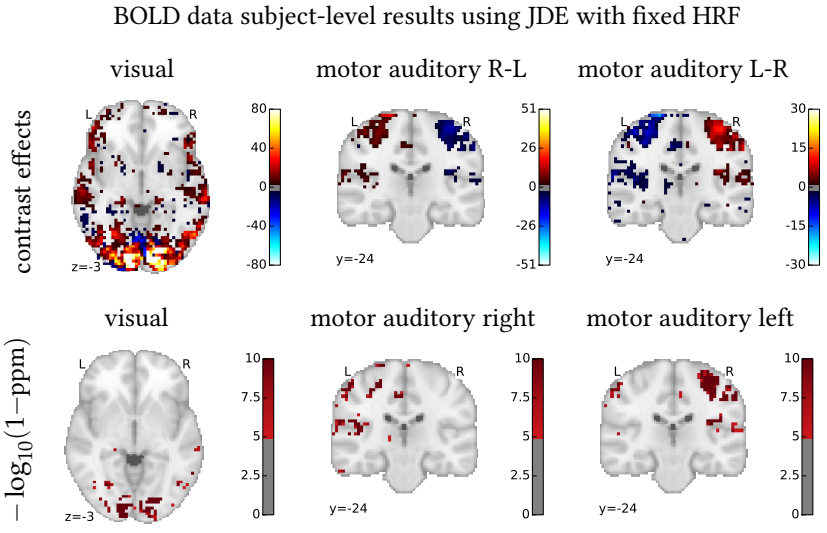



Figure 8.6: Subject level maps of BOLD data computed using JDE, using the canonical HRF. First row are the contrast effect maps $\gamma^t \boldsymbol{a}_j$. Second row are the posterior probability (PPM) maps of each experimental condition $m$ in logarithmic scale using as $\delta$ threshold the intersection of the two Gaussian densities of the GMM imposed in the response levels (active/non-active). They are thresholded to show values from $10^{-5}$ to $10^{-10}$.

For the case of JDE estimating the HRF, we can also check the shapes of the estimated responses. Figure 8.8 shows the HRFs estimated in 8 parcels containing regions expected to elicit evoked activity in the auditory, visual and motor cortices of both hemispheres. As we can see, we get similar shapes for both sessions in most regions, confirming that the estimation

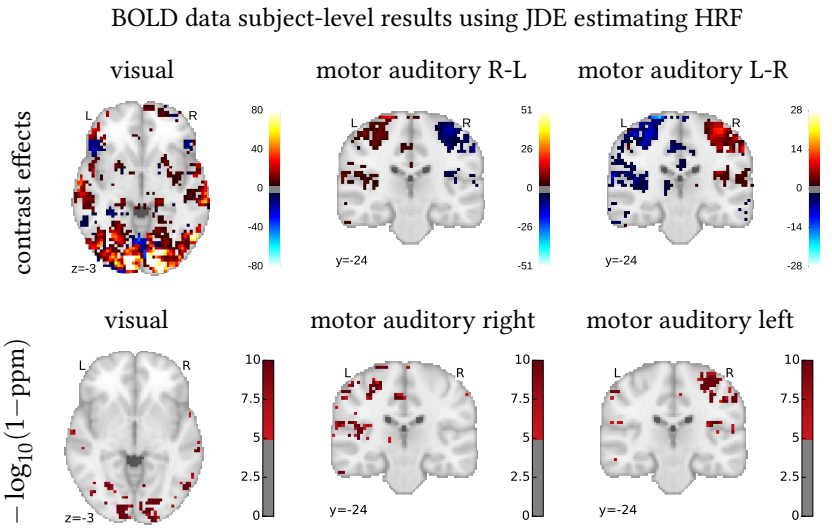



Figure 8.7: Subject level maps of BOLD data computed using JDE and estimating the HRF in each parcel. First row are the contrast effect maps $\gamma^t a_j$. Second row are the posterior probability (PPM) maps of each experimental condition $m$ in logarithmic scale using as $\delta$ threshold the intersection of the two Gaussian densities of the GMM imposed in the response levels (active/non-active). They are thresholded to show values from $10^{-5}$ to $10^{-10}$.

was consistent. The HRFs of the right auditory cortex and the left motor cortex differ slightly between sessions. In the PPM maps, these regions are less active than the rest. Left and right regions have similar HRFs too, and visual regions are all very similar. It is worth noting that auditory cortex HRFs peak slightly before 5 s, visual cortex HRFs slightly after 5 s and motor cortex HRFs peak at 5 s.

### 8.2.2 Group level statistics

Group level statistics were performed on the BOLD data of 13 subjects, considering 2 sessions per subject. Figure 8.9 shows the comparison of group level z-maps using GLM, JDE with the canonical HRF, and JDE estimating the HRF. The group level contrast is done considering random effects analysis (see section 3.2.2). A one-sample t-test is conducted on the contrast effect maps computed as $\gamma^t \beta_j$ on GLM, and $\gamma^t a_j$ on JDE, and the z-maps were corrected for multiple comparisons using the FDR criterion with a threshold 0.05. The clusters of size lower than 50 voxels were also thresholded. We observe large active regions. In the case of the visual contrast we observe that visual occipital cortex, superior colliculus (SC) and lateral geniculate nucleus (LGN) are active. For the motor and auditory contrast, we find activations in the motor and auditory cortices, as expected. Note that when using JDE we have a higher sensitivity to activations, which is consistent with the literature [Handwerker et al., 2004]. We might think that this is due to the spatial smoothing on GLM. However, the same analysis has been performed without spatial smoothing and z-maps had lower or similar values than when smoothing (see subsection 8.3.3), although this is probably due to a worse correspondence across subjects. The same analysis was done using JDE on smoothed data and the results had the same z-score levels as when using non-smoothed data, with smoother activation patterns (see also subsection 8.3.3). In the subject level analysis, the same results were reported too with less noisier maps and more defined clusters. However, the HRF estimation was slightly worse. When we estimate the HRF in JDE, we observe bigger active regions around the LGN

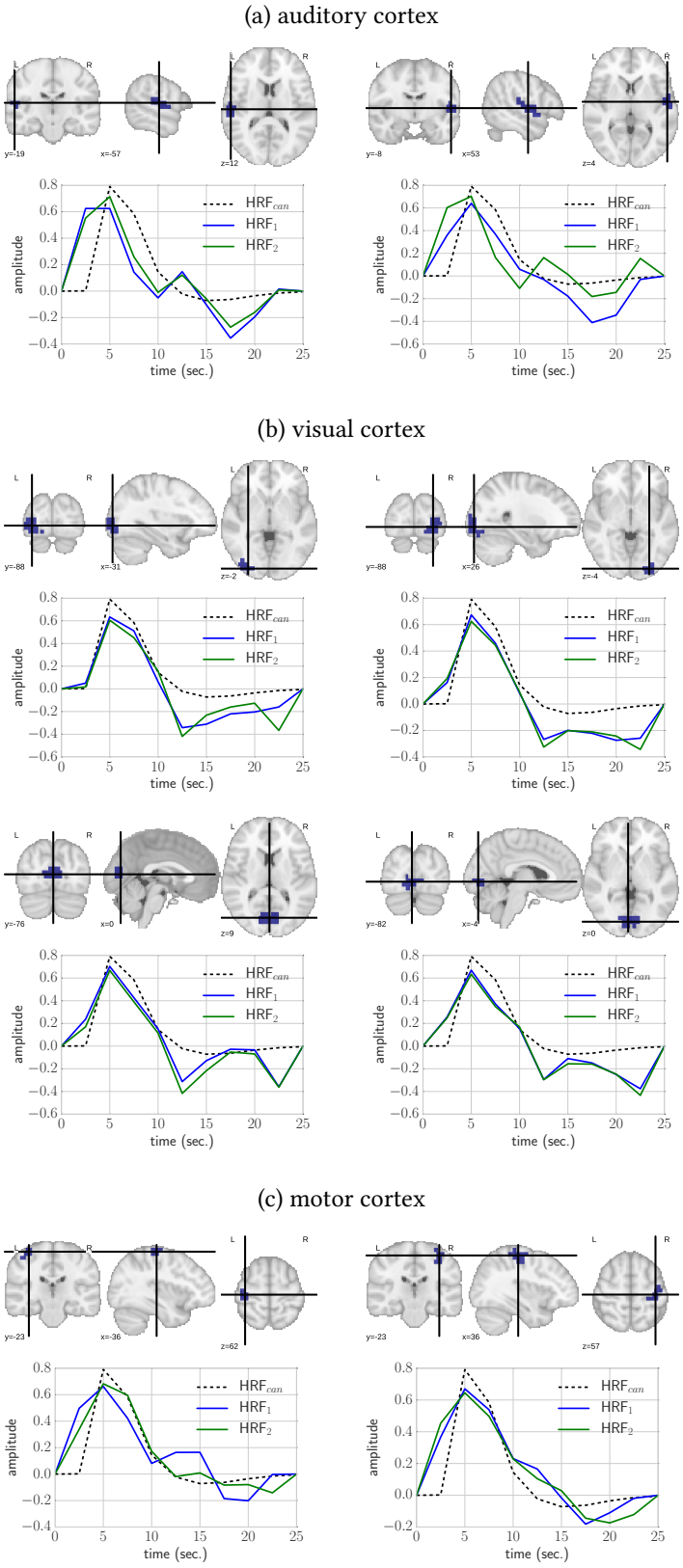



Figure 8.8: Subject level HRFs for one subject. We show 8 regions considering motor, auditory and visual cortex regions, on the left and right side. Four visual cortex regions are considered, including primary visual and high level visual regions. Two different sessions were analysed and plotted together in blue and green. HRFs are very similar in both sessions for a single region, confirming that the estimation is consistent. Left and right regions have similar HRFs too.

and SC in figure 8.9.

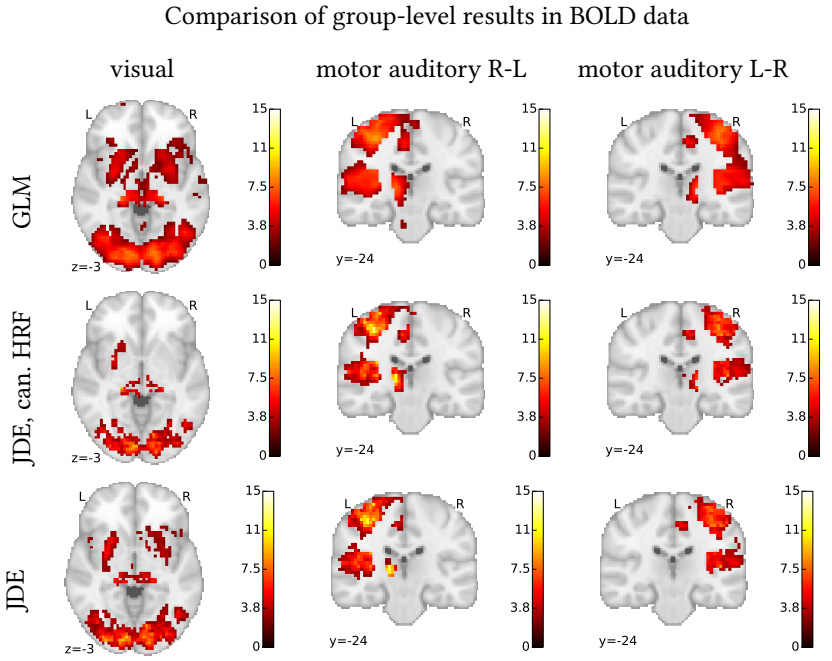



Figure 8.9: Comparison of group level z-maps of BOLD data computed using different methods. From top to bottom: GLM, JDE with the canonical HRF, and JDE estimating the HRF. Columns correspond to the contrasts shown: visual, motor-auditory right minus left, and motor-auditory left minus right. Note that FDR-corrected maps are one-sided.

These maps represent the group effect for each contrast. We can check the correlation of these maps with the subject level maps to be able to detect outliers[10]. Figure 8.10 shows the mean cross-correlation matrix across subjects on the left. We compute, for each subject, the cross-correlation matrix of the subject activation maps and the group activation maps, considering the different contrasts. Then we average all subject matrices to get the mean cross-correlation matrix across subjects (figure 8.10 left). Note here that the correlation was computed between subject-level z-maps of the contrast $\gamma^t \beta_j$ and group-level z-maps computed from $\gamma^t \beta_j$ for GLM, and subject-level contrast effect maps $\gamma^t a_j$ and group-level z-maps computed from $\gamma^t a_j$ for JDE. That is the reason why the scores are not comparable between GLM and JDE, but they are across subjects when using the same method. We just want to see a coherence in the group results for each method, and detect potential outliers: subjects or sessions that differ very much from the rest of the group. We also check that the correlation between contrasts (figure 8.10 left) is 0 when conditions are supposed to elicit activity in different regions, and equally correlated when we do left-right and right-left contrasts.

[10] There exist statistical approaches to mitigate the effect of outliers such as Wilcoxon signed-rank test with non-parametric estimation of the distribution under $H_0$.

We can define a similarity score per subject[11] as the correlation between subject and leave-one-out group maps, averaged across contrasts. It is computed as the mean of the diagonal entries in the cross-correlation matrix. For the correlation, the subject level being compared is not used to compute the group map. Figure 8.10 on the right shows the similarity scores per subject and session. There is one subject session that is a bit different from the rest. The subject level results of this subject session were checked and it shows activations in the expected regions.

[11] This analysis was first performed in the context of the INRIA Parietal retreat, in a joint work with E. Dohmatob and B. Thirion, for the analysis of OpenfMRI datasets.

Figure 8.11 shows the estimated HRFs for the 8 same regions that we considered in the subject level analysis: auditory cortex left and right, visual

cortex left and right for primary and high level visual, and motor cortex left and right. Left and right mean estimated HRFs for the auditory and motor cortex are quite similar. In the auditory cortex, the HRF peaks around 6 s and it has a deeper undershoot. In the motor cortex, the HRFs peak at 5 s and the undershoot is not very deep. The HRFs of the visual cortex peak around 6 s with a wide peak, and pronounced previous dip and posterior undershoot. Note that there is negative bump around second 22.5, that could be considered an artifact, but that it is consistent across subjects and similar visual regions. This could be due to the spectral characteristics of the paradigm: we might not be sampling some frequencies and that could affect the HRF results of the regions that are activated during a certain experimental condition, as it could be the visual.

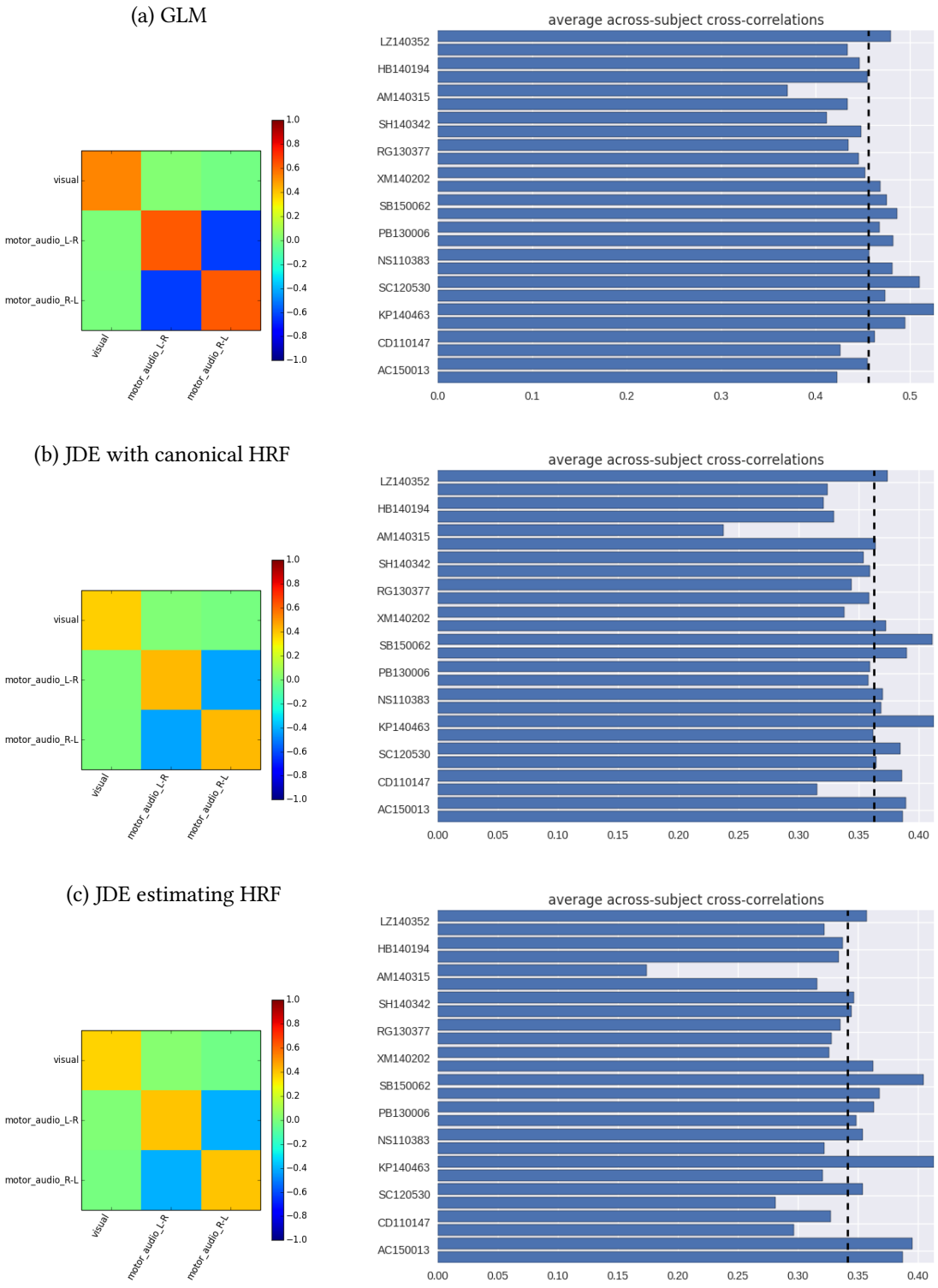
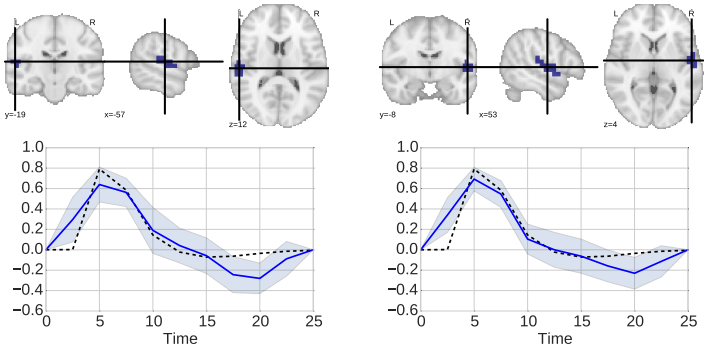
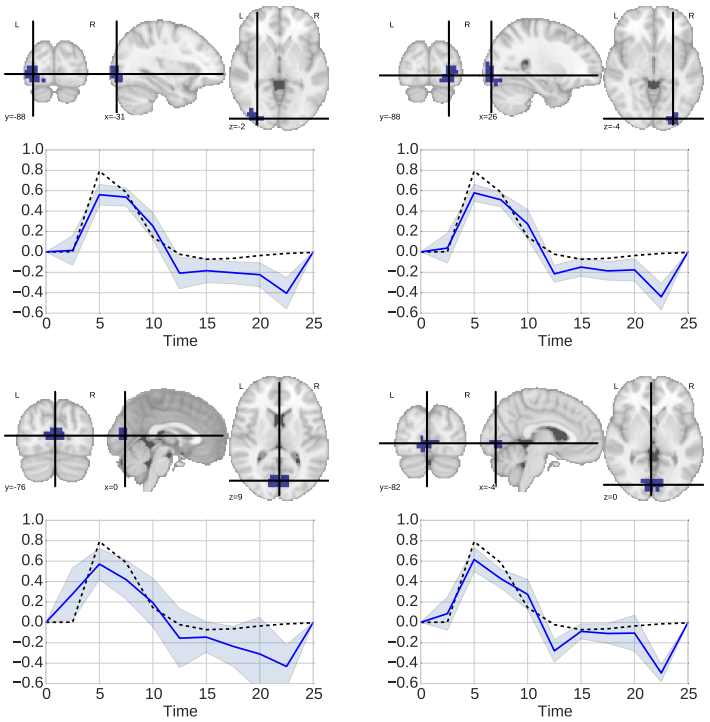



Figure 8.10: Similarity between subject and group maps. On the left side, we have the mean cross-correlation between the subject level maps and the group level maps. For each subject and session, we can define a score as the subject/leave-one-out-group map correlation, averaged across contrasts. On the right side of the figure, we depict the scores of each subject and session. The black dashed line shows the median of the scores of all subjects and sessions. Note that the subject that we used in the subject level analysis (SB150062) has a high score, so it is representative of the group in that sense.

(a) auditory cortex




(b) visual cortex




(c) motor cortex

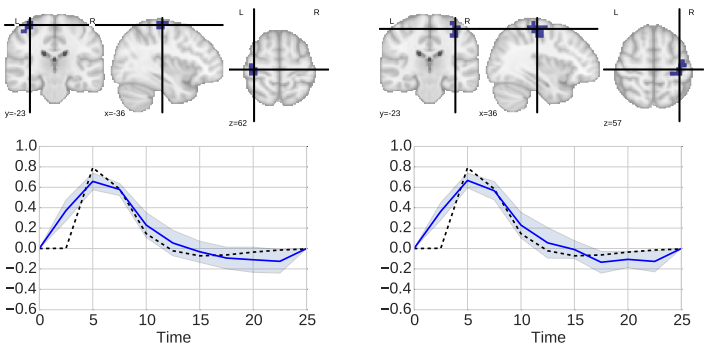



Figure 8.11: HRF estimated by JDE from BOLD data for some parcels of the auditory, visual and motor cortex regions, that we expect to be activated. The blue line is the mean over subjects and sessions and the blue shadow surrounding it the standard deviation over subjects and sessions. A dashed black line draws the canonical HRF sampled at the same rate as the estimated one.

## 8.3 Functional ASL data analysis

After the analysis of BOLD data, we now move to functional ASL data analysis. The design matrix has been constructed by considering the same conditions as in BOLD. For these conditions, we have hemodynamic and perfusion regressors. The perfusion regressor considers the control tag alternation. See figure 8.12.

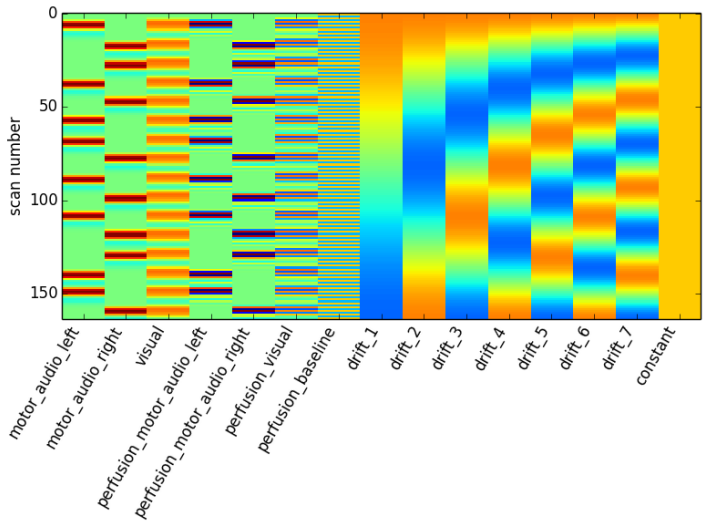



Figure 8.12: Design matrix used in the General Linear Model. The 3 task related hemodynamic regressors, 3 task related perfusion regressors, a control tag regressor to capture the perfusion baseline, 7 polynomial drifts and a constant regressor to capture the mean.

As in the case of BOLD data, ASL data is analysed with GLM, JDE using the canonical HRF and a PRF derived from the canonical HRF using the physiological link $\mathbf{\Omega}$, and JDE estimating both HRF and PRF, and injecting $\mathbf{\Omega}$ as prior knowledge on the PRF (see chapters 5 and 6). In this analysis we used the VEM solution using $\mathbf{\Omega}$ as a stochastic constraint prior. For ASL, we obtain maps corresponding to the hemodynamic component and maps corresponding to the perfusion component. We also analyse the baseline perfusion.

### 8.3.1 Single subject

Subject-level results for ASL data are illustrated on the same subject as the one used for BOLD imaging data.

Figure 8.13 shows the subject level z-maps of ASL data computed using GLM (on $\gamma^t \beta_j$ contrast maps). Significant activation is observed in the expected regions in the hemodynamic component in figure 8.13(a), and the significance does not change when we correct for multiple comparisons. The FDR corrected z-maps look cleaner though. Perfusion maps are generally visibly noisier than hemodynamic maps. In the perfusion component in figure 8.13(b), activations are not very significant before correction and they disappear after the correction for multiple comparisons. This is the reason why it is not displayed in the figure. This happens with all the sessions and subjects except for 5 out of 26 sessions, in which we can see some small active clusters. This might be due to the lower magnitude of this component, or to the smoothing of the effect during the preprocessings. One would need more data either on the subject level or on the group level

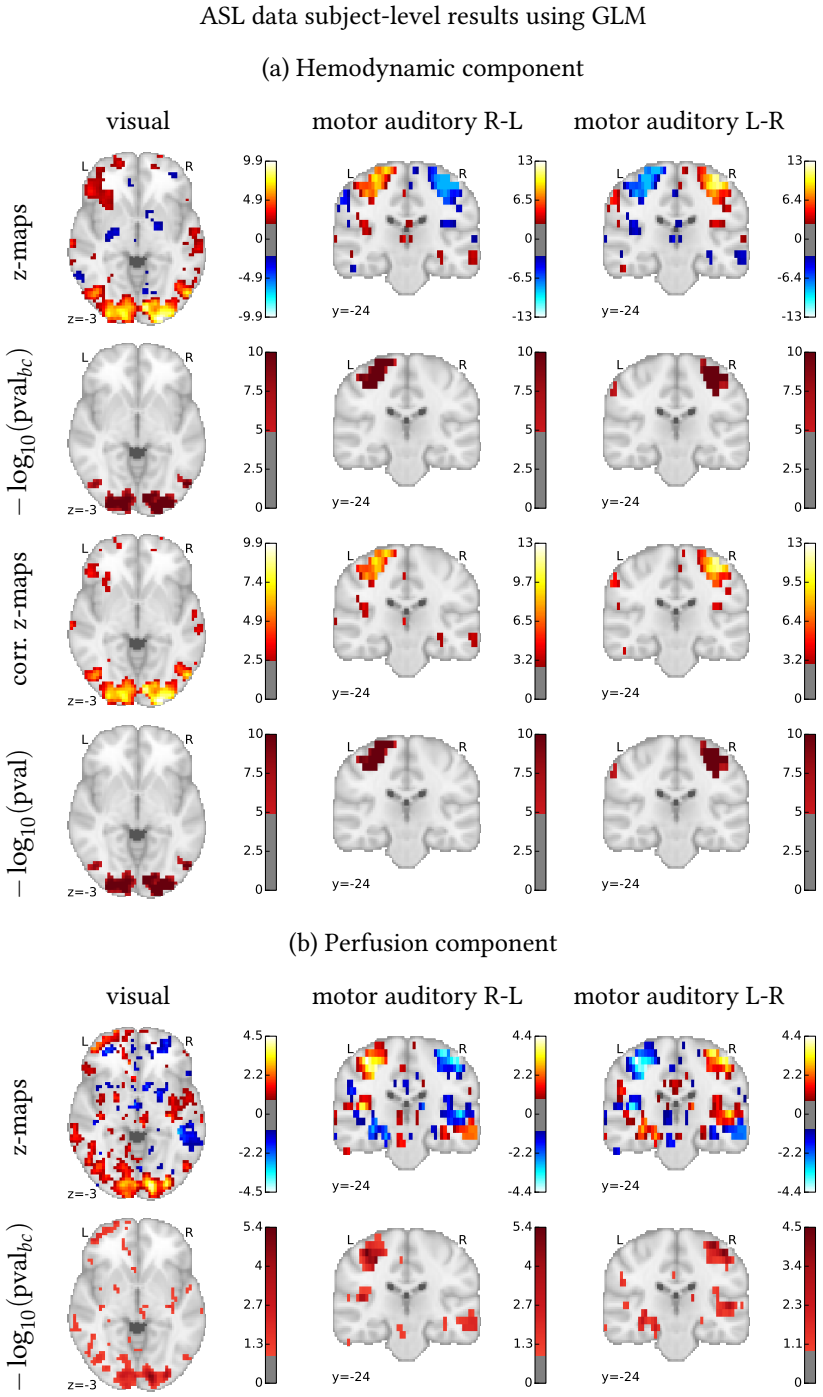



Figure 8.13: Subject level z-maps of ASL data computed using GLM for the (a) hemodynamic component and the (b) perfusion component, both from the ASL signal. Columns indicate different contrasts: visual, motor and auditory R-L, and motor and auditory L-R. For a given component, the first row contains the z-maps, the second row the negative log p-values of those z-maps, the third row depicts the z-maps for both components corrected for multiple comparisons using FDR to 0.05, and the fourth row the negative log p-values of the corrected z-maps. The p-value maps before correction are referred to as $\text{pval}_{bc}$. We observe that there is significant activation in the expected regions in the hemodynamic component and that the significance does not change when we correct for multiple comparisons. In the perfusion component, activations are not very significant before correction and they disappear after the correction for multiple comparisons. For this reason, these maps are not shown.

to gain in statistical power. Note that the scale of the negative log p-values shown has been modified for the perfusion component to show p-values from 0.1 to $10^{-5}$.

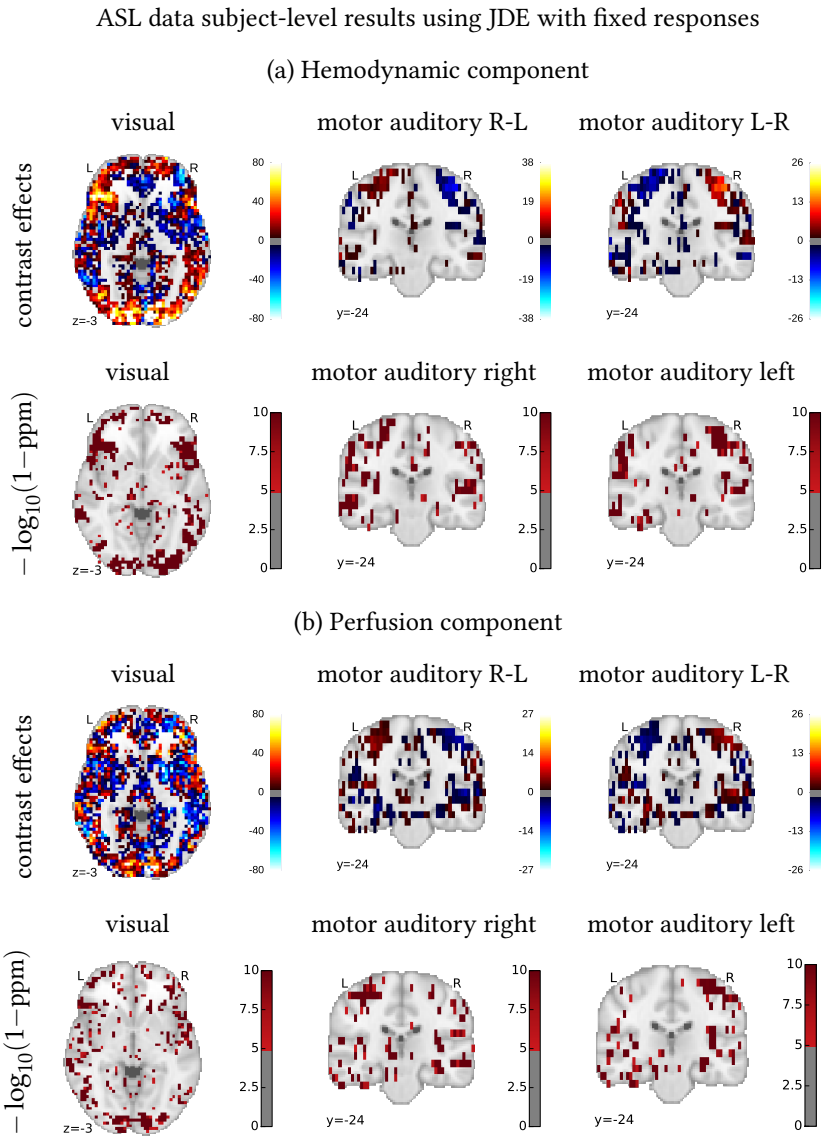



Figure 8.14: Subject level maps of ASL data computed using JDE with fixed HRF and PRF shapes for the (a) hemodynamic component and the (b) perfusion component, both from the ASL signal. Columns correspond to contrasts or experimental conditions. For each component, first and second rows display effect maps and PPM maps $-\log_{10}(1-\text{ppm})$ using a certain threshold. We observe task-related activations in the occipital cortex for the visual contrast, in the motor cortex for the motor auditory condition, and some also in the auditory cortex. In the visual contrast, we have some unexpected activations in the frontal lobe in the hemodynamic component that disappear almost completely in the perfusion component. PPM maps show that activity in visual, motor and auditive cortices is significant.

When analysing the data with JDE, as before, we use effect contrast maps and PPM. For this analysis, non-smoothed data and an MRF model for spatial regularization were used. The threshold $\delta$ chosen to create the PPMs is the same one as in BOLD for the hemodynamic component. For the perfusion component, we used the 5% of the parcel perfusion baseline mean. Figure 8.14 shows results for JDE using the canonical HRF and the PRF computed from the canonical HRF and the physiological link $\boldsymbol{g} = \boldsymbol{\Omega h}$ as response functions for the hemodynamic and perfusion components. Figure 8.14(a) illustrates the effect contrast maps $\boldsymbol{\gamma}^t \boldsymbol{a}_j$ and the PPMs of the hemodynamic component. Activations are much less defined than in BOLD data, but it is visible that the main detected regions correspond to those also detected by the BOLD analysis. We get activations in the occipital cortex for the visual contrast, and in the motor cortex for the motor auditory

condition. We also observe some activity in the right auditive cortex. In the visual contrast, we have some unexpected activations in the frontal lobe in the hemodynamic component that disappear almost completely in the perfusion component (see figure 8.14(b)). PPM maps show that activity in visual, motor and auditive cortices is significant. The respective perfusion maps in figure 8.14(b) display noisier effect contrast maps for the perfusion component $\gamma^t c_j$ as compared to the hemodynamic one. However, the PPM maps show that activations are significant in the expected regions too. Results are noisier than in BOLD, even for the hemodynamic component, as expected.

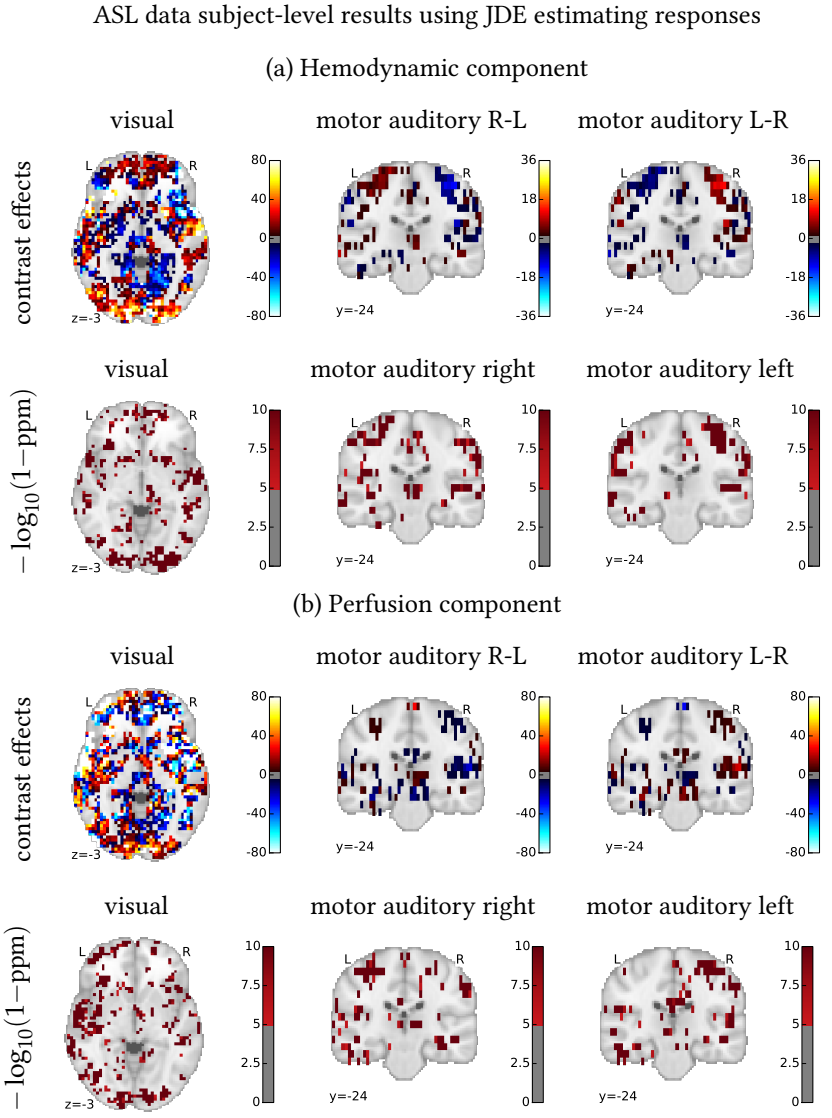



Figure 8.15: Subject level maps of ASL data computed using JDE and estimating an HRF per region for the (a) hemodynamic component and the (b) perfusion component, both from the ASL signal. Columns correspond to contrasts or experimental conditions. For each component, first and second rows display effect maps and PPM maps $-\log_{10}(1-\text{ppm})$ using a certain threshold. Similarly to JDE with fixed canonical HRF and to GLM, we get activations in the visual and motor cortices. However, results are quite noisy, as in the case of JDE with canonical HRF.

Compared to JDE with fixed HRF and PRF, estimating HRF and PRF responses within the JDE framework yields to similar results. Figure 8.15 shows the effect contrast maps and PPMs for hemodynamic $\gamma^t a_j$ (see figure 8.15(a)) and perfusion $\gamma^t c_j$ (see figure 8.15(b)) components for JDE estimating the responses. The biggest difference between the maps of both hemodynamic and perfusion components is the removal of some "parasite" activations in the frontal lobe. PPM results on JDE estimating or using a

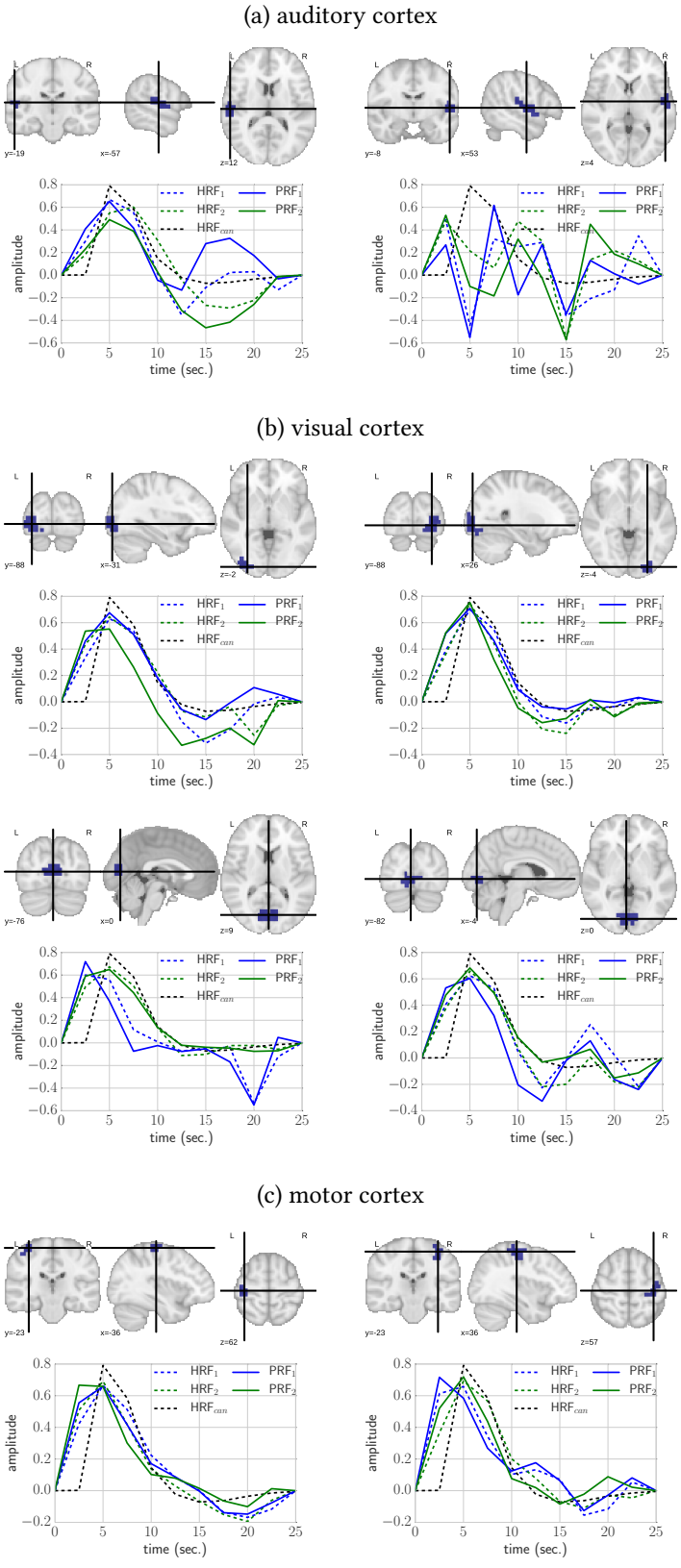



Figure 8.16: Subject level HRFs computed from a subject in 8 regions that we expect to be activated from the auditory, visual and motor cortices. Here HRFs from different sessions are in blue and green as in BOLD results, and we differentiate HRF and PRF with dashed or continuous lines, respectively. The estimated HRF and PRF are quite similar in a single region. Auditory cortex HRFs are not well estimated.

fixed HRF/PRF give similar significance patterns.

In figure 8.16, it is worth noting that the difference between PRF and HRF is very small and therefore the estimation of the two response functions might not be necessary for this data. This might not be the case in an event related design experiment. Second, the HRF responses peak before the canonical in the motor cortex, and close to the canonical HRF peak (a bit before) in the visual cortex. This is consistent with BOLD results shown in figure 8.16. PRF responses peak slightly before in most of the cases. The responses were not well estimated in auditory regions, especially in the right auditory cortex. In this dataset, a weak effect has been reported in the auditory cortices. This could be due to the fact that the experiment is optimized to get motor and visual responses. The auditory stimulus is monotone and it acts as a cue to trigger motor action, complementary to the visual indicator. Subjects may have paid less attention to these stimuli.

### 8.3.2 Group level statistics

From the subject level effect maps that we get from GLM ($\boldsymbol{\beta}_j$), JDE with fixed response shapes and JDE estimating responses ($\boldsymbol{a}_j, \boldsymbol{c}_j$), we performed a group level analysis as we did for BOLD data. FDR-correction for multiple comparisons with a threshold of 0.05 was applied.

Figure 8.17 shows a comparison across methods of the group level maps for hemodynamic (a) and perfusion (b) components. We observe that JDE results give a more sharper map than GLM results. This also amounts to a gain in specificity: active regions are less spread in JDE with respect to GLM. The GLM visual contrast z-maps contain unexpected activations in the temporal cortex. However, GLM captures auditory cortex too (with a low z-score) in the hemodynamic component, whereas in JDE most activations in auditory cortex are lost when we do the right-left contrast. We might think that this is due to the smoothing in GLM helping to compensate inter-subject variability, but JDE results on smoothed data do not capture activation on the left side either. The estimation of the HRF in JDE does not make much difference for the hemodynamic component. The perfusion component contrast z-maps show much more localized activations no matter which method is used. In these cases, no auditory cortex activations are observed. For the case of the visual perfusion contrast z-map, JDE finds much smaller activated regions. JDE results might be showing a more localized activated region, that is spread by the smoothing in GLM, although it could be also an effect of inter-subject variability. JDE results on smoothed data confirm the detection of bigger clusters than when using non-smoothed data. In this sense, the use of smoothing does not seem to be very appropriate, since one of the interests of fASL is the observation of more localized activations [Luh et al., 2000, Tjandra et al., 2005].

The estimation of the perfusion response function seems to increase the activated region in the visual contrast, but it decreases the activated motor cortex region in the motor auditory contrast. Usually the responses are more easily estimated in the visual cortex than in other regions because of the strength of the signal in those regions. That might be the reason why we get larger active clusters when we estimate the responses in the visual

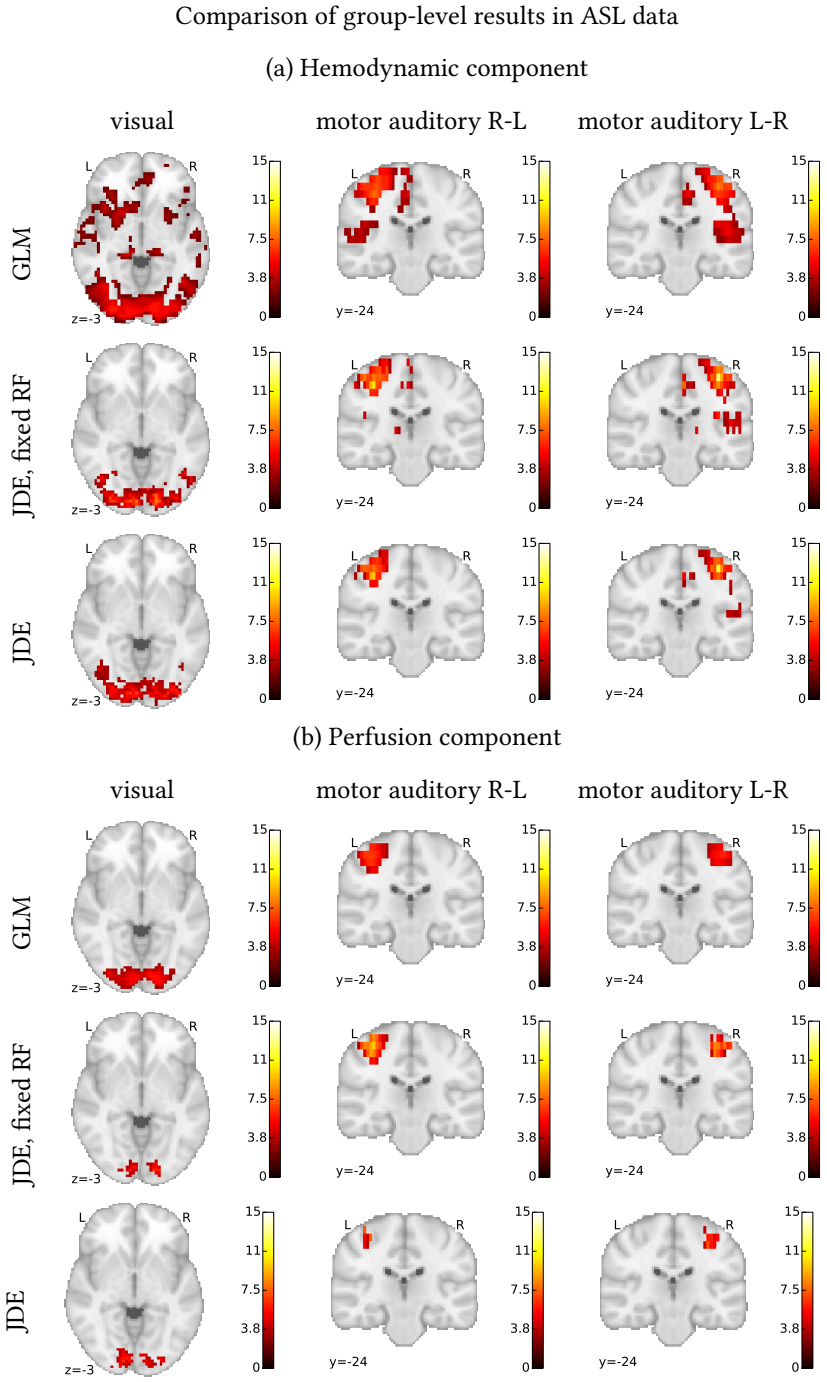
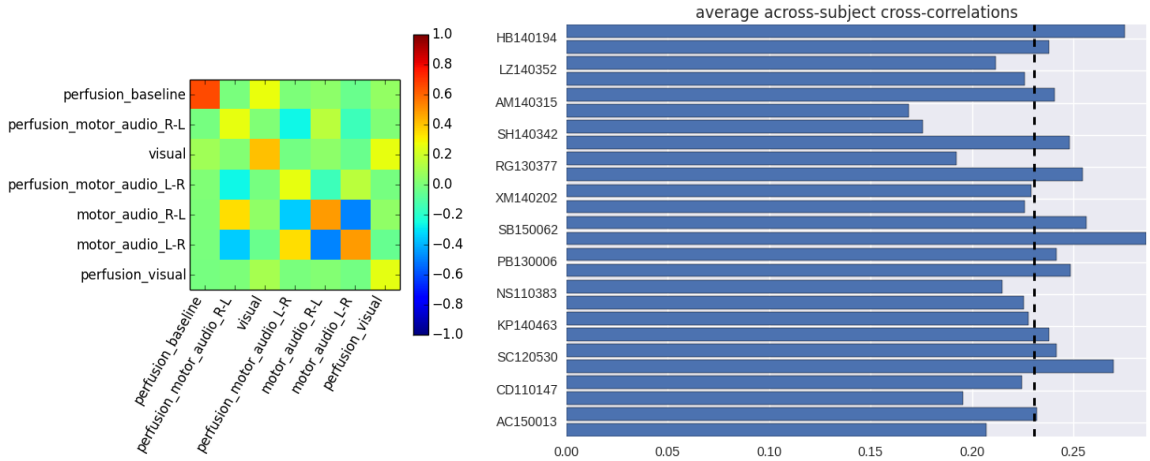
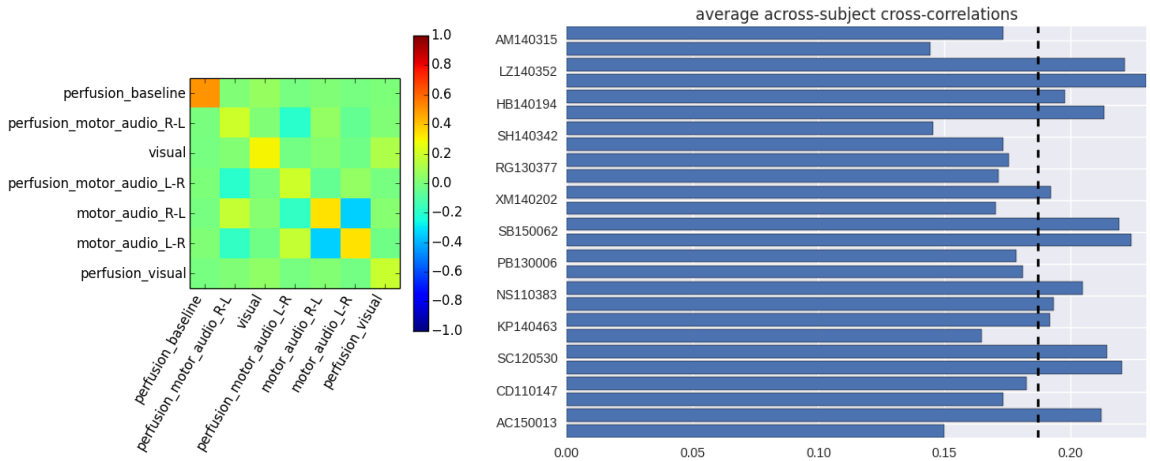



Figure 8.17: Comparison of group level maps of ASL data computed using GLM, JDE with fixed HRF/PRF, and JDE estimating the HRF and PRF. Columns correspond to different contrasts. Rows correspond to (a) the hemodynamic component of ASL and (b) the perfusion component of ASL. Note that active regions are bigger in GLM: this is a direct consequence of the smoothing applied to the used data. In JDE, the activated region is sharper and more localized. In ASL, in contrast to BOLD, estimating HRF and PRF amounts to losing sensitivity to activation. A possible explanation is that the noise level of the data and the block design nature of the experiment make this estimation difficult.

(a) GLM




(b) JDE with canonical HRF




(c) JDE estimating HRF/PRF

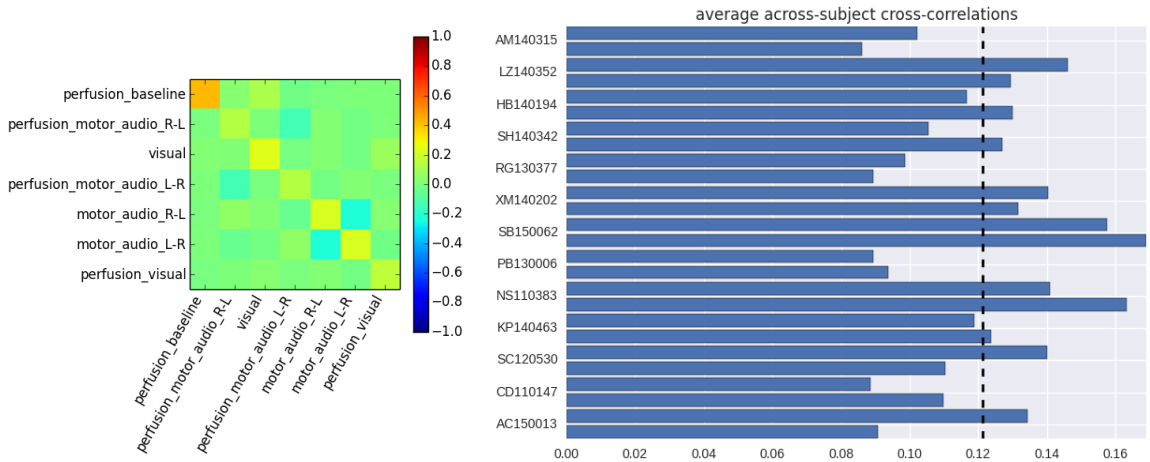



Figure 8.18: Similarity between subject and group maps in ASL. On the left side, we have the mean cross-correlation between the subject level maps and the group level maps. Note a non-zero correlation between hemodynamic and perfusion maps of the same contrast, and also between baseline perfusion and perfusion visual contrast. On the right side, we have the mean correlation between subject and leave-one-out group level maps per subject and session. The black dashed line shows the median of the correlations of all subjects. Note that the subject that we used in the subject level analysis (SB150062) has a high score, so it is representative.

cortex than in other regions.

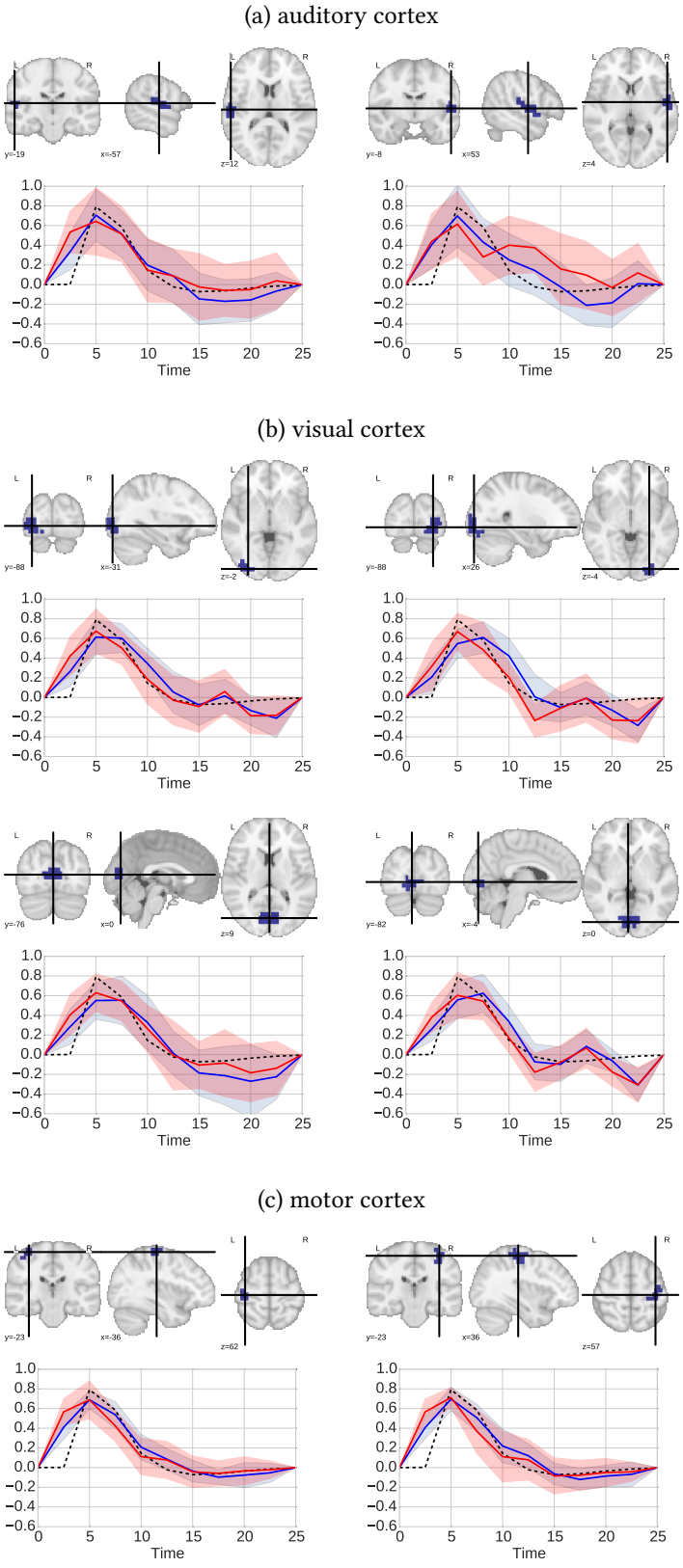



Figure 8.19: **HRF** (in blue) and **PRF** (in red) estimated by JDE from ASL data for some parcels of the auditory, visual and motor cortex regions, that we expect to be activated. The blue and red lines are the mean HRF and mean PRF over subjects and sessions and the blue and red shadows surrounding them are the standard deviation over subjects and sessions. A dashed black line draws the canonical HRF.

Figure 8.18 shows the similarity check of the subject level maps with respect to the leave-one-out group level maps. Recall that these graphs

were computed between subject level z-maps and leave-one-out group level z-maps for GLM, and subject level effect maps and leave-one-out group level z-maps for JDE. The different nature of subject level maps and group level maps leads to a lower correlation in JDE. One can observe that there are no outliers. However, there is a larger variability across subjects when we estimate the responses. This is also reflected in the activation maps, which show more restricted active regions. Note that, as one can expect, there is a non zero correlation between hemodynamic and perfusion components, when we compare a certain contrast in figure 8.18 on the left. We also observe a non-zero correlation between baseline perfusion and visual perfusion.

The HRF and PRF estimated responses for the 8 regions already shown for other analyses are depicted in figure 8.19. The mean PRF peaks slightly before than the mean HRF in all cases. Consistently with the BOLD data results, the motor cortex HRF responses peak before than the canonical and have a wider peak. The visual cortex HRF responses peak around the canonical or a bit later, as in BOLD. The auditory cortex responses peak around 5 s and there is a higher peak difference between PRF and HRF. It is worth mentioning that left and right regions have similar shapes, and that all visual regions have similar results. As reported in single-subject results, responses for the right auditory cortex have shapes that we would not consider physiologically plausible.

### 8.3.3 The impact of the smoothing

In the previous sections, a comparison of JDE with non-smooth data and GLM with smooth data has been done. The reason why is that the smoothing is intrinsic in the JDE model through a Markov Random Field on a hidden variable $\boldsymbol{q}$. However, this comparison might be not fair to one of the methods, if the amount of smoothing in both methods is not the same. For this reason, in this section we compare both methods on smooth and non-smooth data. Figure 8.20 shows, for both components, the effect of the smoothing for the methods GLM and JDE with fixed responses. In the hemodynamic component, we observe more spread active regions when using smooth data for both methods. Using smooth and non-smooth data, activity is more significant when using JDE. Note also more "parasite" activations in GLM results. In the perfusion component, we also observe more significant activity when using JDE in both cases (on smooth and non-smooth data). Moreover, when using GLM we lose most activations. As in the hemodynamic component, activation patterns are more spread when using smooth data.

With this comparison, we want to show how the smoothing in GLM is justified. In JDE, the use of both smooth and non-smooth data deliver significant activation patterns. The choice of using non-smooth data in this case is to keep the higher localization of the perfusion component, which is interesting. Moreover, the intrinsic smoothing in JDE seems to be comparable to the previous smoothing in GLM. Future works could be dedicated to study this smoothing effect in more detail.

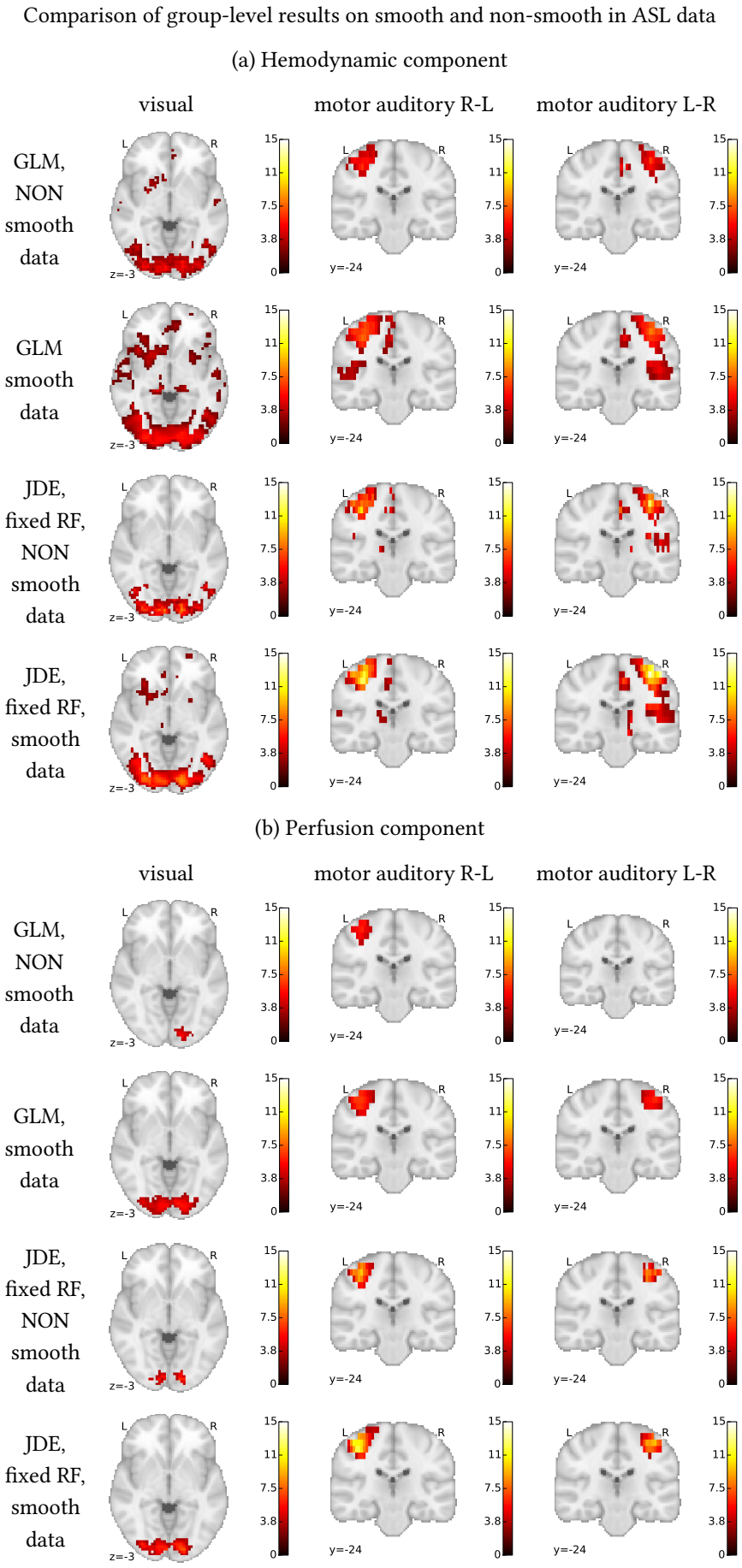



Figure 8.20: Comparison of group level maps of ASL data computed using GLM and JDE with fixed HRF/PRF, on previously smooth and on non-smooth data. The comparison is done for the (a) hemodynamic and (b) perfusion components of ASL. Rows in each component correspond to GLM on non-smooth data, GLM on smooth data, JDE with fixed HRF/PRF on non-smooth data, and JDE with fixed HRF/PRF on smooth data. Columns correspond to different contrasts: visual, motor auditory right-left, and motor auditory left-right. Note that active regions are bigger when using smoothing, but that activation is more significant in both cases when using JDE with fixed HRF/PRF. In the perfusion component, using non-smooth data on GLM most of the significant activaty disappears. When using JDE with fixed responses, we have higher significant activity when using smooth and non-smooth data, but the activity is more localized when using non-smooth data.

## 8.4 Baseline perfusion in basal and functional ASL

The most important part of the ASL signal, and the one that provides a quantitative measure of cerebral blood flow is the perfusion baseline component $\Delta M = M_{control} - M_{tag}$. From the baseline perfusion, we obtain the cerebral blood flow (CBF) maps with the expression (2.2) in section 2.4.2. Recall that CBF quantification is done with this expression [Alsop et al., 2015]:

$$CBF[mL/100g/\text{min}] = \frac{6000\lambda\Delta M}{2\alpha TI_1 M_0}\exp\left(\frac{TI_2}{T_{1,\text{blood}}}\right) \tag{8.3}$$

$$= \frac{6000\lambda\Delta M}{M_0}\xi, \tag{8.4}$$

$$\text{where } \xi = \frac{1}{2\alpha TI_1}\exp\left(\frac{TI_2}{T_{1,blood}}\right) \tag{8.5}$$

Here $\Delta M = M_{control} - M_{tag}$ is the averaged control-tag magnetization; $M_0$ is the relaxed magnetization; $\lambda = 0.9$ mL/g is the averaged brain/blood partition coefficient; $\alpha = 0.98$ is the labeling efficiency; $T_{1,blood} = 1650$ ms is the longitudinal relaxation time at 3T. See section 2.4.2 for further details. Time difference between slices [Nöth et al., 2006] needs to be considered and it is done before any other preprocessing. The division by the relaxed magnetization $M_0$ corrects for signal variations during the acquisition. For low TRs, a correction by $1/(1-\exp(-TR/T_{1,tissue}))$ needs to be applied, $T_{1,tissue}$ being the assumed $T_1$ of gray matter. Figure 8.21 shows the basal CBF map and the functional session CBF map for one subject session.

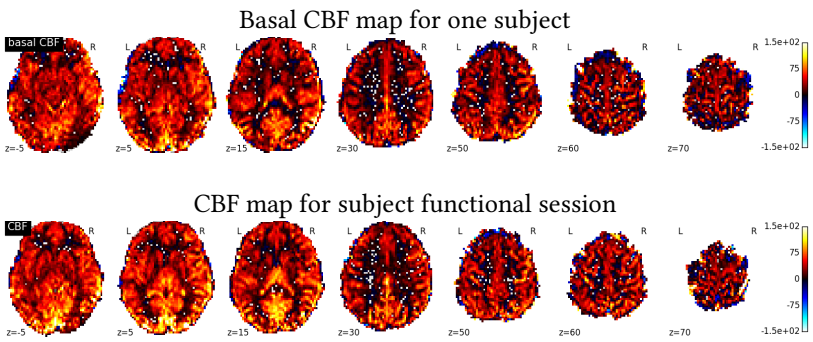



Figure 8.21: Basal CBF and CBF of a functional ASL session.

In the analysis of HEROES functional ASL data, scale correction during the pre-processing step corrects for the time difference between the slice acquisition: $\xi = \frac{1}{2\alpha TI_1}\exp\left(\frac{TI_2}{T_{1,blood}}\right)$. The output of our analyses corresponds to the difference $\Delta M$ multiplied by this scale factor $\xi$. Figure 8.22 shows the perfusion maps estimated with GLM, JDE with fixed HRF/PRF responses, and JDE estimating HRF/PRF. We observe already that they are quite different between GLM and JDE, because of the smoothing of the data. GLM perfusion map has mean 6.99 and standard deviation 3.84. JDE perfusion map has mean 7.58 and standard deviation 4.96 when using canonical HRF, and mean 7.56 and standard deviation 5.15 when estimating the HRF. For GLM the mean but mainly the standard deviation are low because of the smoothing effect. Note that the standard deviation is a bit larger when we estimate the HRF in JDE, as we observed in the activation

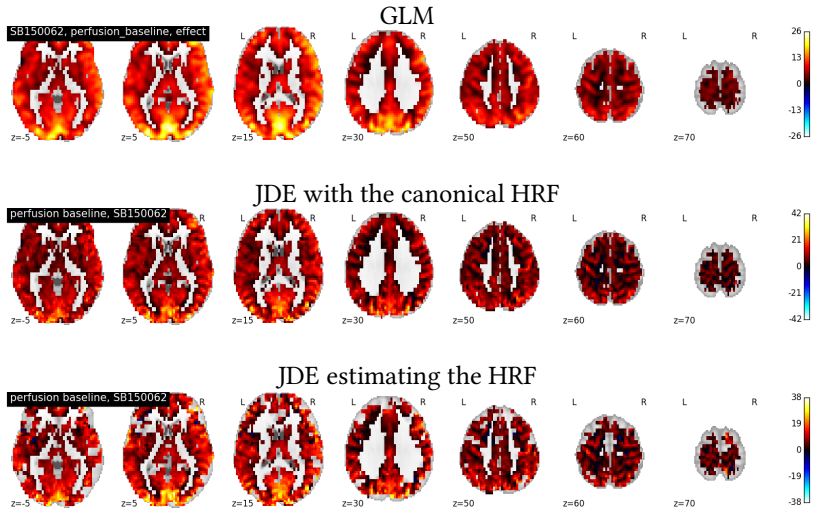



Figure 8.22: Perfusion baseline of a subject for GLM, JDE with canonical HRF, and JDE estimating the HRF. Note in JDE some regions with 0 value that come from numerical issues during the computation.

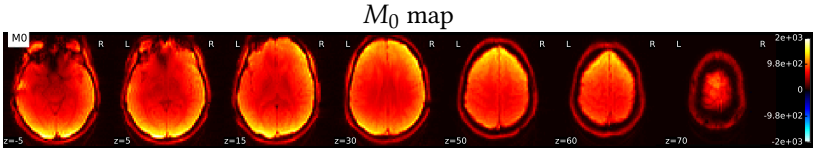



Figure 8.23: $M_0$ magnetization map for the functional ASL session of a subject.

maps in figure 8.18.

After the analysis, we can apply the remaining correction to get CBF maps from the perfusion: $\frac{6000\lambda}{M_0}\xi\Delta M$. To get proper results, this has to be done before applying a spatial normalization. We need the $M_0$ map, that is depicted in figure 8.23.

### 8.4.1 Group analysis of cerebral blood flow

In this section, we compare the CBF maps from 13 subjects and the mean and standard deviation of the CBF maps across subjects. In figure 8.24, we observe the gray matter CBF values for all subjects in subject space (before normalization). [Alsop et al., 2015] states that, as a general rule, gray matter CBF values from $40 - 100$ mL/100g/min can be normal. All subjects have mean values from 30 to 60 mL/100g/min. Some are below the threshold of what [Alsop et al., 2015] considers normal values.

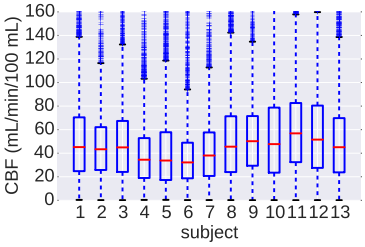



Figure 8.24: Boxplot showing the basal gray matter CBF for all subjects, considering data in subject space.

Figure 8.25 shows the cerebral blood flow (CBF) for the basal CBF acquisition (see figure 8.25(a)), and the CBF extracted from the functional ASL sessions (see figure 8.25(b)). We show the mean and the standard deviation over subjects and session, in MNI space.

The subject CBF map mean and standard deviation across subjects is 38.97 and 24.0 respectively for the basal CBF, and 43.68 and 22.72 respectively for the functional ASL sessions, when we just consider the gray matter regions and after normalization. It is worth mentioning that a higher mean but a lower standard deviation are found in the functional ASL session CBF. Note that the region with higher standard deviation lies in the occipital region: higher values are observed in the occipital lobe in the first slice. This

is due partly to the fact that it is close to the limit of the acquired volume, and some subjects have this region cut. One subject has been discarded to do the mean for this reason. Note that this is also consistent with the non-zero correlation between visual and baseline perfusion in figure 8.18.

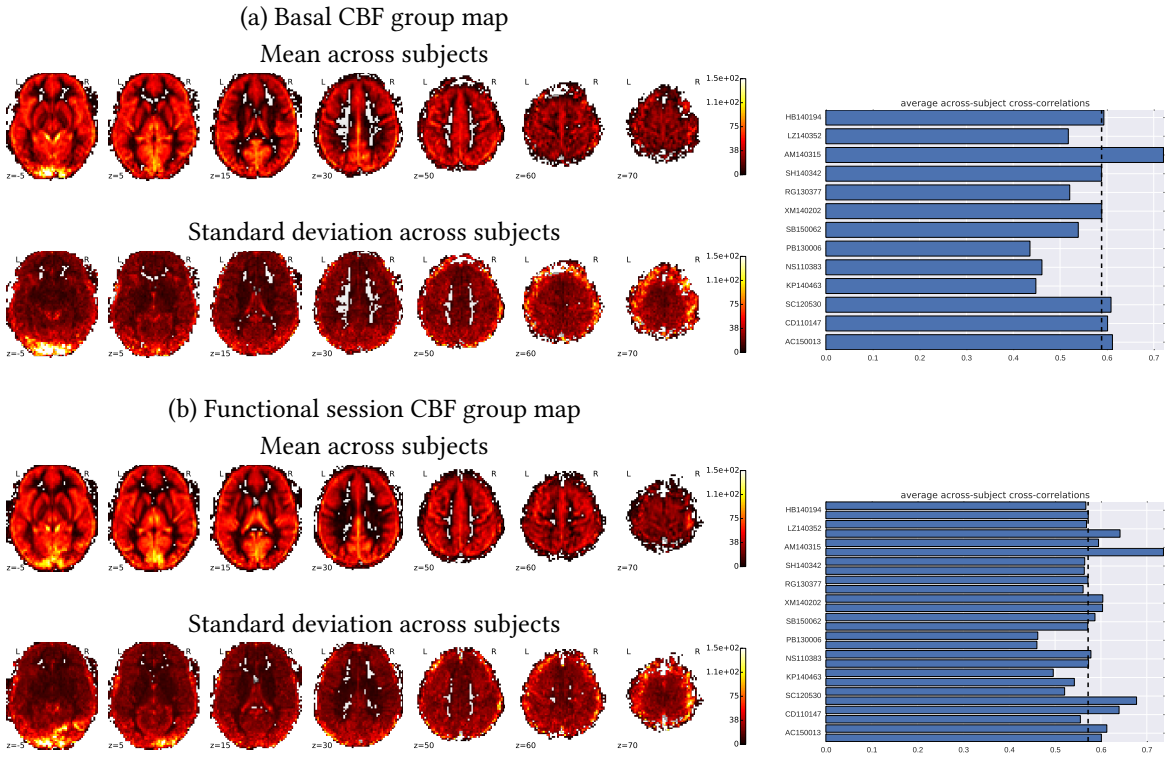



Figure 8.25: Cerebral blood flow mean and standard deviation of a group of 13 subjects. (a) Basal CBF, acquired at rest. (b) Baseline CBF extracted from functional ASL session (2 fASL sessions/subject). On the right side, the correlation of the subject-level and group-level CBF maps.

## 8.5 Discussion of the results

In this chapter, we have analysed a dataset with BOLD fMRI, ASL fMRI, and basal ASL data corresponding to 13 subjects. The experimental paradigm used for fMRI is a mini-block design and consists of visual, auditory and motor tasks. We have analysed them both with GLM, JDE with fixed responses and JDE estimating HRF and PRF responses. All three methods give results which illustrate activity in expected regions. CBF maps give reasonable values at the subject and group levels, according to the literature [Pimentel et al., 2013, Raoult et al., 2011]. Note here that JDE has just been compared to the simplest GLM, but that some flexibility can be given to GLM, for example through the addition of other regressors to capture delay and dispersion.

Results on BOLD and ASL fMRI data show significant activation, at the subject and group levels, in motor, visual and auditory cortices. Auditory regions do not show task-specific CBF variations in the perfusion component of fASL. In the hemodynamic component of the group level GLM results, we find some activation in the auditory cortex, although it

is much lower than in other involved regions. The lower activity in the auditory cortex, also seen in BOLD data, might be due to the fact that the experiment has been optimized for visual and motor tasks, and the auditory task is just a complementary task acts as a cue to trigger motor action, complementary to the visual indicator. Subject level GLM results do not show significant activation in the perfusion component after FDR-correction for multiple comparisons. However, when we use JDE, PPM maps show significant activation. To compare GLM and JDE results, we compared (1−p-values) of GLM with PPM maps of JDE. These maps would be equivalent if non-informative priors were used in JDE. However, this is not the case and therefore they do not represent exactly the same quantity but they can be roughly compared.

When comparing the methods used for the analysis, the main difference is that GLM needs smoothed data and JDE does not (although smoothed data can also be used). As we have seen, a multivariate method as JDE is more sensitive to voxel activity when we compare it to a univariate method as GLM in most of the cases. However, the use of smoothed data in GLM allows to decrease spatial inter-subject variability and noise effects. Especially in the analysis of fASL, using smoothed data can cause a loss of specificity and localization of activations. Moreover, the smoothing incurs a loss of detail in regional changes of perfusion baseline, and therefore of regional CBF, when estimating it. This makes the correspondence between fASL and basal CBF measured at rest more difficult, and it decreases the effects that make the fASL signal interesting: it incurs a loss in localization and it introduces an error in the quantification of task-related CBF variation.

When we compare results of BOLD and ASL fMRI, that we have summarized in figure 8.26, it is worth noting that BOLD data shows more active regions than the hemodynamic component of the fASL acquisition. This is coherent with the existing literature [Pimentel et al., 2013] and it is due to the fact that the TE used in the ASL protocol (TE $=$ 11 ms) is not optimal for the BOLD contrast, which is maximized for TE $=$ 30 ms at 3T [Ogawa et al., 1993, Tjandra et al., 2005]. The perfusion component in fASL also yields more localized activations than BOLD, as already reported in the literature for the motor cortex [Pimentel et al., 2013, Raoult et al., 2011]. This happens when using both methods in the analysis: GLM and JDE with fixed HRF/PRF responses. It is worth noting that in JDE we obtain more localized and stronger effects than when using GLM, in general.

As regards HRF estimation, we observe that for the case of BOLD data activation results improve when estimating the HRF, consistently with the literature [Handwerker et al., 2004]. However, in ASL data we detect less activity when estimating the HRF and PRF. This might be due to the fact that the signal is noisy and this block design setting does not make the response estimation an easy task. In an event-related design setting, in which estimation of the responses is usually easier and in which statistical power is lower, the estimation of HRF and PRF responses could be key in the detection of task-related activity.

In this chapter, we analysed BOLD and ASL fMRI data with different methods and observed different activation patterns depending on the analysis made. Further works should assess the impact of the spatial

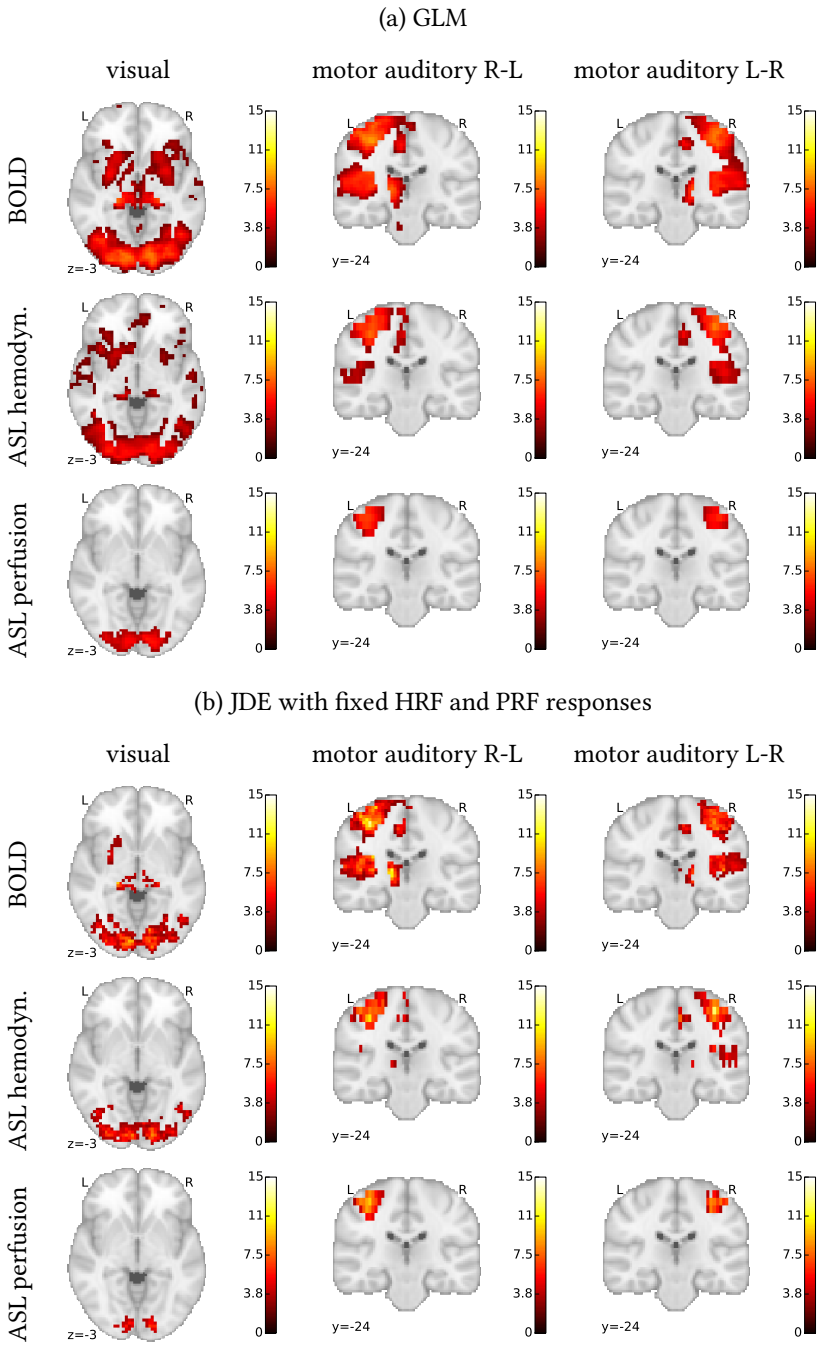



Figure 8.26: Comparison of group level maps of BOLD, and the hemodynamic and perfusion components of ASL data computed using (a) GLM and (b) JDE with fixed responses. Columns correspond to different contrasts. In this summary figure, we observe the more localized activation of the perfusion ASL component with respect to the hemodynamic ASL component and to BOLD in both methods. In general, JDE finds more localized activated regions than GLM.

modelling and of the use of multivariate models instead of univariate models. Results in this chapter lead to the conclusion that a multivariate approach without the need for data smoothing could be best suited for the analysis of fASL, due to the nature of this data.

# 9 Conclusion

Throughout this thesis we aim at improving the understanding of BOLD and ASL fMRI signals, and at providing methods for the neuroscientific community to study brain function and the neurovascular coupling.

In order to better understand BOLD fMRI data, we relied on Bayesian models in form of the joint detection estimation (JDE) approach and its temporal modelling of hemodynamics. We showed that the modelling of the existing spatial correlation in the BOLD signal through a Markov random field yields an improvement of the sensitivity to detect regional activation with respect to the classical general linear model. The modelling of temporal hemodynamics in this context yields an even higher sensitivity for detecting evoked activity in BOLD, which is intuitive: we are capturing the spatial variability of the hemodynamic response function, making it possible to calibrate activation estimation to the exact moment where the effect occurs. Moreover, the estimation of the so-called hemodynamic response function is interesting in itself since it reflects differences between regions and subjects. In the context of this thesis, a multi-session model allowing the weighting of the sessions according to the noise level was proposed to increase the statistical power of the model.

In order to better understand ASL fMRI data and considering the similarities between BOLD and ASL signals, we relied again on Bayesian models to estimate the parameters of interest. We showed that ASL data is noisier and more difficult to model than BOLD. The perfusion measure that provides this data is the main reason for using ASL in functional MRI. Perfusion gives a more localized measure, it can be quantitative, and it is directly related to brain activity. However, this component of the signal is very small and its estimation relies on the subtraction of two images acquired at different moments (control and tag). A joint detection estimation approach had already been proposed to temporally model perfusion and hemodynamic responses, relying on Markov Chain Monte Carlo (MCMC) optimization strategy for the parameter estimation. In the context of this thesis, we focused on an improved modelling of the perfusion response function, the estimation of which has been shown to be difficult [Vincent et al., 2013a].

To this aim we investigated physiological models and derived a linear operator from the extended Balloon model that relates perfusion and hemodynamic responses. In the JDE Bayesian setting, we injected this relationship as *a priori* information to inform the perfusion response from the hemodynamic response function.

We also proposed a faster optimization strategy for the parameter estimation in functional ASL JDE analysis: a variational expectation maximization solution. VEM has been shown to give similar results to MCMC with much lower computational times. VEM and MCMC solutions were developed in python in the package PyHRF[1].

[1] pyhrf.org

We investigated the variants of the Balloon model and the parameters involved, and their impact when used as prior knowledge in the context of JDE. It has been shown that, although the estimation when using MCMC converges using different sets of parameters, the convergence is faster depending on the parameters used. For this reason an assumption was made: the parameters that make the convergence faster are the closest to the real ones. This is interesting in itself: estimating the physiological parameters is a field of research. We observed that the set of physiological parameters used has more impact than the variants of the Balloon and hemodynamic models.

Finally, we validated the performance of these methods on a dataset acquired at CEA/Neurospin during this thesis. Classical methods were compared to JDE VEM solution with BOLD and fASL data, and subject and group level results were put into perspective. It was shown that JDE can provide a better sensitivity to activation and a more localized activation detection. We observed that the estimation of temporal responses does improve detection results in BOLD but not in fASL in this setting. However, the estimation of hemodynamics and perfusion responses can be interesting independently of the activation detection. The quantification of the cerebral blood flow was also performed on this data set. With JDE and GLM, we extracted the perfusion baseline of the functional ASL signal. Because of the smoothing in GLM, the estimated perfusion baseline is smoothed and it has a lower correspondence with the basal CBF measured at rest. Although JDE provides some advantages with respect to GLM, it also introduces variability with the introduction of prior knowledge and the impact of this variability in group level studies might be difficult to quantify. Moreover, the computational cost of GLM is much lower than JDE and this might be important depending on the application. It is worth noting that most of the computational load in JDE comes from the estimation of the responses, and that if this is not done the computational load reduces considerably. In the end, one should choose the method of interest depending on the application.

## 9.1 Perspectives

This thesis treats several aspects of fMRI data analysis and more specifically functional ASL data analysis. The study of fASL signal through the development of data analysis methods and evaluation of real data has given rise to other questions that could be addressed in future work.

### 9.1.1 Short term perspectives

#### Improve the estimation of the perfusion response

Results on the HEROES dataset showed the performance of JDE in a block experimental design setting. The estimation of the perfusion response

function is very challenging and a block experimental design is not optimized for the temporal modelling of response shapes. However, a block experimental design increases the statistical power of the activation maps. Further analyses on fast event-related designs need to be performed. In any case, the estimation of a temporal perfusion response is interesting independently of the activation detection.

#### Physiological priors in Bayesian modelling

The estimation of the physiological parameters is interesting in itself. In this thesis, it has been shown that the physiological parameters have an impact on the convergence of the JDE algorithm. For this reason, if we could know which is the correct set of physiological parameters when running JDE, we would speed up convergence and attain a higher level of accuracy. This can be done with optimization techniques such as differential evolution, already used in [Mesejo et al., 2015, 2016] to estimate physiological parameters from BOLD time series. Following their lead, the estimation of physiological parameters could be done from the response shapes estimated from JDE. Using the correct set of physiological parameters, we could refine the estimation of the temporal responses in JDE by injecting the correct parameters in the physiological prior and adding a physiological parameter estimation step in the VEM algorithm. This way, the physiological prior becomes an adaptive physiological prior that updates along with the VEM convergence.

We showed that physiological priors can improve the temporal modelling of task-related responses when used in a Bayesian setting. This opens a door for the use of other physiological priors in a Bayesian modelling.

#### Introduction of basal perfusion as *a priori* knowledge in JDE

The main motivation for functional ASL data analysis is the possibility of quantifying the results and having actual units of blood flow. As we saw in the HEROES dataset, we can measure basal ASL by scanning the subject at rest, and functional ASL in an experimental setting. From both of them we can measure cerebral blood flow (CBF), although a higher confidence can be put in the basal CBF. The baseline perfusion is also estimated in JDE, and it would be interesting to inject the basal perfusion as prior knowledge into the Bayesian model. This can be easily done by injecting basal perfusion as the prior mean of the modelled baseline in the JDE approach, that we set to 0: $\alpha_j \sim \mathcal{N}(\alpha_{basal,j}, v_\alpha)$, where $\alpha_{basal,j}$ would be the basal perfusion measured at rest in voxel $j$. Recall that we showed that the perfusion maps from basal CBF measured at rest and during a functional experiment are reasonably similar.

### 9.1.2 Long term perspectives

#### Combination of BOLD and ASL analysis

The combination of BOLD and ASL is probably the best approach if we want to take advantage of both signals. Combining BOLD and ASL, one

could access other physiological parameters such as cerebral metabolic rate of oxygen (CMRO2).

In the context of JDE, this could be done with a hierarchical model similar to the one presented in chapter 5, in which a "true" HRF is assumed, and the HRF to estimate is considered the "noisy" HRF. The PRF is coupled to the "true" HRF through the HRF/PRF link $\boldsymbol{\Omega}$. When we consider BOLD and ASL signals, this is actually true: the "true" HRF or the best estimate would be the one estimated from BOLD signal, and the "noisy" HRF would be the one estimated from the hemodynamic component of the ASL signal. The estimation of the HRF and PRF responses has been shown to be the most difficult part in the analysis of fASL (see chapter 8). In BOLD, on the contrary, we observed an improvement in activation detection with respect to using fixed responses. For this reason, using the HRF estimates of the BOLD signal in the ASL signal analysis seems a reasonable perspective. We could also consider the correct set of physiological parameters when using the HRF/PRF link in ASL, if we estimate them.

Another possibility is using the activation state maps contained in $\boldsymbol{q}$ from the BOLD signal in the analysis of the hemodynamic component of the ASL signal. The activation state maps of the perfusion component could be therefore independent to the other component. In this work we showed that the activation maps are different for both components even if using the same activation state maps. However, given the low magnitude of the perfusion component, we might want to keep the coupling through the activation states $\boldsymbol{q}$.

The ASL signal is very noisy, and driving the estimation of some of its components with appropriate prior knowledge has led to a more robust estimation of the rest of the quantities. The JDE Bayesian setting gives the possibility to do this flexibly.

### Application to clinical research

Functional ASL has already been used in clinical applications for diseases such as schizophrenia [Kindler et al., 2013], multiple sclerosis [D'haeseleer et al., 2013] and Alzheimer's [Bron et al., 2014]. Its use can also be extended to drug studies [Wang et al., 2011a, Nordin et al., 2013] and population studies in general.

The absolute measure of cerebral blood flow and the fact that it is highly reproducible allow us to directly compare fASL results between subjects and in longitudinal studies. The analysis of functional ASL in a clinical context is interesting to detect functional anomalies or to follow-up a progressive cerebral recovery or degeneration. Since ASL is especially useful for neurovascular diseases, the estimation of a perfusion response function as in JDE becomes interesting: it provides additional temporally resolved information on the differences between pathological and non-pathological cases. Moreover, the fact that JDE provides localized measures due to the modelling of spatial correlation makes JDE appropriate for precise activation detection.

An interesting application could be vasoreactivity studies [Krainik et al., 2013] with BOLD and ASL, which could lead to disentangle the purely

vascular component of the BOLD response. The link between the kinetics measured with vasoreactivity BOLD and ASL, and with activation BOLD can reveal a purely vaso-motor component of vascular alteration observed in several pathologies. JDE has already been used for the analysis of vasoreactivity ASL in [Vincent et al., 2013b], demonstrating its potential for this particular application.

## 9.2 Concluding remarks

This thesis has been centered around the development of tools for the neuroscientific community to analyse functional ASL data. We proposed a framework to analyse this data in which the introduction of prior knowledge as physiological information becomes possible. More reliable information coming from other modalities can be easily introduced, too, in the form of prior. With the analysis of real data, we assessed the strong and weak points of the developed JDE method with respect to classical methods. Although the estimation of hemodynamic and perfusion responses can be difficult in JDE and does not seem to improve the detection of activation, the estimation in itself can be of interest. When fixing the responses in JDE, results improve in sensitivity and specificity when focusing on activation detection.

It is important to retain that the use of functional ASL data is envisaged for clinical research settings, in which having a quantitative and reproducible measure is important and even necessary for longitudinal studies. The use of precise tools in this setting becomes a necessity. However, stability and user-friendliness of these tools is crucial if we want the clinical research community to use them and benefit from them. For this reason, considerable effort is currently being put into making the PyHRF software ready for use in a clinical context in a collaboration between Inria, CEA/Neurospin, the *Grenoble Institute of Neuroscience* (GIN), and the *Centre Hospitalier Universitaire* (CHU) Grenoble Alpes.

From a clinical research point of view, the investigation of methods that can deliver quantified perfusion and functionally relevant variations is very interesting. Clinical studies using BOLD and ASL data could bring some light into the understanding of neurovascular and neurodegenerative diseases through longitudinal studies.

# A Notation

## A.1 Basics

| | |
|---|---|
| $n : 1...N$ | Scan indexes |
| $j : 1...J$ | Voxel indexes |
| $m : 1...M$ | Experimental condition indexes |
| $\boldsymbol{y}_j \in \mathbb{R}^N$ | Data fMRI time series acquired in voxel $j$ |
| $\boldsymbol{X}^m$ | $N \times (D+1)$ binary "onsets matrix" for condition $m$ |

## A.2 Forward model (bLTI)

| | |
|---|---|
| $\boldsymbol{b}_j \in \mathbb{R}^N$ | Gaussian noise vector in voxel $j$ |
| $d : 0...D$ | HRF coefficient indexes |
| $\boldsymbol{h} \in \mathbb{R}^{D+1}$ | HRF to be estimated |
| $a_j^m$ | HRL for voxel $j$ and condition $m$ |
| $\boldsymbol{g} \in \mathbb{R}^{D+1}$ | PRF to be estimated |
| $c_j^m$ | PRL for voxel $j$ and condition $m$ |
| $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_F]$ | Low frequency orthogonal matrix of size $N \times F$ |
| $\boldsymbol{\ell}_j \in \mathbb{R}^F$ | Nuisance parameter vector for voxel $j$ |
| $\alpha_j$ | Scalar modeling perfusion baseline at voxel $j$ |
| $\boldsymbol{w} \in \mathbb{R}^N$ | encodes the difference in magnetization signs between control and tag |
| $\boldsymbol{W} = \text{diag}(\boldsymbol{w})$ | diagonal matrix with $\boldsymbol{w}$ as diagonal components |

## A.3 Probabilistic model

| | |
|---|---|
| $\boldsymbol{\theta}$ | Set of all hyperparameters |
| $v_{b_j}$ | Noise variance for voxel $j$ |
| $\mu_0^m$ | Mean value of HRLs for non-active voxels in condition $m$ |
| $\mu_1^m$ | Mean value of HRLs for active voxels in condition $m$ |
| $\eta_0^m$ | Mean value of PRLs for non-active voxels in condition $m$ |
| $\eta_1^m$ | Mean value of PRLs for active voxels in condition $m$ |
| $v_0^m$ | Variance of HRLs for non-active voxels in condition $m$ |
| $v_1^m$ | Variance of HRLs for active voxels in condition $m$ |
| $\rho_0^m$ | Variance of PRLs for non-active voxels in condition $m$ |
| $\rho_1^m$ | Variance of PRLs for active voxels in condition $m$ |
| $\boldsymbol{q}$ | Vector of allocation variables coding for voxel states |
| $i$ | class index in the Gaussian mixture model (GMM) |
| $v_\alpha$ | variance of the perfusion baseline |
| $v_\ell$ | variance of the drifts |
| $v_h$ | variance of hemodynamic response function $\boldsymbol{h}$ |
| $\lambda_h$ | hyperparameter on $v_h$ |
| $v_g$ | variance of perfusion response function $\boldsymbol{g}$ |
| $\lambda_g$ | hyperparameter on $v_g$ |
| $\beta^m$ | interaction parameter of the 2-class Potts model for experimental condition $m$ |
| $\lambda_\beta$ | hyperparameter on $\beta^m$ |

## A.4 Physiological model

| | |
|---|---|
| $f(t)$: | cerebral blood flow (CBF) |
| $h(t)$: | BOLD signal |
| $E_0$: | Oxygen extraction fraction at baseline |
| $u(t)$ | neural activity (defined such that u(t) = 1 in response to stimulus) |
| $\psi(t)$ | flow inducing signal |
| $f_{in}(t)$ | normalized inflow to the balloon (CBF) |
| $f_{out}(t)$ | normalized outflow to the balloon (CBF) |
| $v(t)$ | normalized cerebral blood volume (CBV) |
| $\xi(t)$ | normalized deoxyhemoglobin (deoxyHb) content |
| $\eta$ | neural efficacy |
| $\tau_\psi$ | time constant for signal decay/elimination |
| $\tau_f$ | time constant for auto-regulatory feedback from blood flow |
| $\tau_m$ | mean transit time (average time for blood to traverse venous compartment) |
| $\tilde{w}$ | steady state flow parameter (Windkessel) |

# B Variational approximation for the JDE framework: BOLD multiple-session

The full joint model, considering $\boldsymbol{\theta} = \{\boldsymbol{\Gamma}, \boldsymbol{\mu}, \boldsymbol{v}, v_h, \boldsymbol{\beta}, \boldsymbol{\ell}\}$, reads:

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta}) &= p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}; \boldsymbol{\ell}, \boldsymbol{\Gamma})\ p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\ p(\boldsymbol{h}; v_h)\ p(\boldsymbol{q}; \boldsymbol{\beta})\ p(\boldsymbol{\theta}) \\ &= \prod_j \left( \prod_s \left( p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}; \boldsymbol{\ell}_j^s, \boldsymbol{\Gamma}_j^s) \prod_m p(a_j^{m,s} \mid \boldsymbol{q}_j^m; \mu^m, v^m) \right)\ p(\boldsymbol{q}_j; \beta) \right)\ p(\boldsymbol{h}; v_h)\ p(\boldsymbol{\theta}) \end{aligned}$$

## B.1 Likelihood

$$\forall j \in \mathcal{J}, \forall s \in \mathcal{S} \quad \boldsymbol{y}_j^s = \sum_{m=1}^{M} a_j^{m,s} \boldsymbol{X}^{m,s} \boldsymbol{h} + \boldsymbol{P} \boldsymbol{\ell}_j^s + \boldsymbol{b}_j^s, \tag{B.1}$$

where $\boldsymbol{X}^{m,s}$ is the linear function that associates $\boldsymbol{h}$ to $x^{m,s} * \boldsymbol{h}$ and will be different across sessions; $\boldsymbol{P}^s$ will be also different in each session; $\boldsymbol{b}_j^s$ are independent and normally distributed $\boldsymbol{b}_j^s \sim \mathcal{N}(0, (\Gamma_j^s)^{-1})$, where $\boldsymbol{\Gamma}_j^s = \boldsymbol{\Lambda}_j / v_{b_j}^s$. Here we consider White Gaussian noise, so $\boldsymbol{\Lambda}_j = \boldsymbol{I}$. We denote by $\boldsymbol{\ell} = \left\{ \boldsymbol{\ell}_j^s, j \in \mathcal{J}, s \in \mathcal{S} \right\}$ the set of low frequency drifts and by $\boldsymbol{\Gamma} = \left\{ \boldsymbol{\Gamma}_j^s, j \in \mathcal{J}, s \in \mathcal{S} \right\}$ the set of all precision matrices. It comes that:

$$p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}; \boldsymbol{\ell}, \boldsymbol{\Gamma}) = \prod_{s \in S} \prod_{j \in \mathcal{J}} p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}; \boldsymbol{\ell}_j^s, \boldsymbol{\Gamma}_j^s) \tag{B.2}$$

$$\text{with} \quad p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}; \boldsymbol{\ell}_j^s, \boldsymbol{\Gamma}_j^s) \sim \mathcal{N}\left( \sum_{m=1}^{M} a_j^{m,s} \boldsymbol{X}^{m,s} \boldsymbol{h} + \boldsymbol{P}^s \boldsymbol{\ell}_j^s, (\boldsymbol{\Gamma}_j^s)^{-1} \right) \tag{B.3}$$

## B.2 Variational EM

The expectation and maximization steps read as in the single-session model:

$$\textbf{E-H:} \quad \tilde{p}_h^{(r)}(h) \propto \exp\left( \mathrm{E}_{\tilde{p}_a^{(r-1)} \tilde{p}_q^{(r-1)}} \left[ \log p(\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{a}^{(r-1)}, \boldsymbol{q}^{(r-1)}; \boldsymbol{\theta}) \right] \right) \tag{B.4}$$

$$\textbf{E-A:} \quad \tilde{p}_a^{(r)}(a) \propto \exp\left( \mathrm{E}_{\tilde{p}_h^{(r)} \tilde{p}_q^{(r-1)}} \left[ \log p(\boldsymbol{a} \mid \boldsymbol{y}, \boldsymbol{h}^{(r)}, \boldsymbol{q}^{(r-1)}; \boldsymbol{\theta}^{(r)}) \right] \right) \tag{B.5}$$

$$\textbf{E-Q:} \quad \tilde{p}_q^{(r)}(q) \propto \exp\left( \mathrm{E}_{\tilde{p}_a^{(r)} \tilde{p}_h^{(r)}} \left[ \log p(\boldsymbol{q} \mid \boldsymbol{y}, \boldsymbol{a}^{(r)}, \boldsymbol{h}^{(r)}; \boldsymbol{\theta}^{(r)}) \right] \right) \tag{B.6}$$

and the corresponding M-step:

$$\textbf{M:} \quad \boldsymbol{\theta}^{(r+1)} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ \mathrm{E}_{\tilde{p}_a^{(r)} \tilde{p}_h^{(r)} \tilde{p}_q^{(r)}} \left[ \log p(\boldsymbol{y}, \boldsymbol{a}^{(r)}, \boldsymbol{h}^{(r)}, \boldsymbol{q}^{(r)}\ ; \boldsymbol{\theta}) \right]. \tag{B.7}$$

These steps are now given in more details in the next subsections for the multi-session case. For simplification, the $(r)$ and $(r-1)$ superscripts are omitted.

### B.2.1 E-H step

$$\widetilde{p}_h(\boldsymbol{h}) = \mathcal{N}(\tilde{\boldsymbol{m}}_h, \tilde{\boldsymbol{\Sigma}}_h), \text{ where} \tag{B.8}$$

$$\tilde{\boldsymbol{\Sigma}}_h^{-1} = \frac{\boldsymbol{R}^{-1}}{v_h} + \sum_{s\in S}\sum_{j\in J}\sum_{m,m'}\left(\tilde{m}_{a_j^{m,s}}\tilde{m}_{a_j^{m',s}} + \tilde{v}_{a_j^{m,s}a_j^{m',s}}\right)(\boldsymbol{X}^{m,s})^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{X}^{m',s} \tag{B.9}$$

$$\tilde{\boldsymbol{m}}_h = \tilde{\boldsymbol{\Sigma}}_h\sum_{s\in S}\sum_{j\in J}\sum_{m=1}^{M}\tilde{m}_{a_j^{m,s}}(\boldsymbol{X}^{m,s})^{\mathrm{t}}\boldsymbol{\Gamma}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s) \tag{B.10}$$

• If we consider the constraint in the norm of $\boldsymbol{h}$:

$$\tilde{\boldsymbol{h}} = \underset{\boldsymbol{h}\text{ s.t. }\boldsymbol{h}^T\boldsymbol{h}=1}{\arg\max}\left(\mathrm{E}_{\tilde{p}_a}\left[\log p(\boldsymbol{h}\,|\,\boldsymbol{y},\boldsymbol{a};\boldsymbol{\theta})\right]\right),$$

Solving it amounts to minimizing a quadratic function in $\boldsymbol{h}$ under a quadratic constraint, namely $\boldsymbol{h}^T\boldsymbol{h} = 1$. The function to be minimized is

$$(\boldsymbol{h} - \tilde{\boldsymbol{m}}_h)^T\tilde{\boldsymbol{\Sigma}}_h^{-1}(\boldsymbol{h} - \tilde{\boldsymbol{m}}_h) + \lambda\boldsymbol{h}^T\boldsymbol{h}$$

.

### B.2.2 E-A step

$$\widetilde{p}_a(\boldsymbol{a}) = \prod_s\prod_j\mathcal{N}\left(\tilde{\boldsymbol{m}}_{a_j^s}, \tilde{\boldsymbol{\Sigma}}_{a_j^s}\right), \text{ with:} \tag{B.11}$$

$$\tilde{\boldsymbol{\Sigma}}_{a_j^s} = \left(E_{\widetilde{p}_h}\left[(\boldsymbol{X}_h^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{X}_h^s\right] + \sum_{i=0}^{1}\Delta_{i,j}^s\right)^{-1} \tag{B.12}$$

$$\tilde{\boldsymbol{m}}_{a_j^s} = \tilde{\boldsymbol{\Sigma}}_{a_j^s}\left(E_{\widetilde{p}_h}\left[(\boldsymbol{X}_h^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)\right] + \sum_{i=0}^{1}\Delta_{i,j}^s(\boldsymbol{\mu}_i + \mu_s)\right) \text{ where } \Delta_{i,j}^s = \mathrm{diag}_M\left[\frac{\widetilde{p}_{q_j^m}}{v_i^m + v_s}\right] \tag{B.13}$$

$E_{\widetilde{p}_h}\left[(\boldsymbol{X}_h^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{X}_h^s\right]$ is a matrix of dimension $M\times M$ where $\boldsymbol{X}_h^s = [\boldsymbol{X}^{1,s}\boldsymbol{h}|\boldsymbol{X}^{2,s}\boldsymbol{h}|...|\boldsymbol{X}^{m,s}\boldsymbol{h}]$ and each element $(m,m')$ is written:

$$E_{\widetilde{p}_h}\left[(\boldsymbol{X}^{m,s}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{X}^{m',s}\boldsymbol{h}\right] = \tilde{\boldsymbol{m}}_h^{\mathrm{t}}(\boldsymbol{X}^{m,s})^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{X}^{m',s}\tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h(\boldsymbol{X}^{m,s})^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{X}^{m',s}\right) \tag{B.14}$$

$E_{\widetilde{p}_h}\left[(\boldsymbol{X}_h^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)\right]$ is a vector of dimension $M\times 1$ where each element $(m,1)$ is written:

$$E_{\widetilde{p}_h}\left[(\boldsymbol{X}^{m,s}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Gamma}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)\right] = \tilde{\boldsymbol{m}}_h^{\mathrm{t}}(\boldsymbol{X}^{m,s})^{\mathrm{t}}\boldsymbol{\Gamma}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s) \tag{B.15}$$

In the end, we return the level mean over sessions $\bar{\boldsymbol{a}}_j^m$, with $\bar{\boldsymbol{a}}_j = \{\bar{\boldsymbol{a}}_j^m, m = 1..M\}$ and $\tilde{\boldsymbol{m}}_{a_j^s} = \{\tilde{\boldsymbol{m}}_{a_j^{m,s}}, m = 1..M, s = 1..S\}$:

$$\bar{\boldsymbol{a}}_j = \sum_s\frac{\tilde{\boldsymbol{m}}_{a_j^s}}{S} \tag{B.16}$$

• If we consider the constraint in the norm of h: $\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s = [\boldsymbol{X}^{1,s}\tilde{\boldsymbol{h}}|\boldsymbol{X}^{2,s}\tilde{\boldsymbol{h}}|...|\boldsymbol{X}^{m,s}\tilde{\boldsymbol{h}}]$ and the computation of the distribution is directly:

$$\widetilde{p}_a(\boldsymbol{a}) = \prod_s\prod_j\mathcal{N}\left(\tilde{\boldsymbol{m}}_{a_j^s}, \tilde{\boldsymbol{\Sigma}}_{a_j^s}\right), \text{ with:} \tag{B.17}$$

$$\tilde{\boldsymbol{\Sigma}}_{a_j^s} = \left((\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s + \sum_{i=0}^{1}\Delta_{i,j}^s\right)^{-1} \tag{B.18}$$

$$\tilde{\boldsymbol{m}}_{a_j^s} = \tilde{\boldsymbol{\Sigma}}_{a_j^s}\left((\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s) + \sum_{i=0}^{1}\Delta_{i,j}^s(\boldsymbol{\mu}_i + \mu_s)\right) \tag{B.19}$$

Note here that when $v_s$ is high, $\Delta^s_{i,j}$ will be low and $(\boldsymbol{\mu}_i + \boldsymbol{\mu}_s)$ will have less weight in $\tilde{\boldsymbol{m}}_{a_j^s}$. Therefore, the sessions with a low variance $v_s$ will dominate the sum.

### B.2.3 E-Q step

We assume

$$\tilde{p}_{q^m}(\boldsymbol{q}^m) = \prod_{j \in \mathcal{J}} \tilde{p}_{q_j^m}(q_j^m) \tag{B.20}$$

$$\begin{aligned}\widetilde{p}_{q_j^m}(q_j^m) &\propto \exp\left(E_{\widetilde{p}_a, \widetilde{p}_{q_{\setminus j}^m}, \widetilde{p}_{q^{\setminus m}}}\left[\log p(q_j^m \mid q_{\setminus j}^m, q^{\setminus m}, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{\theta})\right]\right) \\ &\propto \exp\left(E_{\widetilde{p}_{a_j^m}}\left[\sum_s \log p(a_j^{m,s} \mid q_j^m, \boldsymbol{\theta})\right] + E_{\widetilde{p}_{q_{\setminus j}^m}}\left[\log p(q_j^m \mid q_{\setminus j}^m; \beta^m)\right]\right)\end{aligned} \tag{B.21}$$

**1)** Computation $E_{\widetilde{p}_{a_j^{m,s}}}\left[\log p(a_j^{m,s} \mid q_j^m, \boldsymbol{\theta})\right]$:

$$\log p(a_j^{m,s} \mid q_j^m, \boldsymbol{\theta}) = (1 - q_j^m) \log \mathcal{N}(a_j^{m,s}; \mu_0^m + \mu_s, v_0^m + v_s) + q_j^m \log \mathcal{N}(a_j^{m,s}; \mu_1^m + \mu_s, v_1^m + v_s) \tag{B.22}$$

considering $E_{\tilde{p}_{a_j^{m,s}}}[a_j^{m,s} a_j^{m',s}] = \tilde{v}_{a_j^{m,s} a_j^{m',s}} + \tilde{m}_{a_j^{m,s}} \tilde{m}_{a_j^{m',s}}$, we get

$$\begin{aligned}E_{\widetilde{p}_{a_j^{m,s}}}\left[\log p(a_j^{m,s} \mid q_j^m, \boldsymbol{\theta})\right] &= (1 - q_j^m)\left(\log \mathcal{N}(\tilde{m}_{a_j^{m,s}}; \mu_0^m + \mu_s, v_0^m + v_s) - \frac{1}{2}\frac{\tilde{v}_{a_j^{m,s} a_j^{m,s}}}{v_0^m + v_s}\right) \\ &\quad + q_j^m\left(\log \mathcal{N}(\tilde{m}_{a_j^{m,s}}; \mu_1^m + \mu_s, v_1^m + v_s) - \frac{1}{2}\frac{\tilde{v}_{a_j^{m,s} a_j^{m,s}}}{v_1^m + v_s}\right)\end{aligned} \tag{B.23}$$

**2)** Computation of $E_{\widetilde{p}_{q_{\setminus j}^m}}\left[\log p(q_j^m \mid q_{\setminus j}^m; \beta^m)\right]$: Considering the *a priori* distribution of variable $\boldsymbol{q}$

$$\log p(q_j^m \mid q_{\setminus j}^m; \beta^m) = \beta^m \sum_{k \in N(j)} \mathbb{1}(q_j^m = q_k^m) + C_{3, q_j^m} \tag{B.24}$$

where $N(j)$ is the neighborhood of $j$ and where $C_{3,q_j^m}$ is a constant in $q_j^m$.

$$E_{\widetilde{p}_{q_{\setminus j}^m}}\left[\log p(q_j^m \mid q_{\setminus j}^m; \beta^m)\right] = C'_{3,q_j^m} + \beta^m \sum_{k \in N(j)} E_{\widetilde{p}_{q_k^m}}\left[\mathbb{1}(q_j^m = q_k^m)\right] \tag{B.25}$$

with $C'_{3,q_j^m} = E_{\widetilde{p}_{q_{\setminus j}^m}}\left[C_{3,q_j^m}\right]$, and $E_{\widetilde{p}_{q_{\setminus j}^m}}\left[\mathbb{1}(q_j^m = q_k^m)\right] = p_{q_k^m}(q_j^m)$. Finally, we get

$$\begin{aligned}\widetilde{p}_{q_j^m}(q_j^m) \propto \exp\Bigg((1 - q_j^m) \sum_s \left(\log \mathcal{N}(\tilde{m}_{a_j^{m,s}}; \mu_0^m + \mu_s, v_0^m + v_s) - \frac{1}{2}\frac{\tilde{v}_{a_j^{m,s} a_j^{m,s}}}{v_0^m + v_s}\right) \\ + q_j^m \sum_s \left(\log \mathcal{N}(\tilde{m}_{a_j^{m,s}}; \mu_1^m + \mu_s, v_1^m + v_s) - \frac{1}{2}\frac{\tilde{v}_{a_j^{m,s} a_j^{m,s}}}{v_1^m + v_s}\right) + \beta^m \sum_{k \in N(j)} \tilde{p}_{q_k^m}(q_j^m)\Bigg)\end{aligned} \tag{B.26}$$

### B.2.4 M-step

As for the E-step above, the superscript $(r)$ is omitted in the following developments about the M-step (E.12).

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathrm{E}_{\tilde{p}_a \tilde{p}_h \tilde{p}_q}\left[\log p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{q}; \boldsymbol{\theta})\right] \\ &= \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Big[\mathrm{E}_{\tilde{p}_a \tilde{p}_h \tilde{p}_q}\left[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}; \boldsymbol{\ell}, \boldsymbol{\Gamma})\right] + \mathrm{E}_{\tilde{p}_a \tilde{p}_q}\left[\log p(\boldsymbol{a} \mid \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\right] \\ &\qquad + \mathrm{E}_{\tilde{p}_q}\left[\log p(\boldsymbol{q}; \boldsymbol{\beta})\right] + \log p(\boldsymbol{\beta}; \lambda_\beta) + \mathrm{E}_{\tilde{p}_h}\left[\log p(\boldsymbol{h}; v_h)\right] + \log p(v_h; \lambda_h)\Big] .\end{aligned}$$

Given the separability of the priors pdfs, it follows that the M-step also divides into separate M-steps:

### M-$(\mu, v)$ step

Updating parameters $\boldsymbol{\mu}$ and $\boldsymbol{v}$ is straightforward since closed-form expressions are available:

$$(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{v}}) = \arg\max_{\boldsymbol{\mu},\boldsymbol{v}} \mathrm{E}_{\tilde{p}_a \tilde{p}_q}\big[\log p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big]$$

In the E-Q step we already computed $\mathrm{E}_{\tilde{p}_a}\big[\log p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big]$ from which it comes:

$$\mathrm{E}_{\tilde{p}_q \tilde{p}_a}\big[\log p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big] = \sum_{m=1}^{M} \sum_{j \in \mathcal{J}} \sum_{i=\{0,1\}} \tilde{p}_{q_j^m}(i) \sum_{s=1}^{S} \left( \log \mathcal{N}(m_{a_j^{m,s}}; \mu_i^m + \mu_s, v_i^m + v_s) - \frac{1}{2} \frac{\tilde{v}_{a_j^{m,s} a_j^{m,s}}}{v_i^m + v_s} \right).$$

This latter expression is similar to the one to be maximized in a standard Gaussian mixture when replacing the observed data by the $m_{a_j^{m,s}}$'s. Not surprisingly, we get:

$$\widehat{(\mu_i^m + \mu_s)} = \frac{\sum_{j \in \mathcal{J}} \tilde{p}_{q_j^m}^{(r)}(i)\; m_{a_j^{m,s}}}{\sum_{j \in \mathcal{J}} \tilde{p}_{q_j^m}^{(r)}(i)} \quad \text{and} \quad \widehat{(v_i^m + v_s)} = \frac{\sum_{j \in \mathcal{J}} \tilde{p}_{q_j^m}^{(r)}(i) \left[ (m_{a_j^{m,s}} - (\mu_i^m + \mu_s))^2 + \tilde{v}_{a_j^{m,s} a_j^{m,s}} \right]}{\sum_{j \in \mathcal{J}} \tilde{p}_{q_j^m}^{(r)}(i)}$$

### M-$v_h$ step

$$\hat{v}_h = \arg\max_{v_h} f(v_h) = \arg\max_{v_h} \left\{ \mathrm{E}_{\tilde{p}_h}\big[\log p(\boldsymbol{h} \,|\, v_h) + \log p(v_h \,|\, \lambda_{v_h})\big] \right\} \tag{B.27}$$

where the *a priori* $p(v_h \,|\, \lambda_{v_h}) = \lambda_{v_h} \exp(-\lambda_{v_h} v_h)$ allows the better estimation of $v_h$ [Chaari et al., 2013].

1. If we use an hyperprior

$$\mathrm{E}_{\tilde{p}_H}\big[\log p(\boldsymbol{h} \,|\, v_h)\big] = C_{v_h} - \frac{D-1}{2} \log v_h - \frac{\tilde{\boldsymbol{m}}_h^{\mathrm{t}} \boldsymbol{R}^{-1} \tilde{\boldsymbol{m}}_h}{2 v_h}$$

$$\frac{\partial f(v_h)}{\partial v_h} = -\frac{D-1}{2 v_h} + \frac{\tilde{\boldsymbol{m}}_h^{\mathrm{t}} \boldsymbol{R}^{-1} \tilde{\boldsymbol{m}}_h}{2 v_h^2} - \lambda_{v_h} = 0$$

and $\tilde{\boldsymbol{m}}_h^{\mathrm{t}} \boldsymbol{R}^{-1} \tilde{\boldsymbol{m}}_h = \mathrm{tr}\left(\tilde{\boldsymbol{m}}_h \tilde{\boldsymbol{m}}_h^{\mathrm{t}} \boldsymbol{R}^{-1}\right)$

$$\Rightarrow \quad v_h = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_{v_h}\left(\tilde{\boldsymbol{m}}_h^{\mathrm{t}} \boldsymbol{R}^{-1} \tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h \boldsymbol{R}^{-1}\right)\right)}}{4\lambda_{v_h}} \tag{B.28}$$

- If we do consider the constraint in the norm of $\boldsymbol{h}$:

$$\Rightarrow \quad v_h = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_{v_h} \mathrm{tr}\left[\left(\tilde{\boldsymbol{h}} \tilde{\boldsymbol{h}}^{\mathrm{t}}\right) \boldsymbol{R}^{-1}\right]}}{4\lambda_{v_h}} \tag{B.29}$$

2. If we do not use an hyperprior:

$$\Rightarrow \quad v_h = \frac{\tilde{\boldsymbol{m}}_h^{\mathrm{t}} R^{-1} \tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h \boldsymbol{R}^{-1}\right)}{D-1} \tag{B.30}$$

- If we consider the constraint in the norm of $\boldsymbol{h}$ when we do not use an hyperprior:

$$\Rightarrow \quad v_h = \frac{\mathrm{tr}\left[\left(\tilde{\boldsymbol{h}} \tilde{\boldsymbol{h}}^{\mathrm{t}}\right) \boldsymbol{R}^{-1}\right]}{D-1} \tag{B.31}$$

### M-$\beta$ step

The maximization over each $\beta^m$ corresponds to the M-step obtained for a standard Hidden MRF model:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^m &= \arg\max_{\beta^m} \mathrm{E}_{\tilde{p}_{q^m}}\left[\log p(\boldsymbol{q}^m;\beta^m)\right] \\ &= \arg\max_{\beta^m} E_{\tilde{p}_{q^m}}[\log p(\boldsymbol{q}^m \mid \beta^m)] + \log p(\beta^m \mid \lambda_\beta) \\ &= \arg\max_{\beta^m} \left\{-\log Z(\beta^m) + \beta^m\left[E_{\tilde{p}_{q^m}}[U(\boldsymbol{q}^m)] - \lambda_\beta\right] + C_{\beta^m}\right\} = \arg\max_{\beta^m} f(\beta^m)\end{aligned}$$

The maximization of $f(\beta^m)$ needs the computation of its derivative with respect to $\beta^m$:

$$\frac{df(\beta^m)}{d\beta^m} = -\frac{d\log Z(\beta^m)}{d\beta^m} + E_{\tilde{p}_{q^m}}[U(\boldsymbol{q}^m)] - \lambda_\beta \tag{B.32}$$

where

$$Z(\beta^m) = \sum_{\boldsymbol{q}^m} \exp(\beta^m U(\boldsymbol{q}^m)) \tag{B.33}$$

$$\mathrm{E}_{\tilde{p}_{q^m}}\left[U(\boldsymbol{q}^m)\right] = \sum_{j\sim k} \mathrm{E}_{\tilde{p}_{q^m}}\left[\mathbb{1}(q_j^m = q_k^m)\right] = \frac{1}{2}\sum_j \sum_{k\in N(j)} \tilde{p}_{q^m}(q_j^m = q_k^m) \tag{B.34}$$

$$= \frac{1}{2}\sum_j \sum_{k\in N(j)} (\tilde{p}^{(r)}_{q_j^m}(0)\tilde{p}^{(r)}_{q_k^m}(0) + \tilde{p}^{(r)}_{q_j^m}(1)\tilde{p}^{(r)}_{q_k^m}(1)) \tag{B.35}$$

$$\frac{d\log Z(\beta^m)}{d\beta^m} = \mathrm{E}_{p_{q^m}}\left[U(\boldsymbol{q}^m)\right] \tag{B.36}$$

where $p_{q^m} = p(\boldsymbol{q}^m;\beta^m)$ the prior MRF defined in section 2.6.
It can be solved using a mean field like approximation as done in [Celeux et al., 2003]:

$$\frac{df(\beta^m)}{d\beta^m} = -E_{\tilde{p}^{MF}}[U(\boldsymbol{q}^m)] + E_{\tilde{p}_{q^m}}[U(\boldsymbol{q}^m)] - \lambda_\beta \tag{B.37}$$

where $\tilde{p}^{MF}(\boldsymbol{q}^m) = \prod_j \tilde{p}^{MF}(q_j^m)$ with $\tilde{p}^{MF}(q_j^m;\beta^m) = \dfrac{\exp(\beta^m \sum_{k\in N(j)} \tilde{p}^{(r-1)}_{q_k^m}(q_j^m))}{\sum_{i\in\{0,1\}} \exp(\beta^m \sum_{k\in N(j)} \tilde{p}^{(r-1)}_{q_k^m}(i))}$
It follows that $\beta^m$ must satisfy the following equation:

$$\sum_j \sum_{k\in N(j)} \sum_{i=\{0,1\}} \left(\tilde{p}^{(r)}_{q_j^m}(i)\ \tilde{p}^{(r)}_{q_k^m}(i) - \tilde{p}^{MF}(q_j^m = i;\beta^m)\ \tilde{p}^{MF}(q_k^m = i;\beta^m)\right) - \lambda_\beta = 0 \tag{B.38}$$

### M-$(\boldsymbol{\ell},\boldsymbol{\Gamma})$ step

$$(\widehat{\boldsymbol{\ell}},\hat{\boldsymbol{\Gamma}}) = \arg\max_{\boldsymbol{\ell},\boldsymbol{\Gamma}} \mathrm{E}_{\tilde{p}_a}\left[\log p(\boldsymbol{y} \mid \boldsymbol{a},\boldsymbol{h};\boldsymbol{\ell},\boldsymbol{\Gamma})\right].$$

This maximization problem factorizes over voxel and session so that for each $j \in \mathcal{J}$ and $s \in \mathcal{S}$, we compute:

$$(\widehat{\boldsymbol{\ell}}_j^s, \hat{v}_{b_j}^s) = \arg\max_{\boldsymbol{\ell}_j^s,\boldsymbol{\Gamma}_j^s} \mathrm{E}_{\tilde{p}_{a_j^s}}\left[\log p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s,\boldsymbol{h};\boldsymbol{\ell}_j^s,\boldsymbol{\Gamma}_j^s)\right].$$

**M-$v_{b_j}$** Considering the independence between $\{\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s,\boldsymbol{h};\boldsymbol{\ell}_j^s,v_{b_j}^s\}_{j=1:J}$ we can write:

$$\hat{v}_{b_j}^s = \arg\max_{v_{b_j}^s} E_{\tilde{p}_{a_j^s}}\left[\log p\left(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s,\boldsymbol{h};\boldsymbol{\ell}_j^s,v_{b_j}^s\right)\right] \tag{B.39}$$

From the definition of likelihood we can compute, considering $\breve{\boldsymbol{y}}_j^s = \boldsymbol{y}_j^s - \sum_m a_j^{m,s} \boldsymbol{X}^{m,s}\boldsymbol{h} - \boldsymbol{P}^s\boldsymbol{\ell}_j^s$:

$$E_{\widetilde{p}_h,\widetilde{p}_{a_j^s}}\Big[\log p\left(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}, \boldsymbol{\ell}_j^s, v_{b_j}^s\right)\Big] = C_{v_{b_j}^s} - \frac{N}{2}\log v_{b_j}^s - \frac{E_{\widetilde{p}_h,\widetilde{p}_{a_j^s}}\Big[(\widetilde{\boldsymbol{y}}_j^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\widetilde{\boldsymbol{y}}_j^s\Big]}{2}$$

And we have:

$$\begin{aligned}\hat{v}_{b_j}^s = \frac{1}{N}\Big[&\tilde{\boldsymbol{m}}_{a_j^s}^{\mathrm{t}} E_{\widetilde{p}_h}\left[(\boldsymbol{X}_{\boldsymbol{h}}^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s\boldsymbol{X}_{\boldsymbol{h}}^s\right]\tilde{\boldsymbol{m}}_{a_j^s} + tr\left(\tilde{\boldsymbol{\Sigma}}_{a_j^s} E_{\widetilde{p}_h}\left[(\boldsymbol{X}_{\boldsymbol{h}}^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s\boldsymbol{X}_{\boldsymbol{h}}^s\right]\right)\\ &-2\tilde{\boldsymbol{m}}_{a_j^s}^{\mathrm{t}} E_{\widetilde{p}_h}\left[(\boldsymbol{X}_{\boldsymbol{h}}^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)\right] + (\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)\Big]\end{aligned} \tag{B.40}$$

• If we constraint the norm of $\boldsymbol{h}$:

$$\begin{aligned}\hat{v}_{b_j}^s = \frac{1}{N}\Big[&\tilde{\boldsymbol{m}}_{a_j^s}^{\mathrm{t}}(\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s\tilde{\boldsymbol{m}}_{a_j^s} + tr\left(\tilde{\boldsymbol{\Sigma}}_{a_j^s}(\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s\right)\\ &-2\tilde{\boldsymbol{m}}_{a_j^s}^{\mathrm{t}}(\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s) + (\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)^{\mathrm{t}}\boldsymbol{\Lambda}_j^s(\boldsymbol{y}_j^s - \boldsymbol{P}^s\boldsymbol{\ell}_j^s)\Big]\end{aligned} \tag{B.41}$$

**M-$\boldsymbol{\ell}$**

For the same reason (independence between $\{\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}, \boldsymbol{\ell}_j^s, v_{b_j}^s\}_{j=1:J}$) we can compute:

$$\hat{\boldsymbol{\ell}}_j^s = \arg\max_{\boldsymbol{\ell}_j^s} E_{\widetilde{p}_{a_j^s},\widetilde{p}_h}\left[\log p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}, \boldsymbol{\ell}_j^s, v_{b_j}^s)\right] \tag{B.42}$$

Considering $\bar{\boldsymbol{y}}_j^s = \boldsymbol{y}_j^s - \sum_m a_j^{m,s}\boldsymbol{X}^{m,s}\boldsymbol{h}$:

$$\begin{aligned}\log p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}, \boldsymbol{\ell}_j^s, v_{b_j}^s) &= C_{\boldsymbol{\ell}_j^s} + \left(\boldsymbol{y}_j^s - \sum_m a_j^{m,s}\boldsymbol{X}^{m,s}\boldsymbol{h} - \boldsymbol{P}^s\boldsymbol{\ell}_j^s\right)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\left(\boldsymbol{y}_j^s - \sum_m a_j^{m,s}\boldsymbol{X}^{m,s}\boldsymbol{h} - \boldsymbol{P}^s\boldsymbol{\ell}_j^s\right)\\ &= C_{\boldsymbol{\ell}_j^s} + (\bar{\boldsymbol{y}}_j^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\bar{\boldsymbol{y}}_j^{s} - (\bar{\boldsymbol{y}}_j^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s\boldsymbol{\ell}_j^s - (\boldsymbol{\ell}_j^s)^{\mathrm{t}}(\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\bar{\boldsymbol{y}}_j^s + (\boldsymbol{\ell}_j^s)^{\mathrm{t}}(\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s\boldsymbol{\ell}_j^s\end{aligned} \tag{B.43}$$

$$\begin{aligned}&E_{\widetilde{p}_{a_j^s},\widetilde{p}_h}\left[\log p(\boldsymbol{y}_j^s \mid \boldsymbol{a}_j^s, \boldsymbol{h}, \boldsymbol{\ell}_j^s, v_{b_j}^s)\right]\\ &= C_{\boldsymbol{\ell}_j^s} + E_{\widetilde{p}_{a_j^s},\widetilde{p}_h}\left[(\bar{\boldsymbol{y}}_j^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\bar{\boldsymbol{y}}_j^s\right] - 2\,(\boldsymbol{\ell}_j^s)^{\mathrm{t}}(\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s E_{\widetilde{p}_{a_j^s},\widetilde{p}_h}\left[\bar{\boldsymbol{y}}_j^s\right] + (\boldsymbol{\ell}_j^s)^{\mathrm{t}}(\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s\boldsymbol{\ell}_j^s\end{aligned} \tag{B.44}$$

By deriving the previous equation with respect to $\boldsymbol{\ell}_j$ we get:

$$\frac{\partial(.)}{\partial\boldsymbol{\ell}_j^s} = -2(\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s E_{\widetilde{p}_{a_j^s},\widetilde{p}_h}\left[\bar{\boldsymbol{y}}_j^s\right] + 2(\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s\boldsymbol{\ell}_j^s = \mathrm{Vect}\,0 \tag{B.45}$$

$$\begin{aligned}\Rightarrow\quad \hat{\boldsymbol{\ell}}_j^s &= ((\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s)^{-1}(\boldsymbol{P}^s)^{\mathrm{t}}E_{\widetilde{p}_{a_j^s},\widetilde{p}_h}\left[\bar{\boldsymbol{y}}_j^s\right]\\ &= ((\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s)^{-1}(\boldsymbol{P}^s)^{\mathrm{t}}\left(\boldsymbol{y}_j^s - \sum_m \tilde{m}_{a_j^{m,s}}\boldsymbol{X}^{m,s}\tilde{\boldsymbol{m}}_h\right)\end{aligned} \tag{B.46}$$

• If we constraint the norm of $\boldsymbol{h}$:

$$\begin{aligned}\Rightarrow\quad \hat{\boldsymbol{\ell}}_j^s &= ((\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s)^{-1}(\boldsymbol{P}^s)^{\mathrm{t}}E_{\widetilde{p}_{a_j^s}}\left[\bar{\boldsymbol{y}}_j^s\right]\\ &= ((\boldsymbol{P}^s)^{\mathrm{t}}\boldsymbol{\Gamma}_j^s\boldsymbol{P}^s)^{-1}(\boldsymbol{P}^s)^{\mathrm{t}}\left(\boldsymbol{y}_j^s - \sum_m \tilde{m}_{a_j^{m,s}}\boldsymbol{X}^{m,s}\tilde{\boldsymbol{h}}\right)\end{aligned} \tag{B.47}$$

# C Linearization of Physiological Model

From the physiological model introduced in the main text that we call extended Balloon model, we get a set of differential equations. We want to linearize this non-linear system with four state variables $\{s, f_{in}, v, q\}$:

$$\begin{cases} \dot{\psi} = \alpha u - \frac{\psi}{\tau_\psi} - \frac{f_{in}-1}{\tau_f} \\ \dot{f_{in}} = \psi \\ \dot{q} = \frac{1}{\tau_m}\left( f_{in}\frac{1-(1-E_0)^{1/f_{in}}}{E_0} - q v^{\frac{1}{\tilde{w}}-1} \right) \\ \dot{v} = \frac{1}{\tau_m}\left( f_{in} - v^{\frac{1}{\tilde{w}}} \right) \end{cases} \tag{C.1}$$

Considering $\{x_1, x_2, x_3, x_4\} = \{\psi, f_{in} - 1, 1 - v, 1 - q\}$, we have:

$$\begin{cases} \dot{x}_1 = \alpha u - \frac{x_1}{\tau_\psi} - \frac{x_2}{\tau_f} \\ \dot{x}_2 = x_1 \\ \dot{x}_3 = -\frac{1}{\tau_m}\left( (x_2 + 1) - (1 - x_3)^{\frac{1}{\tilde{w}}} \right) \\ \dot{x}_4 = -\frac{1}{\tau_m}\left( (x_2 + 1)\frac{1-(1-E_0)^{\frac{1}{x_2+1}}}{E_0} - (1 - x_4)(1 - x_3)^{\frac{1}{\tilde{w}}-1} \right) \end{cases} \tag{C.2}$$

To solve non-linear systems, we can analyze the dynamic system to check if there are steady-states: time derivatives to 0, or our resting state. The linearization goes: we choose a relevant point to make our linear approximation around it, like the resting point $\{\psi, f_{in} - 1, 1 - v, 1 - q\} = \{x_1, x_2, x_3, x_4\} = \{\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}\}$, and then we calculate the Jacobian matrix at that point:

$$J = \begin{bmatrix} \frac{\partial \dot{x}_1}{\partial x_1} & \frac{\partial \dot{x}_1}{\partial x_2} & \frac{\partial \dot{x}_1}{\partial x_3} & \frac{\partial \dot{x}_1}{\partial x_4} \\ \frac{\partial \dot{x}_2}{\partial x_1} & \frac{\partial \dot{x}_2}{\partial x_2} & \frac{\partial \dot{x}_2}{\partial x_3} & \frac{\partial \dot{x}_1}{\partial x_4} \\ \frac{\partial \dot{x}_3}{\partial x_1} & \frac{\partial \dot{x}_3}{\partial x_2} & \frac{\partial \dot{x}_3}{\partial x_3} & \frac{\partial \dot{x}_3}{\partial x_4} \\ \frac{\partial \dot{x}_4}{\partial x_1} & \frac{\partial \dot{x}_4}{\partial x_2} & \frac{\partial \dot{x}_4}{\partial x_3} & \frac{\partial \dot{x}_4}{\partial x_4} \end{bmatrix} = \begin{bmatrix} \frac{-1}{\tau_\psi} & \frac{-1}{\tau_f} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{-1}{\tau_m} & \frac{-1}{\tau_m \tilde{w}} & 0 \\ 0 & -\frac{1}{\tau_m}\left(1 + \frac{(1-E_0)\ln(1-E_0)}{E_0}\right) & -\frac{1-\tilde{w}}{\tilde{w}\tau_m} & \frac{-1}{\tau_m} \end{bmatrix} \tag{C.3}$$

All of them are quite straight-forward. In the case of $\frac{\partial \dot{x}_4}{\partial x_2}$,

$$\frac{\partial \dot{x}_4}{\partial x_2} = \frac{-1}{\tau_m E_0}\left( \left(1 - (1 - E_0)^{\frac{1}{x_2+1}}\right) + (x_2 + 1)\left( (1 - E_0)^{\frac{1}{x_2+1}} \ln(1 - E_0)\frac{1}{(x_2 + 1)^2} \right) \right)$$

if we consider the steady-state point $x_2 = \mathbf{0}$,

$$= -\frac{1}{\tau_m}\left(1 + \frac{(1 - E_0)\ln(1 - E_0)}{E_0}\right) \tag{C.4}$$

After linearization (taylor series: $f(0) + f(x_1, x_2, x_3, x_4) + ...$), we have:

$$\begin{cases} \dot{x}_1 = \alpha u - \frac{x_1}{\tau_\psi} - \frac{x_2}{\tau_f} \\ \dot{x}_2 = x_1 \\ \dot{x}_3 = -\frac{1}{\tau_m}\left( x_2 + \frac{x_3}{\tilde{w}} \right) \\ \dot{x}_4 = -\frac{1}{\tau_m}\left( \left(1 + \frac{(1-E_0)\ln(1-E_0)}{E_0}\right) x_2 - \frac{1+\tilde{w}}{\tilde{w}} x_3 + x_4 \right) \end{cases} \tag{C.5}$$

Considering $\gamma = \frac{1}{\tau_m}\left(1 + \frac{(1-E_0)\ln(1-E_0)}{E_0}\right)$, we can diagonalize the system by considering $\boldsymbol{D}$ the differentiation operator and $\boldsymbol{I}$ the identity matrix. If we just consider the relationship with $\boldsymbol{x}_2$:

$$\begin{cases} (\boldsymbol{D})\{\boldsymbol{x}_2\} = \boldsymbol{x}_1 \\ \left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)\{\boldsymbol{x}_3\} = -\frac{1}{\tau_m}\boldsymbol{x}_2 \\ \left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tau_m}\right)\{\boldsymbol{x}_4\} = -\left(\gamma\boldsymbol{I} - \frac{1-\tilde{w}}{\tilde{w}\tau_m^2}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right)\boldsymbol{x}_2 \end{cases} \tag{C.6}$$

Now, we introduce the BOLD response $h(t)$, which is the HRF when $u(t) = \delta(t)$:

$$\begin{aligned} \boldsymbol{h} &= V_0(k_1(1-\boldsymbol{q}) + k_2(1-\boldsymbol{q}\boldsymbol{v}^{-1}) + k_3(1-\boldsymbol{v})) \\ &= V_0(k_1\boldsymbol{x}_4 + k_2\left(1-(1-\boldsymbol{x}_4)(1-\boldsymbol{x}_3)^{-1}\right) + k_3\boldsymbol{x}_3) \\ &= V_0(k_1\boldsymbol{x}_4 + k_2(\boldsymbol{x}_4-\boldsymbol{x}_3)(1-\boldsymbol{x}_3)^{-1} + k_3\boldsymbol{x}_3) \end{aligned} \tag{C.7}$$

The Jacobian is:

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}_3} & \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}_4} \end{bmatrix} = \begin{bmatrix} V_0(k_3-k_2) & V_0(k_1+k_2) \end{bmatrix} \tag{C.8}$$

And the system of equations:

$$\boldsymbol{h} = V_0 \begin{bmatrix} (k_3-k_2) & (k_1+k_2) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_3 \\ \boldsymbol{x}_4 \end{bmatrix} \tag{C.9}$$

Therefore, after linearization we have

$$\boldsymbol{h} = V_0((k_1+k_2)\boldsymbol{x}_4 + (k_3-k_2)\boldsymbol{x}_3) \tag{C.10}$$

We want to find the relation between $\boldsymbol{h}$ and $\boldsymbol{x}_2 = \boldsymbol{f}_{in} - 1$, as $\boldsymbol{x}_2$ represents the inflow of blood in the ballon.

$$\boldsymbol{h} = -V_0\left((k_1+k_2)\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tau_m}\right)^{-1}\left(\gamma\boldsymbol{I} - \frac{1-\tilde{w}}{\tilde{w}\tau_m^2}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right) + \frac{k_3-k_2}{\tau_m}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right)\boldsymbol{x}_2 \tag{C.11}$$

The relation $\boldsymbol{h} = \boldsymbol{\Omega}^{-1}\boldsymbol{x}_2$ then is:

$$\boldsymbol{\Omega}^{-1} = -V_0\left((k_1+k_2)\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tau_m}\right)^{-1}\left(\gamma\boldsymbol{I} - \frac{1-\tilde{w}}{\tilde{w}\tau_m^2}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right) + \frac{k_3-k_2}{\tau_m}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right) \tag{C.12}$$

if we consider:

$$\begin{cases} \boldsymbol{A} = -\frac{1}{\tau_m}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1} \\ \boldsymbol{B} = -\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tau_m}\right)^{-1}\left(\gamma\boldsymbol{I} - \frac{1-\tilde{w}}{\tilde{w}\tau_m^2}\left(\boldsymbol{D} + \frac{\boldsymbol{I}}{\tilde{w}\tau_m}\right)^{-1}\right) \end{cases}$$

$$\boldsymbol{\Omega} = V_0^{-1}\left((k_1+k_2)\boldsymbol{B} + (k_3-k_2)\boldsymbol{A}\right)^{-1} \tag{C.13}$$

The condition number of $\Omega$ is $cond(\boldsymbol{\Omega}) \geqslant \frac{|\boldsymbol{\Omega}^T\boldsymbol{\Omega}|_{Max}}{|\boldsymbol{\Omega}^T\boldsymbol{\Omega}|_{Min}} = 2.58$

If we consider the non-linear equation of the hemodynamic or BOLD model:

$$\boldsymbol{\Omega} = V_0^{-1}\left(k_1\boldsymbol{B} + k_2(\boldsymbol{B}-\boldsymbol{A})(\boldsymbol{I}-\boldsymbol{A})^{-1} + k_3\boldsymbol{A}\right)^{-1} \tag{C.14}$$

# D Markov Chain Monte Carlo solution for ASL JDE

Here we develop the inference using the following priors for $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\boldsymbol{h} \sim \mathcal{N}(0, v_h \boldsymbol{\Sigma}_h) \tag{D.1}$$

$$\text{and } \boldsymbol{g}|\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{\Omega h}, v_g \boldsymbol{\Sigma}_g) \tag{D.2}$$

## D.1 Likelihood

$$\text{Let } \boldsymbol{r}_j = \boldsymbol{y}_j - \sum_{m=1}^{M} \left( c_j^m \boldsymbol{W} \boldsymbol{X}^m \boldsymbol{g} + a_j^m \boldsymbol{X}^m \boldsymbol{h} \right) - \boldsymbol{P} \boldsymbol{\ell}_j - \alpha_j \boldsymbol{w} = \boldsymbol{b}_j \sim \mathcal{N}(0, v_b I_N)$$

$$\text{The likelihood reads } p(\boldsymbol{y} \,|\, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\ell}, \boldsymbol{\alpha}, v_b) \propto \prod_{j=1}^{J} \left( v_{b_j}^{-N/2} \right) \exp\left( \sum_{j=1}^{J} \frac{-1}{2} \boldsymbol{r}_j^t (v_{b_j} I_N)^{-1} \boldsymbol{r}_j \right) \tag{D.3}$$

where $J$ is the number of voxels, and $\boldsymbol{b}_j$ is the noise. We consider white noise.

## D.2 Joint posterior density

$$\begin{aligned} p(\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{q}, \boldsymbol{\theta} \,|\, \boldsymbol{y}) \propto\; & v_h^{-\frac{D+1}{2}} v_g^{-\frac{D+1}{2}} |\boldsymbol{\Sigma}_h|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_g|^{-\frac{1}{2}} v_\ell^{(-JF-1)/2} v_\alpha^{-J/2-1/2} \prod_{j=1}^{J} \left( v_{b_j}^{(-N-1)/2} \right) \dots \\ & \dots \exp\left( -\frac{\boldsymbol{h}^t \boldsymbol{\Sigma}_h^{-1} \boldsymbol{h}}{2 v_h} - \frac{(\boldsymbol{g} - \Omega \boldsymbol{h})^t \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{g} - \Omega \boldsymbol{h})}{2 v_g} + \sum_{j=1}^{J} \left( -\frac{||\boldsymbol{r}_j||^2}{2 v_{b_j}} - \frac{\alpha_j^2}{2 v_\alpha} - \frac{||\boldsymbol{\ell}_j||^2}{2 v_\ell} \right) \right) \dots \\ & \dots \prod_m \left[ \prod_j \sum_i \left( \mathbb{1}(q_j^m = i) (v_i^m)^{-1/2} (\rho_i^m)^{-1/2} \exp\left( -\frac{(a_j^m - \mu_i^m)^2}{2 v_i^m} - \frac{(c_j^m - \eta_i^m)^2}{2 \rho_i^m} \right) \right) \dots \right. \\ & \left. \dots Z(\beta^m)^{-1} \exp\left( \beta^m \sum_{j \sim j'} \mathbb{1}(q_j^m = q_{j'}^m) \right) \right] \prod_m \prod_i \left[ (v_i^m)^{-1/2} (\rho_i^m)^{-1/2} \right] \frac{2}{3} \mathbb{1}(0 \leqslant \beta \leqslant \beta_{max}) \end{aligned} \tag{D.4}$$

## D.3 Gibbs sampling

To sample the posterior of interest, each variable $x \in \mathcal{X}, (\mathcal{X} = \{\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{\theta}\})$ is sampled using a hybrid Metropolis-Gibbs sampling scheme and posterior mean estimates are computed after a burn in period.

### D.3.1 Partial residual quantities

The following quantities will be used afterwards:

$$\tilde{\boldsymbol{y}}_j = \boldsymbol{y}_j - P\boldsymbol{\ell}_j - \alpha_j \boldsymbol{w} \tag{D.5}$$

$$\widehat{\boldsymbol{y}}_j = \tilde{\boldsymbol{y}}_j - \sum_m^M c_j^m \boldsymbol{W}\boldsymbol{X}^m\boldsymbol{g} \qquad \widehat{\boldsymbol{y}}_j^{\setminus m} = \tilde{\boldsymbol{y}}_j - \sum_m^M c_j^m \boldsymbol{W}\boldsymbol{X}^m\boldsymbol{g} - \sum_{m'\neq m}^M a_j^{m'}\boldsymbol{X}^{m'}\boldsymbol{h} \tag{D.6}$$

$$\bar{\boldsymbol{y}}_j = \tilde{\boldsymbol{y}}_j - \sum_m^M a_j^m \boldsymbol{X}^m\boldsymbol{h} \qquad \bar{\boldsymbol{y}}_j^{\setminus m} = \tilde{\boldsymbol{y}}_j - \sum_m^M a_j^m \boldsymbol{X}^m\boldsymbol{h} - \sum_{m'\neq m}^M c_j^{m'}\boldsymbol{W}\boldsymbol{X}^{m'}\boldsymbol{g} \tag{D.7}$$

### D.3.2 BRL and PRL conditional posterior

$$p(\boldsymbol{a}\,|\,\boldsymbol{y},\boldsymbol{c},\boldsymbol{q},\boldsymbol{h},\boldsymbol{g},\boldsymbol{\alpha},\boldsymbol{\ell},\boldsymbol{\theta}) = \prod_j p(\boldsymbol{a}_j\,|\,\boldsymbol{y}_j,\boldsymbol{c}_j,\boldsymbol{q}_j,\boldsymbol{h},\boldsymbol{g},\boldsymbol{\alpha}_j,\boldsymbol{\ell}_j,\boldsymbol{\theta}) \tag{D.8}$$

From the joint posterior, we can derive

$$p(\boldsymbol{a}_j\,|\,\boldsymbol{y}_j,\boldsymbol{c}_j,\boldsymbol{q}_j,\boldsymbol{h},\boldsymbol{g},\boldsymbol{\alpha}_j,\boldsymbol{\ell}_j,\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\left(\frac{\|\boldsymbol{r}_j\|^2}{v_{b_j}} + \sum_m \frac{(a_j^m-\mu_i^m)^2}{v_i^m}\right)\right\} \quad \text{where } q_j^m = i \tag{D.9}$$

$$\text{Let} \quad \|\boldsymbol{r}_j\|^2 = \|\widehat{\boldsymbol{y}}_j^{\setminus m} - a_j^m\boldsymbol{X}^m\boldsymbol{h}\|^2 = \|\widehat{\boldsymbol{y}}_j^{\setminus m}\|^2 - 2a_j^m(\boldsymbol{X}^m\boldsymbol{h})^T\widehat{\boldsymbol{y}}_j^{\setminus m} + (a_j^m)^2\|\boldsymbol{X}^m\boldsymbol{h}\|^2$$

$$\text{and} \quad \sum_{m=1}^M \frac{(a_j^m-\mu_i^m)^2}{v_i^m} = \frac{(a_j^m-\mu_i^m)^2}{v_i^m} + \sum_{m'\neq m}\frac{(a_j^{m'}-\mu_i^{m'})^2}{v_i^{m'}}$$

Thus

$$p(\boldsymbol{a}_j\,|\,\boldsymbol{y}_j,\ldots,\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\left(\frac{\|\widehat{\boldsymbol{y}}_j^{\setminus m}\|^2}{v_{b_j}} - \frac{2a_j^m(\boldsymbol{X}^m\boldsymbol{h})^T\widehat{\boldsymbol{y}}_j^{\setminus m}}{v_{b_j}} + \frac{(a_j^m)^2\|\boldsymbol{X}^m\boldsymbol{h}\|^2}{v_{b_j}} + \frac{(a_j^m-\mu_i^m)^2}{v_i^m} + \sum_{m'\neq m}\frac{(a_j^{m'}-\mu_i^{m'})^2}{v_i^{m'}}\right)\right\}$$

The expression is factorized over voxels, but still not over conditions. By considering $a_j^m$ conditioned to knowing $a_j^{m'\neq m}$, we can ignore the cases when condition $m' \neq m$ and factorize this expression as:

$$\begin{aligned}
p(a_j^m\,|\,a_j^{m'\neq m}\boldsymbol{y}_j,\ldots,\boldsymbol{\theta}) &\propto \exp\left\{-\frac{1}{2}\left(-\frac{2a_j^m(\boldsymbol{X}^m\boldsymbol{h})^T\widehat{\boldsymbol{y}}_j^{\setminus m}}{v_{b_j}} + \frac{(a_j^m)^2\|\boldsymbol{X}^m\boldsymbol{h}\|^2}{v_{b_j}} + \frac{(a_j^m-\mu_i^m)^2}{v_i^m}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(-\frac{2a_j^m(\boldsymbol{X}^m\boldsymbol{h})^T\widehat{\boldsymbol{y}}_j^{\setminus m}}{v_{b_j}} + \frac{(a_j^m)^2\|\boldsymbol{X}^m\boldsymbol{h}\|^2}{v_{b_j}} + \frac{(a_j^m)^2}{v_i^m} - \frac{2a_j^m\mu_i^m}{v_i^m}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[(a_j^m)^2\left(\frac{\|\boldsymbol{X}^m\boldsymbol{h}\|^2}{v_{b_j}} + \frac{1}{v_i^m}\right) - 2a_j^m\left(\frac{\mu_i^m}{v_i^m} + \frac{(\boldsymbol{X}^m\boldsymbol{h})^T\widehat{\boldsymbol{y}}_j^{\setminus m}}{v_{b_j}}\right) + C\right]\right\}
\end{aligned}$$

$$\text{Let} \quad \tilde{\boldsymbol{\Sigma}}_a^{-1} = \left(\frac{\|\boldsymbol{X}^m\boldsymbol{h}\|^2}{v_{b_j}} + \frac{1}{v_i^m}\right), \text{ so } \tilde{\boldsymbol{\Sigma}}_a = \frac{v_{b_j}v_i^m}{\|\boldsymbol{X}^m\boldsymbol{h}\|^2 v_i^m + v_{b_j}}.$$

$$\text{Let} \quad \tilde{\boldsymbol{m}}_a = \tilde{\boldsymbol{\Sigma}}_a\left(\frac{\mu_i^m}{v_i^m} + \frac{(\boldsymbol{X}^m\boldsymbol{h})^T\widehat{\boldsymbol{y}}_j^{\setminus m}}{v_{b_j}}\right) = \tilde{\boldsymbol{\Sigma}}_a\left(\frac{(\boldsymbol{X}^m\boldsymbol{h})^T\widehat{\boldsymbol{y}}_j^{\setminus m}v_i^m + \mu_i^m v_{b_j}}{v_i^m v_{b_j}}\right).$$

We can identify this expression to a posterior Gaussian density:

$$p(a_j^m\,|\,q_j^m = i,\ldots) \sim \mathcal{N}\left(\frac{\boldsymbol{\epsilon}_m^T\widehat{\boldsymbol{y}}_j^{\setminus m}v_i^m + v_{b_j}\mu_i^m}{(v_{i,j}^m)^{-1}v_{b_j}v_i^m}, v_{i,j}^m\right), \text{with } \boldsymbol{\epsilon}_m = \boldsymbol{X}^m\boldsymbol{h} \text{ and } v_{i,j}^m = \frac{v_{b_j}v_i^m}{\boldsymbol{\epsilon}_m^T\boldsymbol{\epsilon}_m v_i^m + v_{b_j}} \tag{D.10}$$

Similarly,

$$p(\boldsymbol{c}\,|\,\boldsymbol{y},\boldsymbol{a},\boldsymbol{q},\boldsymbol{h},\boldsymbol{g},\boldsymbol{\alpha},\boldsymbol{\ell},\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\sum_j\left(\frac{\|\tilde{\boldsymbol{y}}_j\|^2}{v_{b_j}}+\sum_m\frac{(a_j^m-\eta_i^m)^2}{2\rho_i^m}\right)\right\} \quad \text{where } q_j^m=i \tag{D.11}$$

$$p(c_j^m\,|\,q_j^m=i,\ldots)\sim\mathcal{N}\left(\frac{\boldsymbol{\epsilon}'_m{}^T\bar{\boldsymbol{y}}_j^{\setminus m}\rho_i^m+v_{b_j}\eta_i^m}{(\rho_{i,j}^m)^{-1}v_{b_j}\rho_i^m},\rho_{i,j}^m\right), \text{with } \boldsymbol{\epsilon}'_m=\boldsymbol{W}\boldsymbol{X}^m\boldsymbol{g} \text{ and } \rho_{i,j}^m=\frac{v_{b_j}\rho_i^m}{\boldsymbol{\epsilon}'_m{}^T\boldsymbol{\epsilon}'_m\rho_i^m+v_{b_j}} \tag{D.12}$$

### D.3.3 Labels conditional posterior

$$p(q_j^m\,|\,\boldsymbol{y},\boldsymbol{a},\boldsymbol{c},\boldsymbol{h},\boldsymbol{g},\boldsymbol{\alpha},\boldsymbol{\ell},\boldsymbol{\theta}) \propto \prod_m\left[\prod_j\sum_i\left(\mathbb{1}(q_j^m=i)(v_i^m)^{-1/2}(\rho_i^m)^{-1/2}\ldots\right.\right.$$
$$\left.\left.\ldots\exp\left(-\frac{(a_j^m-\mu_i^m)^2}{2v_i^m}-\frac{(c_j^m-\eta_i^m)^2}{2\rho_i^m}\right)\right)\exp\left(\beta^m\sum_{j\sim j'}\mathbb{1}(q_j^m=q_{j'}^m)\right)\right] \tag{D.13}$$

The conditional posterior of $\boldsymbol{q}^m$ is identified to an asymmetric Ising field. Considering $N(j)$ the neighbors of $j$:

$$p(q_j^m\,|\,\boldsymbol{y},\ldots,\boldsymbol{\theta}) \propto \exp\Big[\beta^m\sum_{k\in N(j)}\mathbb{1}(q_j^m=q_k^m)-\sum_{i=0}^{1}\mathbb{1}(q_j^m=i)\Big(\frac{\log(v_i^m\rho_i^m)}{2}+\frac{(a_j^m-\mu_i^m)^2}{2v_i^m}+\frac{(c_j^m-\eta_i^m)^2}{2\rho_i^m}\Big)\Big] \tag{D.14}$$

### D.3.4 BRF conditional posterior

$$p(\boldsymbol{h}\,|\,\boldsymbol{y},\boldsymbol{a},\boldsymbol{c},\boldsymbol{q},\boldsymbol{g},\boldsymbol{\alpha},\boldsymbol{\ell},\boldsymbol{\theta}) \propto \exp\left(\frac{-\boldsymbol{h}^t\boldsymbol{\Sigma}_h^{-1}\boldsymbol{h}}{2v_h}-\frac{(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})^t\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})}{2v_g}-\sum_{j=1}^{J}\frac{||\boldsymbol{r}_j||^2}{2v_{b_j}}\right) \tag{D.15}$$

Let $||\boldsymbol{r}_j||^2=||\widehat{\boldsymbol{y}}_j-\boldsymbol{S}_j\boldsymbol{h}||=(\widehat{\boldsymbol{y}}_j-\boldsymbol{S}_j\boldsymbol{h})^t(\widehat{\boldsymbol{y}}_j-\boldsymbol{S}_j\boldsymbol{h})=(\boldsymbol{S}_j\boldsymbol{h})^t(\boldsymbol{S}_j\boldsymbol{h})-(\boldsymbol{S}_j\boldsymbol{h})^t\widehat{\boldsymbol{y}}_j-\widehat{\boldsymbol{y}}_j^t(\boldsymbol{S}_j\boldsymbol{h})+||\widehat{\boldsymbol{y}}_j||^2$,
where $\boldsymbol{S}_j=\sum_{m=1}^M a_j^m\boldsymbol{X}^m$. Then

$$\propto\exp\left(\frac{-1}{2}\left[\frac{\boldsymbol{h}^t\boldsymbol{\Sigma}_h^{-1}\boldsymbol{h}}{v_h}+\frac{(\boldsymbol{h}^t\boldsymbol{\Omega}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\boldsymbol{h}-(\boldsymbol{\Omega}\boldsymbol{h})^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g}-\boldsymbol{g}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\boldsymbol{h}+\boldsymbol{g}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g})}{v_g}\right.\right.$$
$$\left.\left.+\sum_{j=1}^{J}\frac{(\boldsymbol{h}^t(\boldsymbol{S}_j^t\boldsymbol{S}_j)\boldsymbol{h}-(\boldsymbol{S}_j\boldsymbol{h})^t\widehat{\boldsymbol{y}}_j-\widehat{\boldsymbol{y}}_j^t(\boldsymbol{S}_j\boldsymbol{h})+||\widehat{\boldsymbol{y}}_j||^2)}{v_{b_j}}\right]\right)$$
$$\propto\exp\left(\frac{-1}{2}\left[\boldsymbol{h}^t\left(v_h^{-1}\boldsymbol{\Sigma}_h^{-1}+\boldsymbol{\Omega}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}v_g^{-1}+\sum_{j=1}^{J}(\boldsymbol{S}_j^t\boldsymbol{S}_j)v_{b_j}^{-1}\right)\boldsymbol{h}-2\boldsymbol{h}^t\left(\boldsymbol{\Omega}^t\boldsymbol{\Sigma}_g^{-1}v_g^{-1}\boldsymbol{g}+\sum_{j=1}^{J}\boldsymbol{S}_j^t\widehat{\boldsymbol{y}}_jv_{b_j}^{-1}\right)\right]\right)$$
$$\propto\exp\left(\frac{-1}{2}\left[\boldsymbol{h}^t\tilde{\boldsymbol{\Sigma}}_h^{-1}\boldsymbol{h}-2\boldsymbol{h}^t\left(\boldsymbol{\Omega}^t\boldsymbol{\Sigma}_g^{-1}v_g^{-1}\boldsymbol{g}+\sum_{j=1}^{J}\boldsymbol{S}_j^t\widehat{\boldsymbol{y}}_jv_{b_j}^{-1}\right)\right]\right)$$

Thus $p(\boldsymbol{h}|\boldsymbol{y},\ldots,\boldsymbol{\theta})$ can be identified to a posterior Gaussian density: $\mathcal{N}(\tilde{\boldsymbol{m}}_h,\tilde{\boldsymbol{\Sigma}}_h)$, where

$$\tilde{\boldsymbol{\Sigma}}_h=\left(v_h^{-1}\boldsymbol{\Sigma}_h^{-1}+\boldsymbol{\Omega}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}v_g^{-1}+\sum_{j=1}^{J}(\boldsymbol{S}_j^t\boldsymbol{S}_j)v_{b_j}^{-1}\right)^{-1} \tag{D.16}$$

$$\tilde{\boldsymbol{m}}_h=\tilde{\boldsymbol{\Sigma}}_h\left(\boldsymbol{\Omega}^t\boldsymbol{\Sigma}_g^{-1}v_g^{-1}\boldsymbol{g}+\sum_{j=1}^{J}\boldsymbol{S}_j^t\widehat{\boldsymbol{y}}_jv_{b_j}^{-1}\right) \tag{D.17}$$

### D.3.5 PRF conditional posterior

$$p(\boldsymbol{g} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{q}, \boldsymbol{h}, \boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{\theta}) \propto \exp\left(\frac{-(\boldsymbol{g}-\Omega\boldsymbol{h})^t \boldsymbol{\Sigma}_g^{-1}(\boldsymbol{g}-\Omega\boldsymbol{h})}{2v_g} - \sum_{j=1}^{J} \frac{||\boldsymbol{r}_j||^2}{2v_{b_j}}\right) \tag{D.18}$$

Let $||\boldsymbol{r}_j||^2 = ||\bar{\boldsymbol{y}}_j - \check{\boldsymbol{S}}_j\boldsymbol{g}|| = (\bar{\boldsymbol{y}}_j - \check{\boldsymbol{S}}_j\boldsymbol{g})^t(\bar{\boldsymbol{y}}_j - \check{\boldsymbol{S}}_j\boldsymbol{g}) = (\check{\boldsymbol{S}}_j\boldsymbol{g})^t(\check{\boldsymbol{S}}_j\boldsymbol{g}) - (\check{\boldsymbol{S}}_j\boldsymbol{g})^t\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{y}}_j^t(\check{\boldsymbol{S}}_j\boldsymbol{g}) + ||\bar{\boldsymbol{y}}_j||^2$,
where $\check{\boldsymbol{S}}_j = \sum_{m=1}^{M} c_j^m \boldsymbol{W}\boldsymbol{X}^m$

$$\begin{aligned}
&\propto \exp\left(\frac{-1}{2}\left[\frac{(\boldsymbol{h}^t\boldsymbol{\Omega}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\boldsymbol{h} - (\boldsymbol{\Omega}\boldsymbol{h})^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g} - \boldsymbol{g}^t\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\Omega}\boldsymbol{h}) + \boldsymbol{g}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g})}{v_g}\right.\right.\\
&\qquad \left.\left. + \sum_{j=1}^{J} \frac{(\boldsymbol{g}^t(\check{\boldsymbol{S}}_j^t\check{\boldsymbol{S}}_j)\boldsymbol{g} + (\check{\boldsymbol{S}}_j\boldsymbol{g})^t\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{y}}_j^t(\check{\boldsymbol{S}}_j\boldsymbol{g}) + ||\bar{\boldsymbol{y}}_j||^2)}{v_{b_j}}\right]\right)\\
&\propto \exp\left(\frac{-1}{2}\left[\boldsymbol{g}^t\left(v_g^{-1}\boldsymbol{\Sigma}_g^{-1} + \sum_{j=1}^{J}(\check{\boldsymbol{S}}_j^t\check{\boldsymbol{S}}_j)v_{b_j}^{-1}\right)\boldsymbol{g} - \boldsymbol{g}^t\left(\boldsymbol{\Sigma}_g^{-1}v_g^{-1}\Omega\boldsymbol{h} + \sum_{j=1}^{J} v_{b_j}^{-1}\check{\boldsymbol{S}}_j^t\bar{\boldsymbol{y}}_j\right)\right]\right)\\
&\propto \exp\left(\frac{-1}{2}\left[\left(\boldsymbol{g} - \tilde{\boldsymbol{\Sigma}}_g\left(\boldsymbol{\Sigma}_g^{-1}v_g^{-1}\Omega\boldsymbol{h} + \sum_{j=1}^{J} v_{b_j}^{-1}\check{\boldsymbol{S}}_j^t\bar{\boldsymbol{y}}_j\right)\right)^t \tilde{\boldsymbol{\Sigma}}_g^{-1}\left(\boldsymbol{g} - \tilde{\boldsymbol{\Sigma}}_g\left(\boldsymbol{\Sigma}_g^{-1}v_g^{-1}\Omega\boldsymbol{h} + \sum_{j=1}^{J} v_{b_j}^{-1}\check{\boldsymbol{S}}_j^t\bar{\boldsymbol{y}}_j\right)\right)\right]\right)
\end{aligned}$$

Thus $p(\boldsymbol{g}|\boldsymbol{y}, ..., \boldsymbol{\theta})$ can be identified to a posterior Gaussian density: $\mathcal{N}(\tilde{\boldsymbol{m}}_g, \tilde{\boldsymbol{\Sigma}}_g)$, where

$$\tilde{\boldsymbol{\Sigma}}_g = \left(v_g^{-1}\boldsymbol{\Sigma}_g^{-1} + \sum_{j=1}^{J}(\check{\boldsymbol{S}}_j^t\check{\boldsymbol{S}}_j)v_{b_j}^{-1}\right)^{-1} \tag{D.19}$$

$$\tilde{\boldsymbol{m}}_g = \boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}_g^{-1}v_g^{-1}\Omega\boldsymbol{h} + \sum_{j=1}^{J} v_{b_j}^{-1}\check{\boldsymbol{S}}_j^t\bar{\boldsymbol{y}}_j\right). \tag{D.20}$$

### D.3.6 Perfusion Baseline conditional posterior

$$p(\boldsymbol{\alpha} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{q}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\ell}, \boldsymbol{\theta}) \propto \prod_{j=1}^{J} \exp\left(\frac{-\alpha_j^2}{2v_\alpha} - \frac{||\boldsymbol{r}_j||^2}{2v_{b_j}}\right) \tag{D.21}$$

Let $||\boldsymbol{r}_j||^2 = ||\boldsymbol{z}_\alpha - \alpha_j\boldsymbol{w}||^2 = (\boldsymbol{z}_\alpha - \alpha_j\boldsymbol{w})^t(\boldsymbol{z}_\alpha - \alpha_j\boldsymbol{w}) = \alpha_j^2 N - 2\alpha_j\boldsymbol{w}^t\boldsymbol{z}_\alpha + ||\boldsymbol{z}_\alpha^2||$
where $\boldsymbol{z}_\alpha = \boldsymbol{y}_j - \sum_{m=1}^{M}\left[c_j^m\boldsymbol{W}\boldsymbol{X}^m\boldsymbol{g} + a_j^m\boldsymbol{X}^m\boldsymbol{h}\right] - \boldsymbol{P}\boldsymbol{\ell}_j$ is constant due to the given information.

$$\propto \prod_{j=1}^{J} \exp\left(\frac{-\alpha_j^2}{2v_\alpha} - \frac{\alpha_j^2 N - 2\alpha_j\boldsymbol{w}^t\boldsymbol{z}_\alpha + ||\boldsymbol{z}_\alpha||^2}{2v_{b_j}}\right) \propto \prod_{j=1}^{J} \exp\left(\frac{-1}{2}\,\frac{\alpha_j^2(v_{b_j} + Nv_\alpha) - \alpha_j\boldsymbol{w}^t\boldsymbol{z}_\alpha v_\alpha + ||\boldsymbol{z}||^2 v_\alpha}{v_\alpha v_{b_j}}\right)$$

Let $\sigma_\alpha^{-2} = \frac{v_{b_j} + Nv_\alpha}{v_{b_j}v_\alpha}$

$$\propto \prod_{j=1}^{J} \exp\left(\frac{-1}{2}\sigma_\alpha^{-2}(\alpha_j^2 + \frac{\alpha_j 2\boldsymbol{w}^t\boldsymbol{z}_\alpha}{v_{b_j}}\sigma_\alpha^2)\right) \propto \prod_{j=1}^{J} \exp\left(\frac{-1}{2}\sigma_\alpha^{-2}(\alpha_j + \frac{\boldsymbol{w}^t\boldsymbol{z}_\alpha}{v_{b_j}}\sigma_\alpha^2)^2\right) \propto \prod_{j=1}^{J} \exp\left(\frac{-(\alpha_j - \mu_\alpha)^2}{2\sigma_\alpha^2}\right)$$

Letting $\mu_\alpha = \frac{\boldsymbol{w}^t\boldsymbol{z}_\alpha}{v_{b_j}}\sigma_\alpha^2 = \frac{\boldsymbol{w}^t\boldsymbol{z}_\alpha v_\alpha}{v_{b_j} + Nv_\alpha}$. Thus $p(\boldsymbol{\alpha}|\boldsymbol{y}, ..., \boldsymbol{\theta})$ can be identified to a posterior Gaussian density:

$$\alpha_j \propto \mathcal{N}\left(\frac{\boldsymbol{w}^t\boldsymbol{z}_\alpha v_\alpha}{v_{b_j} + Nv_\alpha}, \frac{v_\alpha v_{b_j}}{v_{b_j} + Nv_\alpha}\right) \tag{D.22}$$

### D.3.7 Drift Coefficients conditional posterior

$$p(\boldsymbol{\ell} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{q}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\theta}) \propto \prod_{j=1}^{J} \exp\left(\frac{-||\boldsymbol{\ell}_j||^2}{2v_{\boldsymbol{\ell}}} - \frac{||\boldsymbol{r}_j||^2}{2v_{b_j}}\right) \tag{D.23}$$

Let $||\boldsymbol{r}_j||^2 = ||\boldsymbol{z}_{\ell_j} - \boldsymbol{P}\boldsymbol{\ell}_j||^2 = (\boldsymbol{z}_{\ell_j} - \boldsymbol{P}\boldsymbol{\ell}_j)^t(\boldsymbol{z}_{\ell_j} - \boldsymbol{P}\boldsymbol{\ell}_j) = (\boldsymbol{P}\boldsymbol{\ell}_j)^t(\boldsymbol{P}\boldsymbol{\ell}_j) - (\boldsymbol{P}\boldsymbol{\ell}_j)^t\boldsymbol{z}_{\ell_j} - \boldsymbol{z}_{\ell_j}^t(\boldsymbol{P}\boldsymbol{\ell}_j) + ||\boldsymbol{z}_{\ell_j}^2||$
where $\boldsymbol{z}_{\ell_j} = \boldsymbol{y}_j - \sum_{m=1}^{M}\left[c_j^m \boldsymbol{W}\boldsymbol{X}^m\boldsymbol{g} + a_j^m \boldsymbol{X}^m\boldsymbol{h}\right] - \alpha_j\boldsymbol{w}$ is constant due to the given information.

$$\propto \prod_{j=1}^{J} \exp\left(\frac{-1}{2}\left[\boldsymbol{\ell}_j^t(I_F v_{\boldsymbol{\ell}}^{-1} + \boldsymbol{P}^t\boldsymbol{P}v_{b_j}^{-1})\boldsymbol{\ell}_j - (\boldsymbol{P}\boldsymbol{\ell}_j)^t\boldsymbol{z}_{\ell_j}v_{b_j}^{-1} - \boldsymbol{z}_{\boldsymbol{\ell}_j}^t(\boldsymbol{P}\boldsymbol{\ell}_j)v_{b_j}^{-1}\right]\right)$$

Let $\Sigma_\ell^{-1} = (v_{\boldsymbol{\ell}}^{-1} + v_{b_j}^{-1})$. Since $\boldsymbol{P}$ is an orthonormal basis, $\boldsymbol{P}^t\boldsymbol{P} = I_F$.

$$\propto \prod_{j=1}^{J} \exp\left(\frac{-1}{2}\,(\boldsymbol{\ell}_j - \Sigma_\ell\boldsymbol{P}^t\boldsymbol{z}_{\ell_j}v_{b_j}^{-1})^t\Sigma_\ell^{-1}(\boldsymbol{\ell}_j - \Sigma_\ell\boldsymbol{P}^t\boldsymbol{z}_{\ell_j}v_{b_j}^{-1})\right)$$

Let $\mu_\ell = (\Sigma_\ell\boldsymbol{P}^t\boldsymbol{z}_{\ell_j}v_{b_j}^{-1})$. Thus $p(\boldsymbol{\ell}|\boldsymbol{y}, ..., \boldsymbol{\theta})$ can be identified to a posterior Gaussian density:

$$\boldsymbol{\ell}_j \propto \mathcal{N}\left(\Sigma_\ell\boldsymbol{P}^t\boldsymbol{z}_{\ell_j}v_{b_j}^{-1},\ I_F(v_{\boldsymbol{\ell}}^{-1} + v_{b_j}^{-1})^{-1}\right) \tag{D.24}$$

Vectors $\boldsymbol{\ell}_j$ are independent. As long as there is posterior independence, ie the expression is a product over sub-variables. The sub-variables can be sampled in parallel.

### D.3.8 Hyperparameter conditional posterior

Let $\widetilde{\boldsymbol{\theta}}$ be the set of all hyperparameters except the one in question.

#### Mixture model parameters for the BOLD component $\theta_{\boldsymbol{a}} = (\mu_{i,m}, v_{i,m})$

$$p(\mu_{i,m} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{q}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\ell}, \widetilde{\boldsymbol{\theta}}) \propto \prod_{j=1}^{J}\left[\mathbb{1}(q_j^m = i)\exp\left(-\frac{(a_j^m - \mu_{i,m})^2}{2v_{i,m}}\right)\right] \tag{D.25}$$

Let $J_i$ denote a set of vowels belonging to class i, and $|J_i|$ denote the number of voxels belonging to class i.

$$\propto \prod_{j=1}^{J}\left[\mathbb{1}(q_j^m = i)\exp\left(-\frac{1}{2}\left(\frac{\mu_{i,m}^2}{v_{i,m}} - \frac{2\mu_{i,m}a_j^m}{v_{i,m}}\right)\right)\right] \propto \exp\left(-\frac{1}{2}\left(\frac{|J_i|\mu_{i,m}^2}{v_{i,m}} - \frac{2|J_i|\mu_{i,m}\sum_{j\in J_i}a_j^m}{v_{i,m}}\right)\right)$$

$$\text{Let } \sigma_\mu^{-2} = \frac{|J_i|}{v_{i,m}},\ \ \mu_\mu = \sigma_\mu^2\left(\frac{|J_i|\sum_{j\in J_i}a_j^m}{v_{i,m}}\right) = \sum_{j\in J_i}a_j^m$$

Thus, the posterior conditional of $\mu_{i,m}$ can be identified to:

$$p(\mu_{i,m}|\boldsymbol{y}, ..., \widetilde{\boldsymbol{\theta}}) \propto \mathcal{N}\left(\sum_{j\in J_i}a_j^m,\ \frac{v_{i,m}}{|J_i|}\right) \tag{D.26}$$

$$p(v_{i,m} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{q}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\ell}, \widetilde{\boldsymbol{\theta}}) \propto v_{i,m}^{-1/2}\prod_{j=1}^{J}\left[\mathbb{1}(q_j^m = i)v_{i,m}^{-1/2}\exp\left(-\frac{(a_j^m - \mu_{i,m})^2}{2v_{i,m}}\right)\right]$$

$$\propto v_{i,m}^{(-|J_i|-1)/2}\exp\left(-\frac{\sum_{j\in J_i}(a_j^m - \mu_{i,m})^2}{2v_{i,m}}\right)$$

Thus, the posterior conditional of $v_{i,m}$ can be identified to an inverse Gamma function $\mathcal{IG}$:

$$p(v_{i,m}|\boldsymbol{y}, ..., \widetilde{\boldsymbol{\theta}}) \propto \mathcal{IG}\left(\frac{|J_i| - 1}{2},\ \sum_{j\in J_i}\frac{1}{2}(a_j^m - \mu_{i,m})^2\right) \tag{D.27}$$

### Mixture model parameters for the perfusion component $\theta_c = (\eta_{i,m}, \rho_{i,m})$

Following the same steps outlined in for $\boldsymbol{\theta_a}$,

$$p(\eta_{i,m} \mid \boldsymbol{y,a,c,q,h,g,\alpha,\ell},\widetilde{\boldsymbol{\theta}}) \propto \prod_{j=1}^{J}\left[\mathbb{1}(q_j^m = i)\exp\left(-\frac{(c_j^m-\eta_{i,m})^2}{2\rho_{i,m}}\right)\right] \tag{D.28}$$

$$p(\rho_{i,m} \mid \boldsymbol{y,a,c,q,h,g,\alpha,\ell},\widetilde{\boldsymbol{\theta}}) \propto \rho_{i,m}^{-1/2}\prod_{j=1}^{J}\left[\mathbb{1}(q_j^m = i)\rho_{i,m}^{-1/2}\exp\left(-\frac{(c_j^m-\eta_{i,m})^2}{2\rho_{i,m}}\right)\right] \tag{D.29}$$

Thus, the posterior conditional of $\eta_{i,m}$ and $\rho_{i,m}$ can be identified to:

$$p(\eta_{i,m}|\boldsymbol{y},...,\widetilde{\boldsymbol{\theta}}) \propto \mathcal{N}\left(\sum_{j\in J_i} c_j^m,\ \frac{v_{i,m}}{|J_i|}\right) \tag{D.30}$$

$$p(\rho_{i,m}|\boldsymbol{y},...,\widetilde{\boldsymbol{\theta}}) \propto \mathcal{IG}\left(\frac{|J_i|-1}{2},\ \sum_{j\in J_i}\frac{1}{2}(c_j^m-\eta_{i,m})^2\right) \tag{D.31}$$

### HRF and PRF variance $v_h$ and $v_g$

$$p(v_h \mid \boldsymbol{y,a,c,q,h,g,\alpha,\ell},\widetilde{\boldsymbol{\theta}}) \propto v_h^{-(D+1)/2}\exp\left(\frac{-\boldsymbol{h}^t\boldsymbol{\Sigma}_h^{-1}\boldsymbol{h}}{2v_h}\right) \tag{D.32}$$

$$p(v_g \mid \boldsymbol{y,a,c,q,h,g,\alpha,\ell},\widetilde{\boldsymbol{\theta}}) \propto v_g^{-(D+1)/2}\exp\left(\frac{-\boldsymbol{g}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g}}{2v_g}\right) \tag{D.33}$$

Thus, the posterior conditional of $v_h$ and $v_g$ can be identified to:

$$p(v_h|\boldsymbol{y},...,\widetilde{\boldsymbol{\theta}}) \propto \mathcal{IG}\left(\frac{D-1}{2},\ \frac{1}{2}\boldsymbol{h}^t\boldsymbol{\Sigma}_h^{-1}\boldsymbol{h}\right) \tag{D.34}$$

$$p(v_g|\boldsymbol{y},...,\widetilde{\boldsymbol{\theta}}) \propto \mathcal{IG}\left(\frac{D-1}{2},\ \frac{1}{2}\boldsymbol{g}^t\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g}\right) \tag{D.35}$$

### Perfusion baseline variance $v_\alpha$

$$p(v_\alpha \mid \boldsymbol{y,a,c,q,h,g,\alpha,\ell},\widetilde{\boldsymbol{\theta}}) \propto v_\alpha^{-(J+1)/2}\prod_{j=1}^{J}\left[\exp\left(-\frac{\alpha_j^2}{2v_\alpha}\right)\right] \tag{D.36}$$

Thus, the posterior conditional of $v_\ell$ can be identified to:

$$p(v_\alpha|\boldsymbol{y},...,\widetilde{\boldsymbol{\theta}}) \propto \mathcal{IG}\left(\frac{J-1}{2},\ \frac{1}{2}\sum_{j=1}^{J}\alpha_j^2\right) \tag{D.37}$$

### Drift coefficient variance $v_\ell$

$$p(v_\ell \mid \boldsymbol{y,a,c,q,h,g,\alpha,\ell},\widetilde{\boldsymbol{\theta}}) \propto v_{\boldsymbol{\ell}}^{(-JF-1)/2}\prod_{j=1}^{J}\left[\exp\left(-\frac{||\boldsymbol{\ell}_j||^2}{2v_{\boldsymbol{\ell}}}\right)\right] \tag{D.38}$$

Thus, the posterior conditional of $v_\ell$ can be identified to:

$$p(v_{\boldsymbol{\ell}}|\boldsymbol{y},...,\widetilde{\boldsymbol{\theta}}) \propto \mathcal{IG}\left(\frac{JF-1}{2},\ \frac{1}{2}\sum_{j=1}^{J}||\boldsymbol{\ell}_j||^2\right) \tag{D.39}$$

### Noise variance $v_b$

$$p(v_b \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{q}, \boldsymbol{h}, \boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\ell}, \widetilde{\boldsymbol{\theta}}) \propto \prod_{j=1}^{J} \left[ v_{b_j}^{(-N-1)/2} \exp\left( -\frac{||\boldsymbol{r}_j||^2}{2v_{b_j}} \right) \right] \tag{D.40}$$

Thus, the posterior conditional of $v_{b_j}$ can be identified to:

$$p(v_{b_j} | \boldsymbol{y}, ..., \widetilde{\boldsymbol{\theta}}) \propto \mathcal{IG}\left( \frac{N-1}{2}, \frac{1}{2}||\boldsymbol{r}_j||^2 \right) \tag{D.41}$$

# E Variational EM solution for JDE ASL

As introduced in the document, the variational approximation through mean field assumes independence of the probability distributions. The E-step is instead solved over a restricted class of probability distributions that factorize as $\tilde{p}(\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q}) = \tilde{p}_a(\boldsymbol{a})\ \tilde{p}_h(\boldsymbol{h})\ \tilde{p}_c(\boldsymbol{c})\ \tilde{p}_g(\boldsymbol{g})\ \tilde{p}_q(\boldsymbol{q})$. The E-step becomes an approximate E-step that can be further decomposed into five stages. At iteration $(r)$:

$$\textbf{E-A:}\ \ \tilde{p}_a(a) \propto \exp\left(\mathrm{E}_{\tilde{p}_c^{(r-1)}\tilde{p}_q^{(r-1)}\tilde{p}_h^{(r-1)}\tilde{p}_g^{(r-1)}}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{y},\boldsymbol{h}^{(r-1)},\boldsymbol{c}^{(r-1)},\boldsymbol{g}^{(r-1)},\boldsymbol{q}^{(r-1)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.1}$$

$$\textbf{E-C:}\ \ \tilde{p}_c(c) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_q^{(r-1)}\tilde{p}_h^{(r-1)}\tilde{p}_g^{(r-1)}}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{y},\boldsymbol{a}^{(r)},\boldsymbol{h}^{(r-1)},\boldsymbol{g}^{(r-1)},\boldsymbol{q}^{(r-1)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.2}$$

$$\textbf{E-Q:}\ \ \tilde{p}_q^{(r)}(q) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}\tilde{p}_h^{(r-1)}\tilde{p}_g^{(r-1)}}\left[\log p(\boldsymbol{q}\,|\,\boldsymbol{y},\boldsymbol{a}^{(r)},\boldsymbol{h}^{(r-1)},\boldsymbol{c}^{(r)},\boldsymbol{g}^{(r-1)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.3}$$

$$\textbf{E-H:}\ \ \tilde{p}_h(h) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}\tilde{p}_g^{(r-1)}}\left[\log p(\boldsymbol{h}\ |\ \boldsymbol{y},\boldsymbol{a}^{(r)},\boldsymbol{c}^{(r)},\boldsymbol{g}^{(r-1)},\boldsymbol{q}^{(r)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.4}$$

$$\textbf{E-G:}\ \ \tilde{p}_g(g) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}\tilde{p}_h^{(r)}}\left[\log p(\boldsymbol{g}\ |\ \boldsymbol{y},\boldsymbol{a}^{(r)},\boldsymbol{h}^{(r)},\boldsymbol{c}^{(r)},\boldsymbol{q}^{(r)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.5}$$

The corresponding **M-step** reads:

$$\textbf{M:}\ \ \boldsymbol{\theta}^{(r+1)} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\ \mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}\tilde{p}_q^{(r)}\tilde{p}_h^{(r)}\tilde{p}_g^{(r)}}\left[\log p(\boldsymbol{y},\boldsymbol{a}^{(r)},\boldsymbol{h}^{(r)},\boldsymbol{c}^{(r)},\boldsymbol{g}^{(r)},\boldsymbol{q}^{(r)}\ ;\ \boldsymbol{\theta})\right] \tag{E.6}$$

If constraints on the norm of the response functions $\boldsymbol{h}$ and $\boldsymbol{g}$ are imposed, then:

$$\textbf{E-A:}\ \ \tilde{p}_a(a) \propto \exp\left(\mathrm{E}_{\tilde{p}_c^{(r-1)}\tilde{p}_q^{(r-1)}}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{y},\tilde{\boldsymbol{h}}^{(r-1)},\boldsymbol{c}^{(r-1)},\tilde{\boldsymbol{g}}^{(r-1)},\boldsymbol{q}^{(r-1)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.7}$$

$$\textbf{E-C:}\ \ \tilde{p}_c(c) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_q^{(r-1)}}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{y},\boldsymbol{a}^{(r)},\tilde{\boldsymbol{h}}^{(r-1)},\tilde{\boldsymbol{g}}^{(r-1)},\boldsymbol{q}^{(r-1)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.8}$$

$$\textbf{E-Q:}\ \ \tilde{p}_q^{(r)}(q) \propto \exp\left(\mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}}\left[\log p(\boldsymbol{q}\,|\,\boldsymbol{y},\boldsymbol{a}^{(r)},\tilde{\boldsymbol{h}}^{(r-1)},\boldsymbol{c}^{(r)},\tilde{\boldsymbol{g}}^{(r-1)};\boldsymbol{\theta}^{(r)})\right]\right) \tag{E.9}$$

$$\textbf{E-H:}\ \ \tilde{\boldsymbol{h}} = \underset{\boldsymbol{h}\ \text{s.t.}\ \boldsymbol{h}^t\boldsymbol{h}=1}{\arg\max}\ \mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}}\left[\log p(\boldsymbol{h}\ |\ \boldsymbol{y},\boldsymbol{a}^{(r)},\boldsymbol{c}^{(r)},\tilde{\boldsymbol{g}}^{(r-1)},\boldsymbol{q}^{(r)};\boldsymbol{\theta}^{(r)})\right] \tag{E.10}$$

$$\textbf{E-G:}\ \ \tilde{\boldsymbol{g}} = \underset{\boldsymbol{g}\ \text{s.t.}\ \boldsymbol{g}^t\boldsymbol{g}=1}{\arg\max}\ \mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}}\left[\log p(\boldsymbol{g}\ |\ \boldsymbol{y},\boldsymbol{a}^{(r)},\tilde{\boldsymbol{h}}^{(r)},\boldsymbol{c}^{(r)},\boldsymbol{q}^{(r)};\boldsymbol{\theta}^{(r)})\right] \tag{E.11}$$

The corresponding **M-step** reads:

$$\textbf{M:}\ \ \boldsymbol{\theta}^{(r+1)} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\ \mathrm{E}_{\tilde{p}_a^{(r)}\tilde{p}_c^{(r)}\tilde{p}_q^{(r)}}\left[\log p(\boldsymbol{y},\boldsymbol{a}^{(r)},\tilde{\boldsymbol{h}}^{(r)},\boldsymbol{c}^{(r)},\tilde{\boldsymbol{g}}^{(r)},\boldsymbol{q}^{(r)}\ ;\ \boldsymbol{\theta})\right] \tag{E.12}$$

These steps are now given in more details in the next subsections. For simplification, the $(r)$ and $(r-1)$ superscripts are omitted.

## E.1 E-H step

Constant terms in $x$ will be named $C_x$ or $C_{1,x}, C_{2,x}, etc$ when many constants in a formula. Term $\tilde{p}_{q_j^m}(q_j^m = 1)$ will be $\tilde{p}_{q_j^m}(1)$ for simplicity. By convention, vectors are column, and their transpose is $^{\mathsf{t}}$.

Here we denote by $\tilde{\boldsymbol{m}}_{a_j}$ and $\tilde{\boldsymbol{\Sigma}}_{a_j}$ the mean and covariance matrix of $\tilde{p}_{a_j}$ and write $\tilde{m}_{a_j^m}$ for the $m^{th}$ component of vector $\tilde{\boldsymbol{m}}_{a_j}$ ($\tilde{m}_{a_j^m} = \mathrm{E}_{\tilde{p}_{a_j}}\left[a_j^m\right]$) and $\tilde{v}_{a_j^m a_j^{m'}}$ for the $(m, m')$ element of matrix $\tilde{\boldsymbol{\Sigma}}_{a_j}$ ($\tilde{v}_{a_j^m a_j^{m'}} = \mathrm{E}_{\tilde{p}_{a_j}}\left[a_j^m a_j^{m'}\right] - \tilde{m}_{a_j^m}\tilde{m}_{a_j^{m'}}$). We also denote $\boldsymbol{S}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m$ and $\check{\boldsymbol{S}}_j = \sum_{m=1}^{M} c_j^m \boldsymbol{W}\boldsymbol{X}^m$. To simplify, we consider $\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j = \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{w}\alpha_j$ where, for $j \in \mathcal{J}$, $\bar{\boldsymbol{\ell}}_j = \left[\alpha_j, \boldsymbol{\ell}_j\right]^T$ and similarly $\bar{\boldsymbol{\ell}} = [\boldsymbol{\alpha}, \boldsymbol{\ell}]^T$.

$$\begin{aligned}\tilde{p}_{\boldsymbol{h}}(\boldsymbol{h}) &\propto \exp\left(E_{\tilde{p}_a\tilde{p}_c\tilde{p}_g\tilde{p}_q}[\log p(\boldsymbol{h} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\theta})]\right) \\ &\propto \exp\left(E_{\tilde{p}_a\tilde{p}_c\tilde{p}_g}[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\theta})] + \log p(\boldsymbol{h}; v_h) + \log p(\boldsymbol{g}|\boldsymbol{\Omega}\boldsymbol{h}; v_g)\right) \end{aligned} \tag{E.13}$$

**1)** Computing $\log p\left(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\theta}\right)$:
Considering $\tilde{\boldsymbol{y}}_j = \boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j$ and that we consider the terms depending on $\boldsymbol{h}$, we write:

$$\begin{aligned}\log p\left(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\theta}\right) &= C_{1,h} - \frac{1}{2}\sum_{j\in J}\tilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\tilde{\boldsymbol{y}}_j \\ &= C_{2,h} - \frac{1}{2}\sum_{j\in J}\left[\boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\boldsymbol{h} - 2\boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\end{aligned} \tag{E.14}$$

where $\boldsymbol{\Gamma}_j = \frac{1}{v_{b_j}}\boldsymbol{\Lambda}_j$, $C_{1,h} = -\frac{NJ}{2}\log 2\pi + \frac{J}{2}\log|\boldsymbol{\Lambda}_j| - N\sum_{j\in J}\log v_{b_j}$ and $C_{2,h} = C_{1,\boldsymbol{h}} - \frac{1}{2}\sum_{j\in J}(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \boldsymbol{P}\boldsymbol{\ell}_j)$.

**2)** Computing $\log p\left(\boldsymbol{h} \mid v_h\right)$:
Based on *a priori* HRF, $\log p\left(\boldsymbol{h} \mid v_h\right) = C_{3,h} - \frac{1}{2v_h}\boldsymbol{h}^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\boldsymbol{h}$, with $C_{3,h} = -\frac{D+1}{2}\log 2\pi - \frac{D-1}{2}\log v_h - \frac{1}{2}\log|\boldsymbol{\Sigma}_h|$.
**3)** Computing $\log p\left(\boldsymbol{g}|\boldsymbol{\Omega}\boldsymbol{h}; v_g\right)$:
Based on *a priori* PRF, $\log p\left(\boldsymbol{g}|\boldsymbol{\Omega}\boldsymbol{h}; v_g\right) = C_{4,h} - \frac{1}{2v_g}(\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\boldsymbol{h} + \frac{1}{2v_g}2(\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g}$, with $C_{4,h} = -\frac{D+1}{2}\log 2\pi - \frac{D-1}{2}\log v_g - \frac{1}{2}\log|\boldsymbol{\Sigma}_g| - \frac{1}{2v_g}\boldsymbol{g}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g}$.
It follows that:

$$\begin{aligned}\tilde{p}_h(\boldsymbol{h}) \propto \exp\Bigg(-\frac{1}{2}\Bigg[\boldsymbol{h}^{\mathrm{t}}\left(\frac{\boldsymbol{\Sigma}_h^{-1}}{v_h} + \frac{\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}}{v_g} + E_{\tilde{p}_a}\left[\sum_{j\in J}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right]\right)\boldsymbol{h} \\ -2\boldsymbol{h}^{\mathrm{t}}\left(\frac{\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\tilde{\boldsymbol{m}}_g}{v_g} + E_{\tilde{p}_a\tilde{p}_c\tilde{p}_g}\left[\sum_{j\in J}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\right)\Bigg]\Bigg)\end{aligned}$$

$$\begin{aligned}\text{where,}\quad E_{\tilde{p}_a}\left[\sum_{j\in J}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right] &= \sum_{j\in J}\sum_{m,m'}E_{\tilde{p}_{a_j^{m,m'}}}\left[a_j^m a_j^{m'}\right](\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'} \\ &= \sum_{j\in J}\sum_{m,m'}\left(\tilde{m}_{a_j^m}\tilde{m}_{a_j^{m'}} + \tilde{v}_{a_j^m a_j^{m'}}\right)(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'} \\ E_{\tilde{p}_a\tilde{p}_c\tilde{p}_g}\left[\sum_{j\in J}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right] &= \sum_{j\in J}\sum_{m=1}^{M}\tilde{m}_{a_j^m}(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \sum_m \tilde{m}_{c_j^m}\boldsymbol{W}\boldsymbol{X}^m\tilde{\boldsymbol{m}}_g - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\end{aligned} \tag{E.15}$$

So in the end we have:

$$\tilde{p}_{\boldsymbol{h}}(\boldsymbol{h}) = \mathcal{N}(\tilde{\boldsymbol{m}}_h, \tilde{\boldsymbol{\Sigma}}_h), \text{ where} \tag{E.16}$$

$$\tilde{\boldsymbol{\Sigma}}_h^{-1} = \frac{\boldsymbol{\Sigma}_h^{-1}}{v_h} + \frac{\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}}{v_g} + \sum_{j\in J}\sum_{m,m'}\left(\tilde{m}_{a_j^m}\tilde{m}_{a_j^{m'}} + \tilde{v}_{a_j^m a_j^{m'}}\right)(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'} \tag{E.17}$$

$$\tilde{\boldsymbol{m}}_h = \tilde{\boldsymbol{\Sigma}}_h\left(\frac{\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\tilde{\boldsymbol{m}}_g}{v_g} + \sum_{j\in J}\sum_{m=1}^{M}\tilde{m}_{a_j^m}(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \sum_m \tilde{m}_{c_j^m}\boldsymbol{W}\boldsymbol{X}^m\tilde{\boldsymbol{m}}_g - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right) \tag{E.18}$$

• If we consider the constraint in the norm of $\boldsymbol{h}$:

$$\tilde{\boldsymbol{h}} = \underset{\boldsymbol{h}\ \text{s.t.}\ \boldsymbol{h}^t\boldsymbol{h}=1}{\arg\max}\left(\mathrm{E}_{\tilde{p}_a\tilde{p}_c}\left[\log p(\boldsymbol{h}\,|\,\boldsymbol{y},\boldsymbol{a},\boldsymbol{c},\boldsymbol{g};\boldsymbol{\theta})\right]\right)$$

Solving it amounts to minimizing a quadratic function in $\boldsymbol{h}$ under a quadratic constraint, namely $\boldsymbol{h}^T\boldsymbol{h} = 1$. The function to be minimized is

$$(\boldsymbol{h} - \tilde{\boldsymbol{m}}_h)^T\tilde{\boldsymbol{\Sigma}}_h^{-1}(\boldsymbol{h} - \tilde{\boldsymbol{m}}_h) + \lambda\boldsymbol{h}^T\boldsymbol{h}$$

.

## E.2 E-G step

As in the previous step, we denote by $\tilde{m}_{c_j}$ and $\tilde{\boldsymbol{\Sigma}}_{c_j}$ the mean and covariance matrix of $\tilde{p}_{c_j}$ and write $\tilde{m}_{c_j^m}$ for the $m^{th}$ component of vector $\tilde{m}_{c_j}$ ($\tilde{m}_{c_j^m} = \mathrm{E}_{\tilde{p}_{c_j}}\left[c_j^m\right]$) and $\tilde{v}_{c_j^m c_j^{m'}}$ for the $(m,m')$ element of matrix $\tilde{\boldsymbol{\Sigma}}_{c_j}$ ($\tilde{v}_{c_j^m c_j^{m'}} = \mathrm{E}_{\tilde{p}_{c_j}}\left[c_j^m c_j^{m'}\right] - \tilde{m}_{c_j^m}\tilde{m}_{c_j^{m'}}$).

$$\begin{aligned}\tilde{p}_g(\boldsymbol{g}) &\propto \exp\left(E_{\tilde{p}_a\tilde{p}_h\tilde{p}_c\tilde{p}_q}[\log p(\boldsymbol{h}\,|\,\boldsymbol{y},\boldsymbol{a},\boldsymbol{c},\boldsymbol{h};\boldsymbol{\theta})]\right)\\ &\propto \exp\left(E_{\tilde{p}_a\tilde{p}_h\tilde{p}_c}[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g};\boldsymbol{\theta})] + E_{\tilde{p}_h}[\log p(\boldsymbol{g}|\boldsymbol{\Omega}\boldsymbol{h};v_g)]\right)\end{aligned} \tag{E.19}$$

where $\boldsymbol{a}$ does not depend on $\boldsymbol{g}$.

**1)** Computing $\log p\,(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g};\boldsymbol{\theta})$:
We consider the terms depending on $\boldsymbol{g}$, we write:

$$\begin{aligned}\log p\,(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g};\boldsymbol{\theta}) &= C_{1,g} - \frac{1}{2}\sum_{j\in J}\tilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\tilde{\boldsymbol{y}}_j\\ &= C_{2,g} - \frac{1}{2}\sum_{j\in J}\left[\boldsymbol{g}^{\mathrm{t}}\check{\boldsymbol{S}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\check{\boldsymbol{S}}_j\boldsymbol{g} - 2\boldsymbol{g}^{\mathrm{t}}\check{\boldsymbol{S}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\end{aligned} \tag{E.20}$$

where $\boldsymbol{\Gamma}_j = \frac{1}{v_{b_j}}\boldsymbol{\Lambda}_j$, $C_{1,g} = -\frac{NJ}{2}\log 2\pi + \frac{J}{2}\log|\boldsymbol{\Lambda}_j| - N\sum_{j\in J}\log\sigma_j$ and $C_{2,g} = C_{1,h} - \frac{1}{2}\sum_{j\in J}(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)$.

**2)** Computing $\log p\left(\boldsymbol{g}|\boldsymbol{\Omega}\boldsymbol{h};v_g\right)$:
Based on *a priori* PRF, $\log p\left(\boldsymbol{g}|\boldsymbol{\Omega}\boldsymbol{h};v_g\right) = C_{4,g} - \frac{1}{2v_g}\boldsymbol{g}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g} + \frac{1}{2v_g}2\boldsymbol{g}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\boldsymbol{h}$, with $C_{4,g} = -\frac{D+1}{2}\log 2\pi - \frac{D-1}{2}\log v_g(\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\boldsymbol{h} - \frac{1}{2}\log|\boldsymbol{\Sigma}_g|$.
It follows that:

$$\tilde{p}_g(\boldsymbol{g})\propto\exp\left(-\frac{1}{2}\left[\boldsymbol{g}^{\mathrm{t}}\left(\frac{\boldsymbol{\Sigma}_g^{-1}}{v_g} + E_{\tilde{p}_c}\left[\sum_{j\in J}\check{\boldsymbol{S}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\check{\boldsymbol{S}}_j\right]\right)\boldsymbol{g} - 2\boldsymbol{g}^{\mathrm{t}}\left(\frac{\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h}{v_g} + E_{\tilde{p}_a\tilde{p}_h\tilde{p}_c}\left[\sum_{j\in J}\check{\boldsymbol{S}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\right)\right]\right)$$

$$\text{where,} \quad E_{\tilde{p}_c}\left[\sum_{j\in J}\check{\boldsymbol{S}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\check{\boldsymbol{S}}_j\right] = \sum_{j\in J}\sum_{m,m'}E_{\tilde{p}_{c_j^{m,m'}}}\left[c_j^m c_j^{m'}\right](\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}^{m'}$$
$$= \sum_{j\in J}\sum_{m,m'}\left(\tilde{m}_{c_j^m}\tilde{m}_{c_j^{m'}} + \tilde{v}_{c_j^m c_j^{m'}}\right)(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}^{m'}$$

$$E_{\tilde{p}_a\tilde{p}_h\tilde{p}_c}\left[\sum_{j\in J}\check{\boldsymbol{S}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right] = \sum_{j\in J}\sum_{m=1}^{M}\tilde{m}_{c_j^m}(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \sum_m \tilde{m}_{a_j^m}\boldsymbol{X}^m\tilde{\boldsymbol{m}}_h - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) \tag{E.21}$$

So in the end we have:

$$\tilde{p}_g(\boldsymbol{g}) = \mathcal{N}(\tilde{\boldsymbol{m}}_g, \tilde{\boldsymbol{\Sigma}}_g), \text{ where} \tag{E.22}$$

$$\tilde{\boldsymbol{\Sigma}}_g^{-1} = \frac{\boldsymbol{\Sigma}_g^{-1}}{v_g} + \sum_{j\in J}\sum_{m,m'}\left(\tilde{m}_{c_j^m}\tilde{m}_{c_j^{m'}} + \tilde{v}_{c_j^m c_j^{m'}}\right)(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}^{m'} \tag{E.23}$$

$$\tilde{\boldsymbol{m}}_g = \tilde{\boldsymbol{\Sigma}}_g\left(\frac{\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h}{v_g} + \sum_{j\in J}\sum_{m=1}^{M}\tilde{m}_{c_j^m}(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \sum_m \tilde{m}_{a_j^m}\boldsymbol{X}^m\tilde{\boldsymbol{m}}_h - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right) \tag{E.24}$$

• If we consider the constraint in the norm of $\boldsymbol{g}$:

$$\tilde{\boldsymbol{g}} = \underset{\boldsymbol{g}\text{ s.t. }\boldsymbol{g}^t\boldsymbol{g}=1}{\arg\max}\left(\mathrm{E}_{\tilde{p}_c\tilde{p}_a}\left[\log p(\boldsymbol{h}\,|\,\boldsymbol{y},\boldsymbol{a},\boldsymbol{c},\boldsymbol{h};\boldsymbol{\theta})\right]\right),$$

Solving it amounts to minimizing a quadratic function in $\boldsymbol{g}$ under a quadratic constraint, namely $\boldsymbol{g}^T\boldsymbol{g} = 1$. The function to be minimized is

$$(\boldsymbol{g} - \tilde{\boldsymbol{m}}_g)^T\tilde{\boldsymbol{\Sigma}}_g^{-1}(\boldsymbol{g} - \tilde{\boldsymbol{m}}_g) + \lambda\boldsymbol{g}^T\boldsymbol{g}$$

.

## E.3 E-A step

Here, we focus on Eq. (E.7) where the independence of $\boldsymbol{h}$ and $\boldsymbol{q}$ and $\boldsymbol{a} \perp\!\!\!\perp \boldsymbol{h}|\boldsymbol{q}$ leads to:

$$\tilde{p}_a(\boldsymbol{a}) \propto \exp(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_c\tilde{p}_q}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{y},\boldsymbol{h},\boldsymbol{c},\tilde{\boldsymbol{g}},\boldsymbol{q};\boldsymbol{\theta})\right]) \tag{E.25}$$
$$\propto \exp(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_c\tilde{p}_q}\left[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta}) + \log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\theta})\right]) \tag{E.26}$$
$$\propto \exp(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_c}\left[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g};\boldsymbol{\alpha},\boldsymbol{\ell},\boldsymbol{\Gamma})\right] + \mathrm{E}_{\tilde{p}_q}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\mu},\boldsymbol{v})\right]) \tag{E.27}$$

**1)** Computing $E_{\tilde{p}_h\tilde{p}_g\tilde{p}_c}\left[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta})\right]$:
The same way, the log-likelihood reads:

$$\log p\left(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta}\right) = C_{1,a} - \frac{1}{2}\sum_{j\in J}\tilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\tilde{\boldsymbol{y}}_j$$
$$= C_{2,a} - \frac{1}{2}\sum_{j\in J}\left(\boldsymbol{a}_j^{\mathrm{t}}\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}_{\boldsymbol{h}}\boldsymbol{a}_j - 2\boldsymbol{a}_j^{\mathrm{t}}\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right) \tag{E.28}$$

where $C_{1,a} = C_{1,h}$ were already defined and $C_{2,a} = C_{1,a} - \frac{1}{2}\sum_{j\in J}(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)$.
$\boldsymbol{a}_j = [a_j^1, ..., a_j^M]^{\mathrm{t}}$ and $\boldsymbol{X}_{\boldsymbol{h}} = [\boldsymbol{X}^1\boldsymbol{h}|\boldsymbol{X}^2\boldsymbol{h}|...|\boldsymbol{X}^M\boldsymbol{h}]$.

We can now do the mean of the previous term (equation (E.28)) with respec to other variables:

$$E_{\tilde{p}_h\tilde{p}_c}\left[\log p\left(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta}\right)\right] = C_{2,a} - \frac{1}{2}\sum_{j\in J}\left(\boldsymbol{a}_j^{\mathrm{t}}E_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}_{\boldsymbol{h}}\right]\boldsymbol{a}_j - 2\boldsymbol{a}_j^{\mathrm{t}}E_{\tilde{p}_h\tilde{p}_g\tilde{p}_c}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\right) \tag{E.29}$$

$E_{\widetilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}_{\boldsymbol{h}}\right]$ is a matrix of dimension $M\times M$ where each element $(m,m')$ is written:

$$E_{\widetilde{p}_h}\left[(\boldsymbol{X}^m\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{h}\right]=E_{\widetilde{p}_h}\left[\boldsymbol{h}^{\mathrm{t}}(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{h}\right]=\tilde{\boldsymbol{m}}_h^{\mathrm{t}}(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\tilde{\boldsymbol{m}}_h+\mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\tag{E.30}$$

$E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_c}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\check{\boldsymbol{S}}_j\boldsymbol{g}-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]$ is a vector of dimension $M\times 1$ where each element $(m,1)$ is written:

$$\begin{aligned}E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_c}\left[(\boldsymbol{X}^m\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\check{\boldsymbol{S}}_j\boldsymbol{g}-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]&=E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_c}\left[\boldsymbol{h}^{\mathrm{t}}(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\check{\boldsymbol{S}}_j\boldsymbol{g}-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\\&=\tilde{\boldsymbol{m}}_h^{\mathrm{t}}(\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\sum_{m=1}^{M}\tilde{m}_{c_j^m}\boldsymbol{W}\boldsymbol{X}^m\tilde{\boldsymbol{m}}_g-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\end{aligned}\tag{E.31}$$

**2)** Computing $E_{\widetilde{p}_q}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\mu},\boldsymbol{v})\right]$, where $\boldsymbol{\mu},\boldsymbol{v}$ are the GMM parameters of the hemodynamic component of the signal. We can write:

$$\begin{aligned}\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\mu},\boldsymbol{v})&=\sum_m\sum_j\left[(1-q_j^m)\log\left(\mathcal{N}(\mu_0^m,v_0^m)\right)+q_j^m\log\left(\mathcal{N}(\mu_1^m,v_1^m)\right)\right]\\&=C_{3,a}-\frac{1}{2}\sum_m\sum_j\left[(1-q_j^m)\left(\frac{(a_j^m-\mu_0^m)^2}{v_0^m}\right)+q_j^m\left(\frac{(a_j^m-\mu_1^m)^2}{v_1^m}\right)\right]\end{aligned}\tag{E.32}$$

where $C_{3,a}=\sum_m\sum_j\left[(1-q_j^m)\log\frac{1}{\sqrt{2\pi v_0^m}}+q_j^m\log\frac{1}{\sqrt{2\pi v_1^m}}\right]$

$E_{\widetilde{p}_q}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\mu},\boldsymbol{v})\right]$

$$\begin{aligned}&=C'_{3,a}-\frac{1}{2}\sum_m\sum_j\left[E_{\widetilde{p}_{q_j^m}}\left[(1-q_j^m)\right]\left(\frac{(a_j^m-\mu_0^m)^2}{v_0^m}\right)+\ E_{\widetilde{p}_{q_j^m}}\left[q_j^m\right]\left(\frac{(a_j^m-\mu_1^m)^2}{v_1^m}\right)\right]\\&=C'_{3,a}-\frac{1}{2}\sum_m\sum_j\left[\left(1-\widetilde{p}_{q_j^m}(1)\right)\left(\frac{(a_j^m-\mu_0^m)^2}{v_0^m}\right)+\ \widetilde{p}_{q_j^m}(1)\left(\frac{(a_j^m-\mu_1^m)^2}{v_1^m}\right)\right]\end{aligned}\tag{E.33}$$

where $C'_{3,a}=E_{\widetilde{p}_q}\left[C_{3,a}\right]$.
In matrix form, we note: $\boldsymbol{a}_j=[a_j^1,...,a_j^M]^{\mathrm{t}}$, $\boldsymbol{q}_j=[q_j^1,...,q_j^M]^{\mathrm{t}}$, $\boldsymbol{\mu}_i=[\mu_i^1,...,\mu_i^M]^{\mathrm{t}}$ for $i\in\{0,1\}$ and $\Delta_{a,j}^0=\mathrm{diag}\left[\frac{1-\widetilde{p}_{q_j^1}(1)}{v_0^1},...,\frac{1-\widetilde{p}_{q_j^M}(1)}{v_0^M}\right]$, $\Delta_{a,j}^1=\mathrm{diag}\left[\frac{\widetilde{p}_{q_j^1}(1)}{v_1^1},...,\frac{\widetilde{p}_{q_j^M}(1)}{v_1^M}\right]$. The equation (E.49) becomes:

$$\begin{aligned}E_{\widetilde{p}_q}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\mu},\boldsymbol{v})\right]&=C''_{3,a}-\frac{1}{2}\sum_j\sum_{i=0}^{1}(\boldsymbol{a}_j-\boldsymbol{\mu}_i)^{\mathrm{t}}\Delta_{a,j}^i(\boldsymbol{a}_j-\boldsymbol{\mu}_i)\\&=C''_{3,a}+C_{4,a}-\frac{1}{2}\sum_j\left[\boldsymbol{a}_j^{\mathrm{t}}\left(\sum_{i=0}^{1}\Delta_{a,j}^i\right)\boldsymbol{a}_j-2\boldsymbol{a}_j^{\mathrm{t}}\left(\sum_{i=0}^{1}\Delta_{a,j}^i\boldsymbol{\mu}_i\right)\right]\end{aligned}\tag{E.34}$$

where $C''_{3,a}$ is the matrix form of $C'_{3,a}$, $C_{4,a}=-\frac{1}{2}\sum_j\sum_{i=0}^{1}\boldsymbol{\mu}_i^{\mathrm{t}}\Delta_{a,j}^i\boldsymbol{\mu}_i$.
By replacing equations (E.29) and (E.34) in equation (E.27) we get:

$$\widetilde{p}_a(\boldsymbol{a})\propto\exp\left(-\frac{1}{2}\sum_j(\boldsymbol{a}_j-\tilde{\boldsymbol{m}}_{a_j})^{\mathrm{t}}\tilde{\boldsymbol{\Sigma}}_{a_j}^{-1}(\boldsymbol{a}_j-\tilde{\boldsymbol{m}}_{a_j})\right).$$

Then

$$\widetilde{p}_a(\boldsymbol{a}) = \prod_j \mathcal{N}\left(\tilde{\boldsymbol{m}}_{a_j}, \tilde{\boldsymbol{\Sigma}}_{a_j}\right), \text{ with:} \tag{E.35}$$

$$\tilde{\boldsymbol{\Sigma}}_{a_j} = \left(E_{\widetilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}_{\boldsymbol{h}}\right] + \sum_{i=0}^{1}\Delta_{a,j}^{i}\right)^{-1} \tag{E.36}$$

$$\tilde{\boldsymbol{m}}_{a_j} = \tilde{\boldsymbol{\Sigma}}_{a_j}\left(E_{\widetilde{p}_h\widetilde{p}_c\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \check{\boldsymbol{S}}_j\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right] + \sum_{i=0}^{1}\Delta_{a,j}^{i}\boldsymbol{\mu}_i\right) \tag{E.37}$$

• If we consider the constraint in the norm of $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\widetilde{p}_a(\boldsymbol{a}) = \prod_j \mathcal{N}\left(\tilde{\boldsymbol{m}}_{a_j}, \tilde{\boldsymbol{\Sigma}}_{a_j}\right), \text{ with:} \tag{E.38}$$

$$\tilde{\boldsymbol{\Sigma}}_{a_j} = \left(\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}_{\tilde{\boldsymbol{h}}} + \sum_{i=0}^{1}\Delta_{a,j}^{i}\right)^{-1} \tag{E.39}$$

$$\tilde{\boldsymbol{m}}_{a_j} = \tilde{\boldsymbol{\Sigma}}_{a_j}\left(\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - E_{\widetilde{p}_c}[\check{\boldsymbol{S}}_j\boldsymbol{g}] - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) + \sum_{i=0}^{1}\Delta_{a,j}^{i}\boldsymbol{\mu}_i\right) \tag{E.40}$$

with $\boldsymbol{X}_{\tilde{\boldsymbol{h}}} = [\boldsymbol{X}^1\tilde{\boldsymbol{h}}|\boldsymbol{X}^2\tilde{\boldsymbol{h}}|...|\boldsymbol{X}^M\tilde{\boldsymbol{h}}]$

## E.4 E-C step

Here, we focus on Eq. (E.7) where the independence of $\boldsymbol{g}$ and $\boldsymbol{q}$ and $\boldsymbol{c} \perp\!\!\!\perp \boldsymbol{g}|\boldsymbol{q}$ leads to:

$$\tilde{p}_c(\boldsymbol{c}) \propto \exp(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_a\tilde{p}_q}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{y},\boldsymbol{h},\boldsymbol{a},\tilde{\boldsymbol{g}},\boldsymbol{q};\boldsymbol{\theta})\right]) \tag{E.41}$$

$$\propto \exp(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_a\tilde{p}_q}\left[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta}) + \log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\theta})\right]) \tag{E.42}$$

$$\propto \exp(\mathrm{E}_{\tilde{p}_h\tilde{p}_g\tilde{p}_a}\left[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g};\boldsymbol{\alpha},\boldsymbol{\ell},\boldsymbol{\Gamma})\right] + \mathrm{E}_{\tilde{p}_q}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})\right]) \tag{E.43}$$

**1)** Computing $E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_a}\left[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta})\right]$:
The same way, the log-likelihood reads:

$$\begin{aligned}\log p\left(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta}\right) &= C_{1,c} - \frac{1}{2}\sum_{j\in J}\widetilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}}_j \\ &= C_{2,c} - \frac{1}{2}\sum_{j\in J}\left(\boldsymbol{c}_j^{\mathrm{t}}\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\boldsymbol{c}_j - 2\boldsymbol{c}_j^{\mathrm{t}}\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right)\end{aligned} \tag{E.44}$$

where $C_{1,c} = C_{1,g}$ were already defined and $C_{2,c} = C_{1,c} - \frac{1}{2}\sum_{j\in J}(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)$.
$\boldsymbol{c}_j = [c_j^1,...,c_j^M]^{\mathrm{t}}$ and $\boldsymbol{X}_{\boldsymbol{g}} = [\boldsymbol{X}^1\boldsymbol{g}|\boldsymbol{X}^2\boldsymbol{g}|...|\boldsymbol{X}^M\boldsymbol{g}]$.

We can now do the mean of the previous term (equation (E.44)) with respect to other variables:

$$E_{\widetilde{p}_g\widetilde{p}_a}\left[\log p\left(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta}\right)\right] = C_{2,c} - \frac{1}{2}\sum_{j\in J}\left(\boldsymbol{c}_j^{\mathrm{t}}E_{\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\right]\boldsymbol{c}_j - 2\boldsymbol{c}_j^{\mathrm{t}}E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_a}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\right) \tag{E.45}$$

$E_{\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\right]$ is a matrix of dimension $M \times M$ where each element $(m, m')$ is written:

$$E_{\widetilde{p}_g}\left[\boldsymbol{g}^{\mathrm{t}}(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}^{m'}\boldsymbol{g}\right] = \tilde{\boldsymbol{m}}_g^{\mathrm{t}}(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}^{m'}\tilde{\boldsymbol{m}}_g + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}^{m'}\right) \tag{E.46}$$

$E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_a}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{S}_j\boldsymbol{h}-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]$ is a vector of dimension $M\times 1$ where each element $(m,1)$ is written:

$$\begin{aligned}E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_a}\left[(\boldsymbol{W}\boldsymbol{X}^m\boldsymbol{g})^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{S}_j\boldsymbol{h}-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]&=E_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_a}\left[\boldsymbol{g}^{\mathrm{t}}(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{S}_j\boldsymbol{h}-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]\\&=\tilde{\boldsymbol{m}}_g^{\mathrm{t}}(\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\sum_{m=1}^{M}\tilde{m}_{a_j^m}\boldsymbol{X}^m\tilde{\boldsymbol{m}}_h-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\end{aligned}\tag{E.47}$$

**2)** Computing $E_{\widetilde{p}_q}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})\right]$, where $\boldsymbol{\eta},\boldsymbol{\rho}$ are the GMM parameters of the perfusion component of the signal. We can write:

$$\begin{aligned}\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})&=\sum_m\sum_j\left[(1-q_j^m)\log\left(\mathcal{N}(\eta_0^m,\rho_0^m)\right)+q_j^m\log\left(\mathcal{N}(\eta_1^m,\rho_1^m)\right)\right]\\&=C_{3,c}-\frac{1}{2}\sum_m\sum_j\left[(1-q_j^m)\left(\frac{(c_j^m-\eta_0^m)^2}{\rho_0^m}\right)+q_j^m\left(\frac{(c_j^m-\eta_1^m)^2}{\rho_1^m}\right)\right]\end{aligned}\tag{E.48}$$

where $C_{3,c}=\sum_m\sum_j\left[(1-q_j^m)\log\frac{1}{\sqrt{2\pi\rho_0^m}}+q_j^m\log\frac{1}{\sqrt{2\pi\rho_1^m}}\right]$

$E_{\widetilde{p}_q}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})\right]$

$$\begin{aligned}&=C'_{3,c}-\frac{1}{2}\sum_m\sum_j\left[E_{\widetilde{p}_{q_j^m}}\left[(1-q_j^m)\right]\left(\frac{(c_j^m-\eta_0^m)^2}{\rho_0^m}\right)+\ E_{\widetilde{p}_{q_j^m}}\left[q_j^m\right]\left(\frac{(c_j^m-\eta_1^m)^2}{\rho_1^m}\right)\right]\\&=C'_{3,c}-\frac{1}{2}\sum_m\sum_j\left[\left(1-\widetilde{p}_{q_j^m}(1)\right)\left(\frac{(c_j^m-\eta_0^m)^2}{\rho_0^m}\right)+\ \widetilde{p}_{q_j^m}(1)\left(\frac{(c_j^m-\eta_1^m)^2}{\rho_1^m}\right)\right]\end{aligned}\tag{E.49}$$

where $C'_{3,c}=E_{\widetilde{p}_q}\left[C_{3,c}\right]$.
In matrix form, we note: $\boldsymbol{c}_j=[c_j^1,...,c_j^M]^{\mathrm{t}}$, $\boldsymbol{q}_j=[q_j^1,...,q_j^M]^{\mathrm{t}}$, $\boldsymbol{\eta}_i=[\eta_i^1,...,\eta_i^M]^{\mathrm{t}}$ for $i\in\{0,1\}$ and $\Delta_{c,j}^0=\mathrm{diag}\left[\frac{1-\widetilde{p}_{q_j^1}(1)}{v_0^1},...,\frac{1-\widetilde{p}_{q_j^M}(1)}{\rho_0^M}\right]$, $\Delta_{c,j}^1=\mathrm{diag}\left[\frac{\widetilde{p}_{q_j^1}(1)}{\rho_1^1},...,\frac{\widetilde{p}_{q_j^M}(1)}{\rho_1^M}\right]$. The equation (E.49) becomes:

$$\begin{aligned}E_{\widetilde{p}_q}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})\right]&=C''_{3,c}-\frac{1}{2}\sum_j\sum_{i=0}^{1}(\boldsymbol{c}_j-\boldsymbol{\eta}_i)^{\mathrm{t}}\Delta_{c,j}^i(\boldsymbol{c}_j-\boldsymbol{\eta}_i)\\&=C''_{3,c}+C_{4,c}-\frac{1}{2}\sum_j\left[\boldsymbol{c}_j^{\mathrm{t}}\left(\sum_{i=0}^{1}\Delta_{c,j}^i\right)\boldsymbol{c}_j-2\boldsymbol{c}_j^{\mathrm{t}}\left(\sum_{i=0}^{1}\Delta_{c,j}^i\boldsymbol{\eta}_i\right)\right]\end{aligned}\tag{E.50}$$

where $C''_{3,c}$ is the matrix form of $C'_{3,c}$, $C_{4,c}=-\frac{1}{2}\sum_j\sum_{i=0}^{1}\boldsymbol{\eta}_i^{\mathrm{t}}\Delta_{c,j}^i\boldsymbol{\eta}_i$.
By replacing equations (E.45) and (E.50) in equation (E.43) we get:

$$\widetilde{p}_c(\boldsymbol{c})\propto\exp\left(-\frac{1}{2}\sum_j(\boldsymbol{c}_j-\tilde{\boldsymbol{m}}_{c_j})^{\mathrm{t}}\tilde{\boldsymbol{\Sigma}}_{c_j}^{-1}(\boldsymbol{c}_j-\tilde{\boldsymbol{m}}_{c_j})\right).$$

Then

$$\widetilde{p}_c(\boldsymbol{c})=\prod_j\mathcal{N}\left(\tilde{\boldsymbol{m}}_{c_j},\tilde{\boldsymbol{\Sigma}}_{c_j}\right),\text{ with:}\tag{E.51}$$

$$\tilde{\boldsymbol{\Sigma}}_{c_j}=\left(E_{\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\right]+\sum_{i=0}^{1}\Delta_{c,j}^i\right)^{-1}\tag{E.52}$$

$$\tilde{\boldsymbol{m}}_{c_j}=\tilde{\boldsymbol{\Sigma}}_{c_j}\left(E_{\widetilde{p}_a\widetilde{p}_h\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-E_{\widetilde{p}_a}[\boldsymbol{S}_j\boldsymbol{h}]-\bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\right]+\sum_{i=0}^{1}\Delta_{c,j}^i\boldsymbol{\eta}_i\right)\tag{E.53}$$

• If we consider the constraint in the norm of $\boldsymbol{g}$:

$$\widetilde{p}_c(\boldsymbol{c}) = \prod_j \mathcal{N}\left(\tilde{\boldsymbol{m}}_{c_j}, \tilde{\boldsymbol{\Sigma}}_{c_j}\right), \text{ with:} \tag{E.54}$$

$$\tilde{\boldsymbol{\Sigma}}_{c_j} = \left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{W} \boldsymbol{X}_{\tilde{\boldsymbol{g}}} + \sum_{i=0}^{1} \Delta_{c,j}^{i}\right)^{-1} \tag{E.55}$$

$$\tilde{\boldsymbol{m}}_{c_j} = \tilde{\boldsymbol{\Sigma}}_{c_j} \left(\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h} - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j) + \sum_{i=0}^{1} \Delta_{c,j}^{i} \boldsymbol{\eta}_i\right) \tag{E.56}$$

with $\boldsymbol{X}_{\tilde{\boldsymbol{g}}} = [\boldsymbol{X}^1 \tilde{\boldsymbol{g}} | \boldsymbol{X}^2 \tilde{\boldsymbol{g}} | ... | \boldsymbol{X}^M \tilde{\boldsymbol{g}}]$

## E.5 E-Q step

We assume

$$\tilde{p}_{q_m}(\boldsymbol{q}^m) = \prod_{j \in \mathcal{J}} \tilde{p}_{q_j^m}(q_j^m) \tag{E.57}$$

$$\begin{aligned}
\tilde{p}_{q_j^m}(q_j^m) \propto \exp & \left(E_{\tilde{p}_a, \tilde{p}_{q_{\setminus j}^m}, \tilde{p}_{q_{\setminus m}}} \left[\log p(q_j^m \mid q_{\setminus j}^m, q_{\setminus m}, \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\theta})\right]\right) \\
\propto \exp & \left(E_{\tilde{p}_{a_j^m}} \left[\log p(a_j^m \mid q_j^m, \boldsymbol{\theta})\right]\right. \\
& + E_{\tilde{p}_{c_j^m}} \left[\log p(c_j^m \mid q_j^m, \boldsymbol{\theta})\right] \\
& \left. + E_{\tilde{p}_{q_{\setminus j}^m}} \left[\log p(q_j^m \mid q_{\setminus j}^m; \beta^m)\right]\right)
\end{aligned} \tag{E.58}$$

**1)** Computation $E_{\tilde{p}_{a_j^m}} \left[\log p(a_j^m \mid q_j^m, \boldsymbol{\theta})\right]$:

$$\log p(a_j^m \mid q_j^m, \boldsymbol{\theta}) = (1 - q_j^m) \log \mathcal{N}(a_j^m; \mu_0^m, v_0^m) + q_j^m \log \mathcal{N}(a_j^m; \mu_1^m, v_1^m) \tag{E.59}$$

Considering $E_{\tilde{p}_{a_j^m}}[a_j^m a_j^{m'}] = \tilde{v}_{a_j^m a_j^{m'}} + \tilde{m}_{a_j^m} \tilde{m}_{a_j^{m'}}$, we get

$$\begin{aligned}
E_{\tilde{p}_{a_j^m}} \left[\log p(a_j^m \mid q_j^m, \boldsymbol{\theta})\right] = & (1 - q_j^m) \left(\log \mathcal{N}(\tilde{m}_{a_j^m}; \mu_0^m, v_0^m) - \frac{1}{2} \frac{\tilde{v}_{a_j^m a_j^m}}{v_0^m}\right) \\
& + q_j^m \left(\log \mathcal{N}(\tilde{m}_{a_j^m}; \mu_1^m, v_1^m) - \frac{1}{2} \frac{\tilde{v}_{a_j^m a_j^m}}{v_1^m}\right)
\end{aligned} \tag{E.60}$$

**2)** Computation $E_{\tilde{p}_{c_j^m}} \left[\log p(c_j^m \mid q_j^m, \boldsymbol{\theta})\right]$:

$$\log p(c_j^m \mid q_j^m, \boldsymbol{\theta}) = (1 - q_j^m) \log \mathcal{N}(c_j^m; \eta_0^m, \rho_0^m) + q_j^m \log \mathcal{N}(c_j^m; \eta_1^m, \rho_1^m) \tag{E.61}$$

Considering $E_{\tilde{p}_{c_j^m}}[c_j^m c_j^{m'}] = \tilde{v}_{c_j^m c_j^{m'}} + \tilde{m}_{c_j^m} \tilde{m}_{c_j^{m'}}$, we get

$$\begin{aligned}
E_{\tilde{p}_{c_j^m}} \left[\log p(c_j^m \mid q_j^m, \boldsymbol{\theta})\right] = & (1 - q_j^m) \left(\log \mathcal{N}(\tilde{m}_{c_j^m}; \eta_0^m, \rho_0^m) - \frac{1}{2} \frac{\tilde{v}_{c_j^m c_j^m}}{\rho_0^m}\right) \\
& + q_j^m \left(\log \mathcal{N}(\tilde{m}_{c_j^m}; \eta_1^m, \rho_1^m) - \frac{1}{2} \frac{\tilde{v}_{c_j^m c_j^m}}{\rho_1^m}\right)
\end{aligned} \tag{E.62}$$

**3)** Computation of $E_{\widetilde{p}_{q^m_{\setminus j}}}\Big[\log p(q^m_j \,|\, q^m_{\setminus j}; \beta^m)\Big]$: Considering the *a priori* distribution of variable $\boldsymbol{q}$

$$\log p(q^m_j \,|\, q^m_{\setminus j}; \beta^m) = \beta^m \sum_{k \in N(j)} \mathbb{1}(q^m_j = q^m_k) + C_{3,q^m_j} \tag{E.63}$$

where $N(j)$ is the neighborhood of $j$ and where $C_{3,q^m_j}$ is a constant in $q^m_j$.

$$E_{\widetilde{p}_{q^m_{\setminus j}}}\Big[\log p(q^m_j \,|\, q^m_{\setminus j}; \beta^m)\Big] = C'_{3,q^m_j} + \beta^m \sum_{k \in N(j)} E_{\widetilde{p}_{q^m_k}}\Big[\mathbb{1}(q^m_j = q^m_k)\Big] \tag{E.64}$$

with $C'_{3,q^m_j} = E_{\widetilde{p}_{q^m_{\setminus j}}}\Big[C_{3,q^m_j}\Big]$, and $E_{\widetilde{p}_{q^m_{\setminus j}}}\Big[\mathbb{1}(q^m_j = q^m_k)\Big] = p_{\widetilde{q}^m_k}(q^m_j)$.

Finally, we get

$$\begin{aligned}
\widetilde{p}_{q^m_j}(q^m_j) \propto \exp\Bigg( & (1 - q^m_j)\left(\log \mathcal{N}(\tilde{m}_{a^m_j}; \mu^m_0, v^m_0) - \frac{1}{2}\frac{\tilde{v}_{a^m_j a^m_j}}{v^m_0}\right) \\
& + q^m_j \left(\log \mathcal{N}(\tilde{m}_{a^m_j}; \mu^m_1, v^m_1) - \frac{1}{2}\frac{\tilde{v}_{a^m_j a^m_j}}{v^m_1}\right) \\
& + (1 - q^m_j)\left(\log \mathcal{N}(\tilde{m}_{c^m_j}; \eta^m_0, \rho^m_0) - \frac{1}{2}\frac{\tilde{v}_{c^m_j c^m_j}}{\rho^m_0}\right) \\
& + q^m_j \left(\log \mathcal{N}(\tilde{m}_{c^m_j}; \eta^m_1, \rho^m_1) - \frac{1}{2}\frac{\tilde{v}_{c^m_j c^m_j}}{\rho^m_1}\right) \\
& + \beta^m \sum_{k \in N(j)} \widetilde{p}_{q^m_k}(q^m_j)\Bigg)
\end{aligned} \tag{E.65}$$

## E.6 M-step

As for the E-step above, the superscript $(r)$ is omitted in the following developments about the M-step (E.12).

$$\begin{aligned}
\widehat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathrm{E}_{\widetilde{p}_h \widetilde{p}_a \widetilde{p}_g \widetilde{p}_c \widetilde{p}_q}\big[\log p(\boldsymbol{y}, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{q}; \boldsymbol{\theta})\big] \\
&= \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Big[ \mathrm{E}_{\widetilde{p}_h \widetilde{p}_a \widetilde{p}_g \widetilde{p}_c}\big[\log p(\boldsymbol{y} \,|\, \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}; \boldsymbol{\alpha}, \boldsymbol{\ell}, \boldsymbol{\Gamma})\big] + \mathrm{E}_{\widetilde{p}_h \widetilde{p}_a \widetilde{p}_q}\big[\log p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big] + \mathrm{E}_{\widetilde{p}_h}\big[\log p(\boldsymbol{h}; v_h)\big] \\
&\qquad + \mathrm{E}_{\widetilde{p}_c \widetilde{p}_q}\big[\log p(\boldsymbol{c} \,|\, \boldsymbol{q}; \boldsymbol{\eta}, \boldsymbol{\rho})\big] + \mathrm{E}_{\widetilde{p}_g}\big[\log p(\tilde{\boldsymbol{g}}; v_g)\big] + \mathrm{E}_{\widetilde{p}_q}\big[\log p(\boldsymbol{q}; \boldsymbol{\beta})\big] \\
&\qquad + \log p(\beta; \lambda_\beta) + \log p(v_g; \lambda_g) + \log p(v_h; \lambda_h) \Big]
\end{aligned}$$

Given the separability of the priors pdfs, it follows that the M-step also divides into separate M-steps:

### E.6.1 M-$(\mu, v)$ step

Updating parameters $\boldsymbol{\mu}$ and $\boldsymbol{v}$ is straightforward since closed-form expressions are available. It is actually similar to updating parameters of a standard Gaussian mixture:

$$(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{v}}) = \arg\max_{\boldsymbol{\mu}, \boldsymbol{v}} \mathrm{E}_{\widetilde{p}_a \widetilde{p}_q}\big[\log p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big]$$

In the E-Q step we already computed:

$$\mathrm{E}_{\widetilde{p}_a}\big[\log p(\boldsymbol{a} \,|\, \boldsymbol{q}; \boldsymbol{\mu}, \boldsymbol{v})\big] = \sum_{m=1}^{M} \sum_{j \in \mathcal{J}} \sum_{i = \{0,1\}} \mathbb{1}(q^m_j = i)\left(\log \mathcal{N}(\tilde{m}_{a^m_j}; \mu^m_i, v^m_i) - \frac{1}{2}\frac{\tilde{v}_{a^m_j a^m_j}}{v^m_i}\right)$$

from which it comes:

$$\mathrm{E}_{\tilde{p}_q\tilde{p}_a}\big[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\mu},\boldsymbol{v})\big] = \sum_{m=1}^{M}\sum_{j\in\mathcal{J}}\sum_{i=\{0,1\}} \tilde{p}_{q_j^m}(i)\left(\log\mathcal{N}(\tilde{m}_{a_j^m};\mu_i^m,v_i^m) - \frac{1}{2}\frac{\tilde{v}_{a_j^m a_j^m}}{v_i^m}\right)$$

This latter expression is similar to the one to be maximized in a standard Gaussian mixture when replacing the observed data by the $\tilde{m}_{a_j^m}$'s. Not surprisingly, we get:

$$\widehat{\mu}_i^m = \frac{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)\;\tilde{m}_{a_j^m}}{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)} \quad\text{and}\quad \widehat{v}_i^m = \frac{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)\;\left[(\tilde{m}_{a_j^m}-\mu_i^m)^2 + \tilde{v}_{a_j^m a_j^m}\right]}{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)}$$

### E.6.2 M-$(\eta,\rho)$ step

Updating parameters $\boldsymbol{\eta}$ and $\boldsymbol{\rho}$ goes as for $\boldsymbol{\mu}$ and $\boldsymbol{v}$:

$$(\widehat{\boldsymbol{\eta}},\widehat{\boldsymbol{\rho}}) = \arg\max_{\boldsymbol{\eta},\boldsymbol{\rho}} \mathrm{E}_{\tilde{p}_a\tilde{p}_q}\big[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})\big]$$

In the E-Q step we already computed:

$$\mathrm{E}_{\tilde{p}_c}\big[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})\big] = \sum_{m=1}^{M}\sum_{j\in\mathcal{J}}\sum_{i=\{0,1\}} \mathbb{1}(q_j^m=i)\left(\log\mathcal{N}(\tilde{m}_{c_j^m};\eta_i^m,\rho_i^m) - \frac{1}{2}\frac{\tilde{v}_{c_j^m c_j^m}}{\rho_i^m}\right)$$

from which it comes:

$$\mathrm{E}_{\tilde{p}_q\tilde{p}_c}\big[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\eta},\boldsymbol{\rho})\big] = \sum_{m=1}^{M}\sum_{j\in\mathcal{J}}\sum_{i=\{0,1\}} \tilde{p}_{q_j^m}(i)\left(\log\mathcal{N}(\tilde{m}_{c_j^m};\eta_i^m,\rho_i^m) - \frac{1}{2}\frac{\tilde{v}_{a_j^m a_j^m}}{\rho_i^m}\right)$$

Finally, we get:

$$\widehat{\eta}_i^m = \frac{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)\;\tilde{m}_{c_j^m}}{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)} \quad\text{and}\quad \widehat{\rho}_i^m = \frac{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)\;\left[(\tilde{m}_{c_j^m}-\eta_i^m)^2 + \tilde{v}_{c_j^m c_j^m}\right]}{\sum_{j\in\mathcal{J}} \tilde{p}^{(r)}_{q_j^m}(i)}$$

### E.6.3 M-$v_h$ step

$$\hat{v}_h = \arg\max_{v_h} f(v_h) = \arg\max_{v_h}\Big\{\mathrm{E}_{\tilde{p}_h}\big[\log p(\boldsymbol{h}\,|\,v_h) + \log p(v_h\,|\,\lambda_h)\big]\Big\} \tag{E.66}$$

where the *a priori* $p(v_h\,|\,\lambda_h) = \lambda_h\exp(-\lambda_h v_h)$ allows the better estimation of $v_h$ [Chaari et al., 2013].

1. If we use an hyperprior

$$\mathrm{E}_{\tilde{p}_h}\big[\log p(\boldsymbol{h}\,|\,v_h)\big] = C_{v_h} - \frac{D-1}{2}\log v_h - \frac{\mathrm{E}_{\tilde{p}_h}\big[\boldsymbol{h}^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\boldsymbol{h}\big]}{2v_h}$$

$$\frac{\partial f(v_h)}{\partial v_h} = -\frac{D-1}{2v_h} + \frac{\tilde{\boldsymbol{m}}_h^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Sigma}_h^{-1}\right)}{2v_h^2} - \lambda_h = 0$$

where $\mathrm{E}_{\tilde{p}_h}\big[\boldsymbol{h}^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\boldsymbol{h}\big] = \tilde{\boldsymbol{m}}_h^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Sigma}_h^{-1}\right)$ and $\tilde{\boldsymbol{m}}_h^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{m}}_h = \mathrm{tr}\left(\tilde{\boldsymbol{m}}_h\tilde{\boldsymbol{m}}_h^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\right)$

$$\Rightarrow\quad v_h = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_h\left(\tilde{\boldsymbol{m}}_h^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Sigma}_h^{-1}\right)\right)}}{4\lambda_h} \tag{E.67}$$

- If we do consider the constraint in the norm of $\boldsymbol{h}$ when we use a hyperprior:

$$\Rightarrow \quad v_h = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_h \mathrm{tr}\left[\left(\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^{\mathrm{t}}\right)\boldsymbol{\Sigma}_h^{-1}\right]}}{4\lambda_h} \tag{E.68}$$

2. If we do not use an hyperprior:

$$\Rightarrow \quad v_h = \frac{\tilde{\boldsymbol{m}}_h^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Sigma}_h^{-1}\right)}{D-1} \tag{E.69}$$

- If we consider the constraint in the norm of $\boldsymbol{h}$ when we do not use a hyperprior:

$$\Rightarrow \quad v_h = \frac{\mathrm{tr}\left[\left(\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}^{\mathrm{t}}\right)\boldsymbol{\Sigma}_h^{-1}\right]}{D-1} \tag{E.70}$$

### E.6.4 M-$v_g$ step

The same way:

$$\hat{v}_g = \arg\max_{v_g} f(v_g) = \arg\max_{v_g}\left\{\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\left[\log p(\boldsymbol{g}\,|\,v_g) + \log p(v_g\,|\,\lambda_g)\right]\right\} \tag{E.71}$$

where the *a priori* $p(v_g\,|\,\lambda_g) = \lambda_g \exp(-\lambda_g v_g)$ allows the better estimation of $v_g$ [Chaari et al., 2013].

1. If we use an hyperprior

$$\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\left[\log p(\boldsymbol{g}\,|\,v_g)\right] = C_{v_g} - \frac{D-1}{2}\log v_g - \frac{\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\left[(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})\right]}{2v_g}$$

$$\frac{\partial f(v_g)}{\partial v_g} = -\frac{D-1}{2v_g} + \frac{\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\left[(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})\right]}{2v_g^2} - \lambda_g = 0$$

Here:

$$\begin{aligned}
&\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\left[(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{g}-\boldsymbol{\Omega}\boldsymbol{h})\right] = \\
&\quad = \mathrm{E}_{\tilde{p}_h\tilde{p}_g}\left[\boldsymbol{g}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g} - 2\boldsymbol{g}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\boldsymbol{h} + (\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\Omega}\boldsymbol{h})\right] \\
&\quad = \mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{g}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g}\right] - 2\mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{g}^{\mathrm{t}}\right]\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{h}\right] + \mathrm{E}_{\tilde{p}_h}\left[(\boldsymbol{\Omega}\boldsymbol{h})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\Omega}\boldsymbol{h})\right] \\
&\quad = \tilde{\boldsymbol{m}}_g^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\tilde{\boldsymbol{m}}_g + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g\boldsymbol{\Sigma}_g^{-1}\right) - 2\tilde{\boldsymbol{m}}_g^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h^{\mathrm{t}} + (\boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h)^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\right) \\
&\quad = (\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h)^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h) + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g\boldsymbol{\Sigma}_g^{-1}\right) + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\right)
\end{aligned}$$

since $\mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{g}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{g}\right] = \tilde{\boldsymbol{m}}_g^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\tilde{\boldsymbol{m}}_g + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g\boldsymbol{\Sigma}_g^{-1}\right)$ and $\tilde{\boldsymbol{m}}_g^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\tilde{\boldsymbol{m}}_g = \mathrm{tr}\left(\tilde{\boldsymbol{m}}_g\tilde{\boldsymbol{m}}_g^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\right)$.

$$\Rightarrow \quad \hat{v}_g = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_g\left((\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h)^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h) + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g\boldsymbol{\Sigma}_g^{-1} + \tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\Omega}\right)\right)}}{4\lambda_g} \tag{E.72}$$

- If we do consider the constraint in the norm of $\boldsymbol{h}$:

$$\Rightarrow \quad \hat{v}_g = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_g\left((\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{h}}) + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g\boldsymbol{\Sigma}_g^{-1}\right)\right)}}{4\lambda_g} \tag{E.73}$$

- If we do consider the constraint in the norm of $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\Rightarrow \quad \hat{v}_g = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_g \left( (\tilde{\boldsymbol{g}} - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})^{\mathrm{t}} \boldsymbol{\Sigma}_g^{-1} (\tilde{\boldsymbol{g}} - \boldsymbol{\Omega}\tilde{\boldsymbol{h}}) \right)}}{4\lambda_g} \tag{E.74}$$

2. If we do not use an hyperprior:

$$\Rightarrow \quad v_g = \frac{(\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h)^{\mathrm{t}} \boldsymbol{\Sigma}_g^{-1} (\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h) + \mathrm{tr}\left( \tilde{\boldsymbol{\Sigma}}_g \boldsymbol{\Sigma}_g^{-1} + \tilde{\boldsymbol{\Sigma}}_h \boldsymbol{\Omega}^{\mathrm{t}} \boldsymbol{\Sigma}_g^{-1} \boldsymbol{\Omega} \right)}{D-1} \tag{E.75}$$

- If we consider the constraint in the norm of $\boldsymbol{h}$:

$$\Rightarrow \quad v_g = \frac{(\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})^{\mathrm{t}} \boldsymbol{\Sigma}_g^{-1} (\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{h}}) + \mathrm{tr}\left( \tilde{\boldsymbol{\Sigma}}_g \boldsymbol{\Sigma}_g^{-1} \right)}{D-1} \tag{E.76}$$

- If we do consider the constraint in the norm of $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\Rightarrow \quad v_g = \frac{(\tilde{\boldsymbol{g}} - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})^{\mathrm{t}} \boldsymbol{\Sigma}_g^{-1} (\tilde{\boldsymbol{g}} - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})}{D-1} \tag{E.77}$$

### E.6.5 M-$\beta$ step

The maximization over each $\beta^m$ corresponds to the M-step obtained for a standard Hidden MRF model:

$$\begin{aligned}
\widehat{\beta}^m &= \arg\max_{\beta^m} \mathrm{E}_{\tilde{p}_{q^m}} \left[ \log p(\boldsymbol{q}^m; \beta^m) \right] \\
&= \arg\max_{\beta^m} E_{\tilde{p}_{q^m}} \left[ \log p(\boldsymbol{q}^m \mid \beta^m) \right] + \log p(\beta^m \mid \lambda_\beta) \\
&= \arg\max_{\beta^m} \left\{ -\log Z(\beta^m) + \beta^m \left[ E_{\tilde{p}_{q^m}} [U(\boldsymbol{q}^m)] - \lambda_\beta \right] + C_{\beta^m} \right\} \\
&= \arg\max_{\beta^m} f(\beta^m)
\end{aligned}$$

The maximization of $f(\beta^m)$ needs the computation of its derivative with respect to $\beta^m$:

$$\frac{df(\beta^m)}{d\beta^m} = -\frac{d \log Z(\beta^m)}{d\beta^m} + E_{\tilde{p}_{q^m}} [U(\boldsymbol{q}^m)] - \lambda_\beta \tag{E.78}$$

where

$$\mathrm{E}_{\tilde{p}_{q^m}} \left[ U(\boldsymbol{q}^m) \right] = \sum_{j \sim k} \mathrm{E}_{\tilde{p}_{q^m}} \left[ \mathbb{1}(q_j^m = q_k^m) \right] = \frac{1}{2} \sum_j \sum_{k \in N(j)} \tilde{p}_{q^m}(q_j^m = q_k^m) = \frac{1}{2} \sum_j \sum_{k \in N(j)} \sum_{i=\{0,1\}} (\tilde{p}_{q_j^m}^{(r)}(i) \tilde{p}_{q_k^m}^{(r)}(i)) \tag{E.79}$$

$$Z(\beta^m) = \sum_{\boldsymbol{q}^m} \exp(\beta^m U(\boldsymbol{q}^m)) \tag{E.80}$$

$$\frac{d \log Z(\beta^m)}{d\beta^m} = \frac{1}{Z(\beta^m)} \frac{d \log Z(\beta^m)}{d\beta^m} = \frac{\sum_{\boldsymbol{q}^m} U(\boldsymbol{q}^m) \exp(\beta^m U(\boldsymbol{q}^m))}{\sum_{\boldsymbol{q}^m} \exp(\beta^m U(\boldsymbol{q}^m))} = E_{p_{q^m}} [U(\boldsymbol{q}^m)] \tag{E.81}$$

If we use the mean field like approximation as in [Celeux et al., 2003], we can approximate $E_{p_{q^m}} [U(\boldsymbol{q}^m)]$ by $E_{\tilde{p}^{MF}} [U(\boldsymbol{q}^m)]$ and the equation becomes:

$$\frac{df(\beta^m)}{d\beta^m} = -E_{\tilde{p}^{MF}} [U(\boldsymbol{q}^m)] + E_{\tilde{p}_{q^m}} [U(\boldsymbol{q}^m)] - \lambda_\beta \tag{E.82}$$

Since $\widetilde{p}^{MF}(\boldsymbol{q}^m) = \prod_j \widetilde{p}^{MF}(q_j^m)$ and $\widetilde{p}^{MF}(\boldsymbol{q}_j^m;\beta^m) = \frac{\exp(\beta^m \sum_{k\in N(j)} \widetilde{p}_{q_k^m}^{(r-1)}(q_j^m)}{\sum_{l\in\{0,1\}} \exp(\beta^m \sum_{k\in N(j)} \widetilde{p}_{q_k^m}^{(r-1)}(l)}$, we can compute:

$$E_{p^{MF}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_j \sum_{k\in N(j)} \sum_{i=0}^{1} p_j^{MF}(i) p_k^{MF}(i) \tag{E.83}$$

It follows that $\beta^m$ must satisfy the following equation:

$$\frac{1}{2}\sum_j \sum_{k\in N(j)} \sum_{i=\{0,1\}} \left(\widetilde{p}_{q_j^m}^{(r)}(i)\ \widetilde{p}_{q_k^m}^{(r)}(i) - \widetilde{p}^{MF}(q_j^m = i;\beta^m)\ \widetilde{p}^{MF}(q_k^m = i;\beta^m)\right) - \lambda_\beta = 0 \tag{E.84}$$

### E.6.6 M-$(\boldsymbol{\ell}, \boldsymbol{\Gamma}, \boldsymbol{\alpha})$ step

This maximization problem factorizes over voxel so that for each $j \in \mathcal{J}$, we compute:

$$(\widehat{\boldsymbol{\ell}}_j, \hat{v}_{b_j}, \hat{\alpha}_j) = \underset{\alpha_j,\boldsymbol{\ell}_j,\boldsymbol{\Gamma}_j}{\arg\max}\, \mathrm{E}_{\widetilde{p}_h\widetilde{p}_{a_j}\widetilde{p}_g\widetilde{p}_{c_j}}\left[\log p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}; \alpha_j, \boldsymbol{\ell}_j, \boldsymbol{\Gamma}_j)\right]$$

**M-$\boldsymbol{v}_{b_j}$**
Considering the independence between $\{\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}; \alpha_j, \boldsymbol{\ell}_j, v_{b_j}\}_{j=1:J}$ we can write:

$$\hat{v}_{b_j} = \underset{v_{b_j}}{\arg\max}\, \mathrm{E}_{\widetilde{p}_h\widetilde{p}_{a_j}\widetilde{p}_g\widetilde{p}_{c_j}}\left[\log p\left(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}; \alpha_j, \boldsymbol{\ell}_j, v_{b_j}\right)\right] \tag{E.85}$$

From the definition of likelihood we can compute, we know that $\boldsymbol{\Gamma}_j = \frac{1}{v_{b_j}}\boldsymbol{\Lambda}_j$, where $\boldsymbol{\Lambda}_j = \boldsymbol{I}$,

$$\mathrm{E}_{\widetilde{p}_{\boldsymbol{a}_j},\widetilde{p}_{\boldsymbol{c}_j}}\left[\log p\left(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}; \alpha_j, \boldsymbol{\ell}_j, v_{b_j}\right)\right] = C_{v_{b_j}} - \frac{N}{2}\log v_{b_j} - \frac{\mathrm{E}_{\widetilde{p}_h\widetilde{p}_g\widetilde{p}_{\boldsymbol{a}_j}\widetilde{p}_{\boldsymbol{c}_j}}\left[\tilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Lambda}_j\tilde{\boldsymbol{y}}_j\right]}{2v_{b_j}}$$

being $\tilde{\boldsymbol{y}}_j = \boldsymbol{y}_j - \sum_m^M a_j^m \boldsymbol{X}^m\boldsymbol{h} - \sum_m^M c_j^m \boldsymbol{W}\boldsymbol{X}^m\boldsymbol{g} - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j$

**Computing $\tilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Lambda}_j\tilde{\boldsymbol{y}}_j$:**

$$\begin{aligned}\tilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Lambda}_j\tilde{\boldsymbol{y}}_j = {}& (\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) + (\boldsymbol{X}_{\boldsymbol{h}}\boldsymbol{a}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\boldsymbol{h}}\boldsymbol{a}_j \\ & + (\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\boldsymbol{c}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\boldsymbol{c}_j - 2(\boldsymbol{X}_{\boldsymbol{h}}\boldsymbol{a}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) \\ & - 2(\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\boldsymbol{c}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) + 2(\boldsymbol{X}_{\boldsymbol{h}}\boldsymbol{a}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\boldsymbol{c}_j\end{aligned} \tag{E.86}$$

so therefore we can write:
$\mathrm{E}_{\widetilde{p}_h\widetilde{p}_{a_j}\widetilde{p}_g\widetilde{p}_{c_j}}\left[\log p\left(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}, \alpha_j, \boldsymbol{\ell}_j, \sigma_j^2\right)\right]$

$$\begin{aligned}= {}& C_{v_{b_j}} - \frac{N}{2}\log v_{b_j} - \frac{1}{2v_{b_j}}\Big[\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}}\mathrm{E}_{\widetilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\boldsymbol{h}}\right]\tilde{\boldsymbol{m}}_{a_j} \\ & + tr(\tilde{\boldsymbol{\Sigma}}_{a_j}\mathrm{E}_{\widetilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\boldsymbol{h}}\right]) + \tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\mathrm{E}_{\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\right]\tilde{\boldsymbol{m}}_{c_j} \\ & + tr(\tilde{\boldsymbol{\Sigma}}_{c_j}\mathrm{E}_{\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\right]) - 2\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}}\mathrm{E}_{\widetilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\right]\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) \\ & - 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\mathrm{E}_{\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\right]\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) + 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\mathrm{E}_{\widetilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\right]\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\mathrm{E}_{\widetilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}\right]\tilde{\boldsymbol{m}}_{a_j} \\ & + (\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)\Big]\} \\ = {}& C_{v_{b_j}} - \frac{N}{2}\log v_{b_j} - \frac{1}{2v_{b_j}}V_j\end{aligned} \tag{E.87}$$

with

$$\begin{aligned} V_j = {} & \tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}} \mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{X}_{\boldsymbol{h}}\right] \tilde{\boldsymbol{m}}_{a_j} + tr(\tilde{\boldsymbol{\Sigma}}_{a_j} \mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{X}_{\boldsymbol{h}}\right]) + \tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}} \mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{W} \boldsymbol{X}_{\boldsymbol{g}}\right] \tilde{\boldsymbol{m}}_{c_j} \\ & + tr(\tilde{\boldsymbol{\Sigma}}_{c_j} \mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{W} \boldsymbol{X}_{\boldsymbol{g}}\right]) - 2\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}} \mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\right] \boldsymbol{\Lambda}_j (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j) - 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}} \mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\right] \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j) \\ & + 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}} \mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\right] \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j \mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}\right] \tilde{\boldsymbol{m}}_{a_j} + (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j)^{\mathrm{t}} \boldsymbol{\Lambda}_j (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j) \end{aligned} \tag{E.88}$$

where $\mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}\right]$ is a matrix with columns $\boldsymbol{X}^1 \tilde{\boldsymbol{m}}_h | ... | \boldsymbol{X}^M \tilde{\boldsymbol{m}}_h$, $\mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}\right]$ is a matrix with columns $\boldsymbol{X}^1 \tilde{\boldsymbol{m}}_g | ... | \boldsymbol{X}^M \tilde{\boldsymbol{m}}_g$, and $\mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{X}_{\boldsymbol{h}}\right]$ and $\mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{W} \boldsymbol{X}_{\boldsymbol{g}}\right]$ are computed in steps E-A and E-C.

$$\frac{\partial \mathrm{E}_{\tilde{p}_h \tilde{p}_{a_j} \tilde{p}_g \tilde{p}_{c_j}}\left[\log p\left(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}, \alpha_j, \boldsymbol{\ell}_j, v_{b_j}\right)\right]}{\partial {\sigma_j}^2} = -N + \sum_j \frac{1}{\hat{v}_{b_j}} V_j = 0 \Rightarrow \hat{v}_{b_j} = \frac{V_j}{N} \tag{E.89}$$

• If we constraint the norm of $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\begin{aligned} V_j = {} & \tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}} \boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{X}_{\tilde{\boldsymbol{h}}} \tilde{\boldsymbol{m}}_{a_j} tr(\tilde{\boldsymbol{\Sigma}}_{a_j} \boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{X}_{\tilde{\boldsymbol{h}}}) + \tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}} \boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{W} \boldsymbol{X}_{\tilde{\boldsymbol{g}}} \tilde{\boldsymbol{m}}_{c_j} \\ & + tr(\tilde{\boldsymbol{\Sigma}}_{c_j} \boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{W} \boldsymbol{X}_{\tilde{\boldsymbol{g}}}) - 2\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}} \boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}} \boldsymbol{\Lambda}_j (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j) - 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}} \boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j) \\ & + 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}} \boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}} \boldsymbol{W}^{\mathrm{t}} \boldsymbol{\Lambda}_j \boldsymbol{X}_{\tilde{\boldsymbol{h}}} \tilde{\boldsymbol{m}}_{a_j} + (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j)^{\mathrm{t}} \boldsymbol{\Lambda}_j (\boldsymbol{y}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j) \end{aligned} \tag{E.90}$$

**M-$\boldsymbol{\ell}, \boldsymbol{\alpha}$**

Parameters $\boldsymbol{\ell}, \boldsymbol{\alpha}$ are the drift weights $\boldsymbol{\ell}_j$ and perfusion baseline $\alpha_j$ for $j \in \mathcal{J}$. Denote $\bar{\boldsymbol{\ell}}_j = \left[\alpha_j, \boldsymbol{\ell}_j\right]^T$ and similarly $\bar{\boldsymbol{\ell}} = [\boldsymbol{\alpha}, \boldsymbol{\ell}]^T$. Denote also by $\bar{\boldsymbol{P}}$ the matrix $\boldsymbol{P}$ added with a first column equal to $\boldsymbol{w}$.

$$\hat{\bar{\boldsymbol{\ell}}} = \arg\max_{\bar{\boldsymbol{\ell}}} \mathrm{E}_{\tilde{p}_h \tilde{p}_a \tilde{p}_g \tilde{p}_c}\left[\log p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{h}, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\alpha}, \boldsymbol{\ell}, v_{b_j})\right] \tag{E.91}$$

For the same reason (independence between $\{\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}, \alpha_j, \boldsymbol{\ell}_j, v_{b_j}\}_{j=1:J}$) we can compute:

$$\hat{\bar{\boldsymbol{\ell}}}_j = \arg\max_{\bar{\boldsymbol{\ell}}_j} \mathrm{E}_{\tilde{p}_h \tilde{p}_{a_j} \tilde{p}_g \tilde{p}_{c_j}}\left[\log p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}, \alpha_j, \boldsymbol{\ell}_j, v_{b_j})\right] \tag{E.92}$$

$$\begin{aligned} & \log p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}, \alpha_j, \boldsymbol{\ell}_j, v_{b_j}) \\ & = C_{\bar{\boldsymbol{\ell}}_j} + \left(\boldsymbol{y}_j - \boldsymbol{X}_{\boldsymbol{h}} \boldsymbol{a}_j + \boldsymbol{W} \boldsymbol{X}_{\boldsymbol{g}} \boldsymbol{c}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j\right)^{\mathrm{t}} \boldsymbol{\Gamma}_j \left(\boldsymbol{y}_j - \boldsymbol{X}_{\boldsymbol{h}} \boldsymbol{a}_j + \boldsymbol{W} \boldsymbol{X}_{\boldsymbol{g}} \boldsymbol{c}_j - \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j\right) \\ & = C_{\bar{\boldsymbol{\ell}}_j} + \bar{\boldsymbol{\ell}}_j^{\mathrm{t}} \bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j - 2\check{\boldsymbol{y}}_j{}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j \end{aligned} \tag{E.93}$$

where $\check{\boldsymbol{y}}_j = \boldsymbol{y}_j - \left(\boldsymbol{X}_{\boldsymbol{h}} \boldsymbol{a}_j + \boldsymbol{W} \boldsymbol{X}_{\boldsymbol{g}} \boldsymbol{c}_j\right)$.
$\mathrm{E}_{\tilde{p}_h \tilde{p}_{a_j} \tilde{p}_g \tilde{p}_{c_j}}\left[\log p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{c}_j, \boldsymbol{g}; \alpha_j, \boldsymbol{\ell}_j, v_{b_j})\right] = C_{\bar{\boldsymbol{\ell}}_j} - 2\bar{\boldsymbol{\ell}}_j^{\mathrm{t}} \bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \mathrm{E}_{\tilde{p}_h \tilde{p}_{a_j} \tilde{p}_g \tilde{p}_{c_j}}\left[\check{\boldsymbol{y}}_j\right] + \bar{\boldsymbol{\ell}}_j^{\mathrm{t}} \bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j$

By deriving the previous equation with respect to $\bar{\boldsymbol{\ell}}_j$ we get:

$$\frac{\partial(.)}{\partial \bar{\boldsymbol{\ell}}_j} = -2\bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \mathrm{E}_{\tilde{p}_h \tilde{p}_a \tilde{p}_g \tilde{p}_c}\left[\check{\boldsymbol{y}}_j\right] + 2\bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}} \bar{\boldsymbol{\ell}}_j = \mathrm{Vect}\, 0 \tag{E.94}$$

$$\begin{aligned} \Rightarrow \quad \hat{\bar{\boldsymbol{\ell}}}_j & = (\bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}})^{-1} \bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \mathrm{E}_{\tilde{p}_h \tilde{p}_a \tilde{p}_g \tilde{p}_c}\left[\check{\boldsymbol{y}}_j\right] \\ & = (\bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}})^{-1} \bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \left(\boldsymbol{y}_j - \mathrm{E}_{\tilde{p}_h}\left[\boldsymbol{X}_{\boldsymbol{h}}\right] \tilde{\boldsymbol{m}}_{a_j^m} - \boldsymbol{W} \mathrm{E}_{\tilde{p}_g}\left[\boldsymbol{X}_{\boldsymbol{g}}\right] \tilde{\boldsymbol{m}}_{c_j^m}\right) \end{aligned} \tag{E.95}$$

• If we constraint the norm of $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\begin{aligned} \Rightarrow \quad \hat{\bar{\boldsymbol{\ell}}}_j & = (\bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}})^{-1} \bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \mathrm{E}_{\tilde{p}_a \tilde{p}_c}\left[\check{\boldsymbol{y}}_j\right] \\ & = (\bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \bar{\boldsymbol{P}})^{-1} \bar{\boldsymbol{P}}^{\mathrm{t}} \boldsymbol{\Gamma}_j \left(\boldsymbol{y}_j - \boldsymbol{X}_{\tilde{\boldsymbol{h}}} \tilde{\boldsymbol{m}}_{a_j^m} - \boldsymbol{W} \boldsymbol{X}_{\tilde{\boldsymbol{g}}} \tilde{\boldsymbol{m}}_{c_j^m}\right) \end{aligned} \tag{E.96}$$

# F JDE Free energy for ASL, using $\boldsymbol{\Omega}$ constraint

The free energy functional

$$\mathcal{F}(\tilde{p},\boldsymbol{\theta}) \quad = \quad \mathrm{E}_{\tilde{p}}\big[\log p(\boldsymbol{y},\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta})\big] + \mathcal{I}(\tilde{p})$$

where $\tilde{p}(\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q}) = \tilde{p}_a(\boldsymbol{a})\tilde{p}_h(\boldsymbol{h})\tilde{p}_c(\boldsymbol{c})\tilde{p}_g(\boldsymbol{g})\tilde{p}_q(\boldsymbol{q})$ and $\mathrm{E}_{\tilde{p}}[.]$ denotes the expectation with respect to $\tilde{p}$ and $\mathcal{I}(\tilde{p}) = -\mathrm{E}_{\tilde{p}}\big[\log \tilde{p}(\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q})\big]$ is the entropy of $\tilde{p}$.

$$\begin{aligned}
\mathcal{F}(\tilde{p},\boldsymbol{\theta}) \quad &= \quad \mathrm{E}_{\tilde{p}}\big[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta})\; p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\theta_a})\; p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\theta_c})\; p(\boldsymbol{q};\boldsymbol{\beta})p(\boldsymbol{\beta};\lambda_\beta) \\
&\qquad p(\boldsymbol{h};v_h)p(v_h;\lambda_h)\; p(\boldsymbol{g}\,|\,\boldsymbol{h};v_g)p(v_g;\lambda_g)\big] + \mathcal{I}(\tilde{p}_a) + \mathcal{I}(\tilde{p}_h) + \mathcal{I}(\tilde{p}_c) + \mathcal{I}(\tilde{p}_g) + \mathcal{I}(\tilde{p}_q) \\
&= \quad \mathrm{E}_{\tilde{p}_a\tilde{p}_h\tilde{p}_c\tilde{p}_g}\big[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta})\big] + \mathrm{E}_{\tilde{p}_h}\big[\log p(\boldsymbol{h};v_h)\big] + \mathrm{E}_{\tilde{p}_h\tilde{p}_g}\big[\log p(\boldsymbol{g}\,|\,\boldsymbol{h};v_g)\big] \\
&\quad +\mathrm{E}_{\tilde{p}_a\tilde{p}_q}\big[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\theta_a})\big] + \mathrm{E}_{\tilde{p}_c\tilde{p}_q}\big[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\theta_c})\big] + \mathrm{E}_{\tilde{p}_q}\big[\log p(\boldsymbol{q};\boldsymbol{\beta})\big] \\
&\quad +p(\boldsymbol{\beta};\lambda_\beta) + p(v_h;\lambda_h) + p(v_g;\lambda_g) + \mathcal{I}(\tilde{p}_a) + \mathcal{I}(\tilde{p}_h) + \mathcal{I}(\tilde{p}_c) + \mathcal{I}(\tilde{p}_g) + \mathcal{I}(\tilde{p}_q)
\end{aligned}$$

• If we constraint the norm of $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\begin{aligned}
\mathcal{F}(\tilde{p},\boldsymbol{\theta}) \quad &= \quad \mathrm{E}_{\tilde{p}_a\tilde{p}_c}\big[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta})\big] + \log p(\boldsymbol{h};v_h) + \log p(\boldsymbol{g}\,|\,\boldsymbol{h};v_g) \\
&\quad +\mathrm{E}_{\tilde{p}_a\tilde{p}_q}\big[\log p(\boldsymbol{a}\,|\,\boldsymbol{q};\boldsymbol{\theta_a})\big] + \mathrm{E}_{\tilde{p}_c\tilde{p}_q}\big[\log p(\boldsymbol{c}\,|\,\boldsymbol{q};\boldsymbol{\theta_c})\big] + \mathrm{E}_{\tilde{p}_q}\big[\log p(\boldsymbol{q};\boldsymbol{\beta})\big] \\
&\quad +p(\boldsymbol{\beta};\lambda_\beta) + p(v_h;\lambda_h) + p(v_g;\lambda_g) + \mathcal{I}(\tilde{p}_a) + \mathcal{I}(\tilde{p}_c) + \mathcal{I}(\tilde{p}_q) + \lambda_{c1}(\|h\|-1) + \lambda_{c2}(\|g\|-1)
\end{aligned}$$

For the computation of each term above:

## F.1 Likelihood term

$$E_{\tilde{p}_a\tilde{p}_h\tilde{p}_c\tilde{p}_g\tilde{p}_q}\left[\log p(\boldsymbol{y}\,|\,\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q};\boldsymbol{\theta})\right] = -\frac{NJ}{2}\log 2\pi + \frac{J}{2}\log|\boldsymbol{\Lambda}_j| - N\sum_{j\in J}\log v_{b_j} + \frac{1}{2v_{b_j}}\sum_{j\in J}V_j$$

$$\begin{aligned}
V_j = \; &\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}}\mathrm{E}_{\tilde{p}_h}\big[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\boldsymbol{h}}\big]\tilde{\boldsymbol{m}}_{a_j} + tr(\tilde{\boldsymbol{\Sigma}}_{a_j}\mathrm{E}_{\tilde{p}_h}\big[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\boldsymbol{h}}\big]) - 2\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}}\mathrm{E}_{\tilde{p}_h}\big[\boldsymbol{X}_{\boldsymbol{h}}^{\mathrm{t}}\big]\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) \\
&+ \tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\mathrm{E}_{\tilde{p}_g}\big[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\big]\tilde{\boldsymbol{m}}_{c_j} + tr(\tilde{\boldsymbol{\Sigma}}_{c_j}\mathrm{E}_{\tilde{p}_g}\big[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\boldsymbol{g}}\big]) - 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\mathrm{E}_{\tilde{p}_g}\big[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\big]\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) \\
&+ 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\mathrm{E}_{\tilde{p}_g}\big[\boldsymbol{X}_{\boldsymbol{g}}^{\mathrm{t}}\big]\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\mathrm{E}_{\tilde{p}_h}\big[\boldsymbol{X}_{\boldsymbol{h}}\big]\tilde{\boldsymbol{m}}_{a_j} + (\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)
\end{aligned}$$

• If we constraint the norm of $\boldsymbol{h}$ and $\boldsymbol{g}$:

$$\begin{aligned}
V_j = \; &\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}}\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\tilde{\boldsymbol{h}}}\tilde{\boldsymbol{m}}_{a_j} + tr(\tilde{\boldsymbol{\Sigma}}_{a_j}\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\tilde{\boldsymbol{h}}}) - 2\tilde{\boldsymbol{m}}_{a_j}^{\mathrm{t}}\boldsymbol{X}_{\tilde{\boldsymbol{h}}}^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) \\
&+ \tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\tilde{\boldsymbol{g}}}\tilde{\boldsymbol{m}}_{c_j} + tr(\tilde{\boldsymbol{\Sigma}}_{c_j}\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{W}\boldsymbol{X}_{\tilde{\boldsymbol{g}}}) - 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j) \\
&+ 2\tilde{\boldsymbol{m}}_{c_j}^{\mathrm{t}}\boldsymbol{X}_{\tilde{\boldsymbol{g}}}^{\mathrm{t}}\boldsymbol{W}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{X}_{\tilde{\boldsymbol{h}}}\tilde{\boldsymbol{m}}_{a_j} + (\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j)
\end{aligned} \tag{F.1}$$

## F.2 Response function terms

$$\mathrm{E}_{\tilde{p}_h}\big[\log p(\boldsymbol{h}; v_h)\big] = -\frac{D+1}{2}\log 2\pi - \frac{D+1}{2}\log v_h - \frac{1}{2}\log|\boldsymbol{\Sigma}_h| - \frac{1}{2v_h}((\tilde{\boldsymbol{m}}_h)^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{m}}_h) + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Sigma}_h^{-1}\right)$$

$$\begin{aligned}\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\big[\log p(\boldsymbol{g}\,|\,\boldsymbol{h}; v_g)\big] = &-\frac{D+1}{2}\log 2\pi - \frac{D+1}{2}\log v_g - \frac{1}{2}\log|\boldsymbol{\Sigma}_g| \\ &- \frac{1}{2v_g}((\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h)^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{m}}_h)) + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g\boldsymbol{\Sigma}_g^{-1} + \tilde{\boldsymbol{\Sigma}}_h\boldsymbol{\Omega}^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\boldsymbol{\Omega}\right)\end{aligned}$$

where the covariance matrices $\boldsymbol{\Sigma}_h$ and $\boldsymbol{\Sigma}_g$ are $\boldsymbol{R}$ if we want to impose smoothness or 0 if we do not.

• If we use the constraint on the norm of $\boldsymbol{h}$ to be 1:

$$\mathrm{E}_{\tilde{p}_h}\big[\log p(\boldsymbol{h}; v_h)\big] = -\frac{D+1}{2}\log 2\pi - \frac{D+1}{2}\log v_h - \frac{1}{2}\log|\boldsymbol{\Sigma}_h| - \frac{1}{2v_h}((\tilde{\boldsymbol{h}})^{\mathrm{t}}\boldsymbol{\Sigma}_h^{-1}\tilde{\boldsymbol{h}})$$

$$\begin{aligned}\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\big[\log p(\boldsymbol{g}\,|\,\boldsymbol{h}; v_g)\big] = &-\frac{D+1}{2}\log 2\pi - \frac{D+1}{2}\log v_g - \frac{1}{2}\log|\boldsymbol{\Sigma}_g| \\ &- \frac{1}{2v_g}((\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\tilde{\boldsymbol{m}}_g - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})) + \mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g\boldsymbol{\Sigma}_g^{-1}\right)\end{aligned}$$

• If we also use the constraint on the norm of $\boldsymbol{g}$ to be 1, then:

$$\mathrm{E}_{\tilde{p}_h\tilde{p}_g}\big[\log p(\boldsymbol{g}\,|\,\boldsymbol{h}; v_g)\big] = -\frac{D+1}{2}\log 2\pi - \frac{D+1}{2}\log v_g - \frac{1}{2}\log|\boldsymbol{\Sigma}_g| - \frac{1}{2v_g}((\tilde{\boldsymbol{g}} - \boldsymbol{\Omega}\tilde{\boldsymbol{h}})^{\mathrm{t}}\boldsymbol{\Sigma}_g^{-1}(\tilde{\boldsymbol{g}} - \boldsymbol{\Omega}\tilde{\boldsymbol{h}}))$$

If hyperprior on $v_h$ and $v_g$ is used:

$$\begin{aligned}\log p(v_h; \lambda_h) &= \log(\lambda_h) - \lambda_h v_h \\ \log p(v_g; \lambda_g) &= \log(\lambda_g) - \lambda_g v_g\end{aligned}$$

otherwise $\log p(v_h; \lambda_h) = 0$ and $\log p(v_g; \lambda_g) = 0$.

## F.3 Response level terms

$$\begin{aligned}E_{\tilde{p}_a\tilde{p}_q}\left[\log p(\boldsymbol{a}\,|\,\boldsymbol{q}, \boldsymbol{\theta}_a)\right] = \sum_m \sum_j \Bigg\{ &\left[1 - \tilde{p}_{q_j^m}(1)\right]\left[\log\frac{1}{\sqrt{2\pi v_0^m}} - \frac{(\tilde{m}_{a_j^m} - \mu_0^m)^2 + \tilde{\Sigma}_{a_j^{m,m}}}{2v_0^m}\right] \\ &+ \tilde{p}_{q_j^m}(1)\left[\log\frac{1}{\sqrt{2\pi v_1^m}} - \frac{(\tilde{m}_{a_j^m} - \mu_1^m)^2 + \tilde{\Sigma}_{a_j^{m,m}}}{2v_1^m}\right]\Bigg\}\end{aligned}$$

$$\begin{aligned}E_{\tilde{p}_c\tilde{p}_q}\left[\log p(\boldsymbol{c}\,|\,\boldsymbol{q}, \boldsymbol{\theta}_c)\right] = \sum_m \sum_j \Bigg\{ &\left[1 - \tilde{p}_{q_j^m}(1)\right]\left[\log\frac{1}{\sqrt{2\pi \rho_0^m}} - \frac{(\tilde{m}_{c_j^m} - \eta_0^m)^2 + \tilde{\Sigma}_{c_j^{m,m}}}{2\rho_0^m}\right] \\ &+ \tilde{p}_{q_j^m}(1)\left[\log\frac{1}{\sqrt{2\pi \rho_1^m}} - \frac{(\tilde{m}_{c_j^m} - \eta_1^m)^2 + \tilde{\Sigma}_{c_j^{m,m}}}{2\rho_1^m}\right]\Bigg\}\end{aligned}$$

## F.4 Labels term

$$E_{\tilde{p}_q}\left[\log p(\boldsymbol{q}\,|\,\boldsymbol{\beta})\right] = \sum_m \left\{-\log Z(\beta^m) + \beta^m E_{\tilde{p}_{q^m}}\left[U(\boldsymbol{q}^m)\right]\right\}$$

$$U(\boldsymbol{q}^m) = \sum_{j\sim k} \mathbb{1}(q_j^m = q_k^m) = \frac{1}{2}\sum_j \sum_{k\in N(j)} (\boldsymbol{q}_j^m)^{\mathrm{t}} \boldsymbol{q}_k^m$$

$$E_{\tilde{p}_{q^m}}\left[U(\boldsymbol{q}^m)\right] = \frac{1}{2}\sum_j \sum_{k\in N(j)} E_{\tilde{p}_{q_j^m}\tilde{p}_{q_k^m}}\left[\mathbb{1}(q_j^m = q_k^m)\right] = \frac{1}{2}\sum_j \sum_{k\in N(j)} \sum_{i=0}^{1} \tilde{p}_{q_j^m}(i)\tilde{p}_{q_k^m}(i)$$

The mean field approximation consists in fixing aleatory neighbours $\boldsymbol{q}_k^m$ to fixed quantities $\tilde{\boldsymbol{q}}_k^m$, allowing the approximation of $U(\boldsymbol{q}^m)$ to $U^{MF}(\boldsymbol{q}^m)$:

$$U^{MF}(\boldsymbol{q}^m) = \frac{1}{2}\sum_j \sum_{k\in N(j)} (\boldsymbol{q}_j^m)^{\mathrm{t}} \tilde{\boldsymbol{q}}_k^m$$

We compute $Z(\beta^m)$ using mean field approximation:

$$Z(\beta^m) \simeq Z^{MF}(\beta^m) \exp\left(\beta^m E_{p^{MF}}\left[U(\boldsymbol{q}^m) - U^{MF}(\boldsymbol{q}^m)\right]\right)$$

$$p^{MF}(q^m) = \frac{\exp\left(\beta^m U^{MF}(\boldsymbol{q}^m)\right)}{Z^{MF}(\beta^m)} = \prod_j \frac{\exp\left(\beta^m \sum\limits_{k\in N_j} \tilde{p}_{q_k^m}(q_j^m)\right)}{\sum\limits_{i\in\{0,1\}} \exp\left(\beta^m \sum\limits_{k\in N_j} \tilde{p}_{q_k^m}(i)\right)} = \prod_j P_j^{MF}(q_j^m)$$

since

$$Z^{MF}(\beta^m) = \prod_j \sum_{i=0}^{1} \exp\left(\beta^m \sum_{k\in N_j} \tilde{q}_k^m(i)\right)$$

Knowing $p^{MF}$, we can compute

$$E_{p^{MF}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_j \sum_{k\in N(j)} \sum_{i=0}^{1} p_j^{MF}(i) p_k^{MF}(i)$$

$$E_{p^{MF}}[U^{MF}(\boldsymbol{q}^m)] = \sum_j \sum_{k\in N(j)} \sum_{i=0}^{1} \tilde{p}_{q_k^m}(i) p_j^{MF}(i)$$

And so when we put it all together

$$Z(\beta^m) \simeq \prod_j \sum_{i=0}^{1} \exp\left(\beta^m \sum_{k\in N(j)} \tilde{p}_{q_k^m}(i)\right) \exp\left(\beta^m \sum_j \sum_{k\in N(j)} \sum_{i=0}^{1} \left[p_j^{MF}(i)\left(\frac{p_k^{MF}(i)}{2} - \tilde{p}_{q_k^m}(i)\right)\right]\right)$$

Therefore, we finally have

$$
\begin{aligned}
E_{\tilde{p}_q}\left[\log p(\boldsymbol{q}\,|\,\boldsymbol{\beta})\right] = \sum_m \Bigg\{ & -\log\left(\prod_j \sum_{i=0}^{1} \exp\left(\beta^m \sum_{k\in N(j)} \tilde{p}_{q_k^m}(i)\right)\right. \\
& \left.\exp\left(\beta^m \sum_j \sum_{k\in N(j)} \sum_{i=0}^{1}\left[p_j^{MF}(i)\left(\frac{p_k^{MF}(i)}{2} - \tilde{p}_{q_k^m}(i)\right)\right]\right)\right) \\
& + \frac{\beta^m}{2} \sum_j \sum_{k\in N(j)} \sum_{i=0}^{1} \tilde{p}_{q_j^m}(i)\tilde{p}_{q_k^m}(i) \Bigg\} \\
= \sum_m \Bigg\{ & -\sum_j \left\{ \log\left(\sum_{i=0}^{1} \exp\left(\beta^m \sum_{k\in N(j)} \tilde{p}_{q_k^m}(i)\right)\right)\right\} \\
& -\beta^m \sum_j \sum_{k\in N(j)} \sum_{i=0}^{1}\left[p_j^{MF}(i)\left(\frac{p_k^{MF}(i)}{2} - \tilde{p}_{q_k^m}(i)\right) - \frac{1}{2}\tilde{p}_{q_j^m}(i)\tilde{p}_{q_k^m}(i)\right] \Bigg\}
\end{aligned}
$$

$$
\log p(\beta;\lambda_\beta) = \sum_{m=1}^{M} \log p(\beta^m;\lambda_\beta) = M\log\lambda_\beta - \lambda_\beta \sum_{m=1}^{M} \beta^m
$$

## F.5 Entropy term

$$
\begin{aligned}
\mathcal{I}(\tilde{p}) &= -\mathrm{E}_{\tilde{p}}\left[\log \tilde{p}(\boldsymbol{a},\boldsymbol{h},\boldsymbol{c},\boldsymbol{g},\boldsymbol{q})\right] = \mathcal{I}(\tilde{p}_a) + \mathcal{I}(\tilde{p}_c) + \mathcal{I}(\tilde{p}_q) + \mathcal{I}(\tilde{p}_h) + \mathcal{I}(\tilde{p}_g) \\
\mathcal{I}(\tilde{p}_a) &= -\sum_j E_{\tilde{p}_{\boldsymbol{a}_j}}[log(\tilde{p}_{\boldsymbol{a}_j})] = \sum_{j=1}^{J} \log(\sqrt{(2\pi e)^M |\tilde{\boldsymbol{\Sigma}}_{a_j}|}) \\
\mathcal{I}(\tilde{p}_c) &= -\sum_j E_{\tilde{p}_{\boldsymbol{c}_j}}[log(\tilde{p}_{\boldsymbol{c}_j})] = \sum_{j=1}^{J} \log(\sqrt{(2\pi e)^M |\tilde{\boldsymbol{\Sigma}}_{c_j}|}) \\
\mathcal{I}(\tilde{p}_q) &= -\sum_j \sum_m E_{\tilde{p}_{q_j^m}}[\log(\tilde{p}_{q_j^m}(\boldsymbol{q}_j^m))] = -\sum_j \sum_m \sum_{i=0}^{1} \tilde{p}_{q_j^m}(i)\log\tilde{p}_{q_j^m}(i) \\
\mathcal{I}(\tilde{p}_h) &= \log(\sqrt{(2\pi e)^{D+1}|\tilde{\boldsymbol{\Sigma}}_h|}) \\
\mathcal{I}(\tilde{p}_g) &= \log(\sqrt{(2\pi e)^{D+1}|\tilde{\boldsymbol{\Sigma}}_g|})
\end{aligned}
$$

# G VEM solution for JDE ASL using Balloon model prior

The priors used for $\boldsymbol{h}$ and $\boldsymbol{g}$ are

$$\boldsymbol{h} \sim \mathcal{N}(\boldsymbol{h}_{balloon}, v_h \boldsymbol{\Sigma}_h)$$
$$\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{g}_{balloon}, v_g \boldsymbol{\Sigma}_g)$$

being $\boldsymbol{\Sigma}_h = \boldsymbol{R}$ and $\boldsymbol{\Sigma}_g = \boldsymbol{R}$.

## G.1 Changes in the model with respect to the $\Omega$ prior in appendix E

**E-H step:**

$\widetilde{p}_h(\boldsymbol{h}) = \mathcal{N}(\tilde{\boldsymbol{m}}_h, \tilde{\boldsymbol{\Sigma}}_h)$, where

$$\tilde{\boldsymbol{\Sigma}}_h^{-1} = \frac{\boldsymbol{\Sigma}_h^{-1}}{v_h} + \sum_{j\in J}\sum_{m,m'} \left(\tilde{m}_{a_j^m} m_{a_j^{m'}} + \tilde{v}_{a_j^m a_j^{m'}}\right) (\boldsymbol{X}^m)^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'}$$

$$\tilde{\boldsymbol{m}}_h = \tilde{\boldsymbol{\Sigma}}_h \left( \frac{\boldsymbol{\Sigma}_h^{-1} \boldsymbol{h}_{balloon}}{v_h} + \sum_{j\in J}\sum_{m=1}^{M} \tilde{m}_{a_j^m} (\boldsymbol{X}^m)^{\mathrm{t}} \boldsymbol{\Gamma}_j \left( \boldsymbol{y}_j - \sum_{m=1}^{M} \tilde{m}_{c_j^m} \boldsymbol{W}\boldsymbol{X}^m \tilde{\boldsymbol{m}}_g - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j \right) \right)$$

**E-G step:**

$\widetilde{p}_g(\boldsymbol{g}) = \mathcal{N}(\tilde{\boldsymbol{m}}_g, \tilde{\boldsymbol{\Sigma}}_g)$, where

$$\tilde{\boldsymbol{\Sigma}}_g^{-1} = \frac{\boldsymbol{\Sigma}_g^{-1}}{v_g} + \sum_{j\in J}\sum_{m,m'} \left(\tilde{m}_{c_j^m} \tilde{m}_{c_j^{m'}} + \tilde{v}_{c_j^m c_j^{m'}}\right) (\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{W}\boldsymbol{X}^{m'}$$

$$\tilde{\boldsymbol{m}}_g = \tilde{\boldsymbol{\Sigma}}_g \left( \frac{\boldsymbol{\Sigma}_g^{-1} \boldsymbol{g}_{balloon}}{v_g} + \sum_{j\in J}\sum_{m=1}^{M} \tilde{m}_{c_j^m} (\boldsymbol{W}\boldsymbol{X}^m)^{\mathrm{t}} \boldsymbol{\Gamma}_j \left( \boldsymbol{y}_j - \sum_{m=1}^{M} \tilde{m}_{a_j^m} \boldsymbol{X}^m \tilde{\boldsymbol{m}}_h - \bar{\boldsymbol{P}}\bar{\boldsymbol{\ell}}_j \right) \right)$$

**M-($v_h$) step:**

$$v_h = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_h \mathrm{tr}\left[ \left( (\tilde{\boldsymbol{m}}_h - \boldsymbol{h}_{balloon})(\tilde{\boldsymbol{m}}_h - \boldsymbol{h}_{balloon})^{\mathrm{t}} + \tilde{\boldsymbol{\Sigma}}_h \right) \boldsymbol{\Sigma}_h^{-1} \right]}}{4\lambda_h}$$

**M-($v_g$) step:**

$$v_g = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_g \mathrm{tr}\left[ \left( (\tilde{\boldsymbol{m}}_g - \boldsymbol{g}_{balloon})(\tilde{\boldsymbol{m}}_g - \boldsymbol{g}_{balloon})^{\mathrm{t}} + \tilde{\boldsymbol{\Sigma}}_g \right) \boldsymbol{\Sigma}_g^{-1} \right]}}{4\lambda_g}$$

# Bibliography


A.-L. Fouque, P. Ciuciu, and L. Risser. Multivariate spatial Gaussian mixture modeling for statistical clustering of hemodynamic parameters in functional MRI. In *34th Proc. IEEE ICASSP*, pages 445–448, Taipei, Taiwan, Apr. 2009.

A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Muller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux. Machine learning for neuroimaging with scikit-learn. *arXiv preprint arXiv:1412.3919*, 2014.

G. K. Aguirre, J. A. Detre, and J. Wang. Perfusion fMRI for functional neuroimaging. *Int Rev Neurobiol*, 66:213–236, 2005.

M. Albughdadi, L. Chaari, F. Forbes, J.-Y. Tourneret, and P. Ciuciu. Model selection for hemodynamic brain parcellation in fMRI. In *2014 22nd European Signal Processing Conference (EUSIPCO)*, pages 31–35. IEEE, 2014.

M. Albughdadi, L. Chaari, F. Forbes, J.-Y. Tourneret, and P. Ciuciu. Multi-subject joint parcellation detection estimation in functional MRI. In *13th IEEE International Symposium on Biomedical Imaging*, 2016a.

M. Albughdadi, L. Chaari, J.-Y. Tourneret, F. Forbes, and P. Ciuciu. Hemodynamic brain parcellation using a non-parametric Bayesian approach. 2016b.

D. C. Alsop, J. A. Detre, X. Golay, M. Günther, J. Hendrikse, L. Hernandez-Garcia, H. Lu, B. J. MacIntosh, L. M. Parkes, M. Smits, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: A consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magnetic resonance in medicine*, 73(1):102–116, 2015.

D. Attwell, A. M. Buchan, S. Charpak, M. Lauritzen, B. A. MacVicar, and E. A. Newman. Glial and neuronal control of brain blood flow. *Nature*, 468(7321):232–243, 2010.

A. Attyé, M. Villien, F. Tahon, J. Warnking, O. Detante, and A. Krainik. Normalization of cerebral vasoreactivity using BOLD MRI after intravascular stenting. *Human brain mapping*, 35(4):1320–1324, 2014.

B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Hum. Brain Mapp.*, 27(8):678–693, Aug. 2006.

S. Badillo, G. Varoquaux, and P. Ciuciu. Hemodynamic estimation based on consensus clustering. In *IEEE Pattern Recognition in Neuroimaging (PRNI)*, Philadelphia, USA, June 2013a.

S. Badillo, T. Vincent, and P. Ciuciu. Multi-session extension of the joint-detection framework in fMRI. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1512–1515. IEEE, 2013b.

P. B. Barker, X. Golay, and G. Zaharchuk. *Clinical perfusion MRI: techniques and applications*. Cambridge University Press, 2013.

M. J. Beal. *Variational algorithms for approximate Bayesian inference*. University of London, London, 2003.

Y. Behzadi and T. Liu. An arteriolar compliance model of the cerebral blood flow response to neural stimulus. *NeuroImage*, 25(4):1100–1111, May 2005.

P. Bellec, P. Rosa-Neto, O. C. Lyttelton, H. Benali, and A. C. Evans. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *Neuroimage*, 51(3):1126–1139, 2010.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

C. M. Bishop. Pattern recognition. *Machine Learning*, 128, 2006.

S. Blanchard, T. Papadopoulo, C.-G. Bénar, N. Voges, M. Clerc, H. Benali, J. Warnking, O. David, and F. Wendling. Relationship between flow and metabolism in BOLD signals: insights from biophysical models. *Brain topography*, 24(1):40–53, 2011.

K. Brodmann. *Brodmann's: Localisation in the cerebral cortex*. Springer Science & Business Media, 2007.

E. E. Bron, R. M. Steketee, G. C. Houston, R. A. Oliver, H. C. Achterberg, M. Loog, J. C. Swieten, A. Hammers, W. J. Niessen, M. Smits, et al. Diagnostic classification of arterial spin labeling and structural MRI in presenile early stage dementia. *Human brain mapping*, 35(9):4916–4931, 2014.

R. B. Buxton. The physics of functional magnetic resonance imaging (fMRI). *Reports on Progress in Physics*, 76(9): 096601, 2013.

R. B. Buxton, L. R. Frank, E. C. Wong, B. Siewert, S. Warach, and R. R. Edelman. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magnetic Resonance in Medicine*, 40(3):383–396, 1998a.

R. B. Buxton, E. C. Wong, and F. L. R. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.*, 39:855–864, June 1998b.

R. B. Buxton, K. Uludağ, D. J. Dubowitz, and T. T. Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23:S220–S233, 2004.

G. Bydder and I. Young. MR imaging: clinical use of the inversion recovery sequence. *Journal of computer assisted tomography*, 9(4):659–675, 1985.

S. Cantin, M. Villien, O. Moreaud, I. Tropres, S. Keignart, E. Chipon, J.-F. Le Bas, J. Warnking, and A. Krainik. Impaired cerebral vasoreactivity to CO 2 in Alzheimer's disease using BOLD fMRI. *Neuroimage*, 58(2):579–587, 2011.

G. Celeux, F. Forbes, and N. Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pat. Rec.*, 36(1):131–144, 2003.

L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu. Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach. *IEEE Trans. on Medical Imaging*, 32(5):821–837, May 2013.

L. Chaari, S. Badillo, T. Vincent, G. Dehaene-Lambertz, F. Forbes, and P. Ciuciu. Subject-level joint parcellation-detection-estimation in fMRI. 2016.

Y. Chen, H. Wan, J. O'Reardon, D. Wang, Z. Wang, M. Korczykowski, and J. Detre. Quantification of cerebral blood flow as biomarker of drug effect: Arterial spin labeling phMRI after a single dose of oral citalopram. *Clinical Pharmacology & Therapeutics*, 89(2):251–258, 2011.

R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

J. A. Detre and J. Wang. Technical aspects and utility of fMRI using BOLD and ASL. *Clinical Neurophysiology*, 113 (5):621–634, 2002.

J. A. Detre, J. S. Leigh, D. S. Williams, and A. P. Koretsky. Perfusion imaging. *Magnetic resonance in medicine*, 23 (1):37–45, 1992.

J. A. Detre, H. Rao, D. J. Wang, Y. F. Chen, and Z. Wang. Applications of arterial spin labeled MRI in the brain. *Journal of Magnetic Resonance Imaging*, 35(5):1026–1037, 2012.

P. J. Drew, A. Y. Shih, and D. Kleinfeld. Fluctuating and sensory-induced vasodynamics in rodent cortex extend arteriole capacity. *Proceedings of the National Academy of Sciences*, 108(20):8473–8478, 2011.

M. D'haeseleer, R. Beelen, Y. Fierens, M. Cambron, A.-M. Vanbinst, C. Verborgh, J. Demey, and J. De Keyser. Cerebral hypoperfusion in multiple sclerosis is reversible and mediated by endothelin-1. *Proceedings of the National Academy of Sciences*, 110(14):5654–5658, 2013.

R. R. Edelman, B. Siewert, D. Darby, V. Thangaraj, A. Nobre, M. Mesulam, and S. Warach. Qualitative mapping of cerebral blood flow and functional localization with echo-planar MR imaging and signal targeting with alternating radio frequency. *Radiology*, 192(2):513–520, 1994.

M. Fabiani, B. A. Gordon, E. L. Maclin, M. A. Pearson, C. R. Brumback-Peltz, K. A. Low, E. McAuley, B. P. Sutton, A. F. Kramer, and G. Gratton. Neurovascular coupling in normal aging: a combined optical, ERP and fMRI study. *Neuroimage*, 85:592–607, 2014.

P. M. Figueiredo, S. Clare, and P. Jezzard. Quantitative perfusion measurements using pulsed arterial spin labeling: Effects of large region-of-interest analysis. *Journal of magnetic resonance imaging*, 21(6):676–682, 2005.

G. Flandin and W. D. Penny. Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage*, 34 (3):1108–1125, 2007.

R. S. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, S. Zeki, J. T. Ashburner, and W. D. Penny. *Human brain function*. Academic press, 2004.

E. K. Fram, R. J. Herfkens, G. A. Johnson, G. H. Glover, J. P. Karis, A. Shimakawa, T. G. Perkins, and N. J. Pelc. Rapid calculation of T1 using variable flip angle gradient refocused imaging. *Magnetic resonance imaging*, 5(3):201–208, 1987.

A. Frau-Pascual, T. Vincent, F. Forbes, and P. Ciuciu. Hemodynamically informed parcellation of cerebral fMRI data. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2079–2083. IEEE, 2014a.

A. Frau-Pascual, T. Vincent, J. Sloboda, P. Ciuciu, and F. Forbes. Physiologically informed Bayesian analysis of ASL fMRI data. In *Bayesian and grAphical Models for Biomedical Imaging*, pages 37–48. Springer, 2014b.

A. Frau-Pascual, F. Forbes, and P. Ciuciu. Physiological models comparison for the analysis of ASL fMRI data. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 1348–1351. IEEE, 2015a.

A. Frau-Pascual, F. Forbes, and P. Ciuciu. Variational physiologically informed solution to hemodynamic and perfusion response estimation from ASL fMRI data. In *Pattern Recognition in NeuroImaging (PRNI), 2015 International Workshop on*, pages 57–60. IEEE, 2015b.

A. Frau-Pascual, F. Forbes, and P. Ciuciu. Comparison of stochastic and variational solutions to ASL fMRI data analysis. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 85–92. Springer, 2015c.

K. Friston and W. Penny. Posterior probability maps and SPMs. *Neuroimage*, 19(3):1240–1249, 2003.

K. Friston, P. Fletcher, O. Joseph, A. Holmes, M. Rugg, and R. Turner. Event-related responses in fMRI: characterising differential responses. *Neuroimage*, 7:30–40, 1998.

K. J. Friston, A. Mechelli, R. Turner, and C. J. Price. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage*, 12:466–477, June 2000.

K. J. Friston, D. E. Glaser, R. N. Henson, S. Kiebel, C. Phillips, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Applications. *Neuroimage*, 16(2):484–512, 2002a.

K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage*, 16(2):465–483, 2002b.

G. Flandin, F. Kherif, X. Pennec, D. Rivière, N. Ayache, and J.-B. Poline. A new representation of fMRI data using anatomo-functional constraints. In *Proc. 8th HBM*, Sendai, Japan, June 2002.

D. Garcia, C. De Bazelaire, and D. Alsop. Pseudo-continuous flow driven adiabatic inversion for arterial spin labeling. In *Proc Int Soc Magn Reson Med*, volume 13, page 37, 2005.

C. J. Gauthier, C. Madjar, L. Desjardins-Crépeau, P. Bellec, L. Bherer, and R. D. Hoge. Age dependence of hemodynamic response characteristics in human functional magnetic resonance imaging. *Neurobiology of aging*, 34(5):1469–1485, 2013.

C. R. Genovese, N. A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Introducing Markov chain Monte Carlo. *Markov chain Monte Carlo in practice*, 1:19, 1996.

G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9:416–429, 1999.

X. Golay and M. Guenther. Arterial spin labelling: final steps to make it a clinical reality. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 25(2):79–82, 2012.

X. Golay, J. Hendrikse, and T. C. Lim. Perfusion imaging using arterial spin labeling. *Topics in Magnetic Resonance Imaging*, 15(1):10–27, 2004.

C. Gössl, L. Fahrmeir, and D. Auer. Bayesian modeling of the hemodynamic response function in BOLD fMRI. *Neuroimage*, 14(1):140–148, 2001.

M. Grade, J. H. Tamames, F. Pizzini, E. Achten, X. Golay, and M. Smits. A neuroradiologist's guide to arterial spin labeling MRI in clinical practice. *Neuroradiology*, 57(12):1181–1202, 2015.

F. Hamzei, R. Knab, C. Weiller, and J. Röther. The influence of extra-and intracranial artery disease on the BOLD signal in fMRI. *Neuroimage*, 20(2):1393–1399, 2003.

D. A. Handwerker, J. M. Ollinger, and D. Mark. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21:1639–1651, 2004.

L. Harrison, W. D. Penny, and K. Friston. Multivariate autoregressive modeling of fMRI time series. *Neuroimage*, 19(4):1477–1491, 2003.

M. Havlicek, K. J. Friston, J. Jan, M. Brazdil, and V. D. Calhoun. Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. *NeuroImage*, 56(4):2109–2128, 2011.

R. Henson, M. Rugg, and K. J. Friston. The choice of basis functions in event-related fMRI. *Neuroimage*, 13(6): 149–149, 2001.

L. Hernandez-Garcia, H. Jahanian, and D. B. Rowe. Quantitative analysis of arterial spin labeling fMRI data using a general linear model. *Magnetic Resonance Imaging*, 28(7):919–927, 2010.

A. P. Holmes, R. Blair, J. Watson, and I. Ford. Nonparametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow & Metabolism*, 16(1):7–22, 1996.

C. Huneau, H. Benali, and H. Chabriat. Investigating human neurovascular coupling using functional neuroimaging: A critical review of dynamic models. *Frontiers in neuroscience*, 9, 2015.

C. Iadecola and M. Nedergaard. Glial regulation of the cerebral microvasculature. *Nature neuroscience*, 10(11): 1369–1376, 2007.

M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith. FSL. *Neuroimage*, 62(2):782–790, 2012.

M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

S. D. Kamvar, D. Klein, and C. D. Manning. Interpreting and extending classical agglomerative clustering algorithms using a model-based approach. In *ICML*, 2002.

S. Kemeny, F. Q. Ye, R. Birn, and A. R. Braun. Comparison of continuous overt speech fMRI using BOLD and arterial spin labeling. *Human brain mapping*, 24(3):173–183, 2005.

I. Khalidov, J. Fadili, F. Lazeyras, D. Van De Ville, and M. Unser. Activelets: Wavelets for sparse representation of hemodynamic responses. *Signal Processing*, 91(12):2810–2821, December 2011.

J. H. Kim, R. Khan, J. K. Thompson, and D. Ress. Model of the transient neurovascular response based on prompt arterial dilation. *Journal of Cerebral Blood Flow & Metabolism*, 33(9):1429–1439, 2013.

J. Kindler, P. Homan, K. Jann, A. Federspiel, R. Flury, M. Hauf, W. Strik, T. Dierks, and D. Hubl. Reduced neuronal activity in language-related regions after transcranial magnetic stimulation therapy for auditory verbal hallucinations. *Biological psychiatry*, 73(6):518–524, 2013.

A. Krainik, M. Hund-Georgiadis, S. Zysset, and D. Y. Von Cramon. Regional impairment of cerebrovascular reactivity and BOLD signal in adults after stroke. *Stroke*, 36(6):1146–1152, 2005.

A. Krainik, M. Villien, I. Tropres, A. Attyé, L. Lamalle, J. Bouvier, J. Pietras, S. Grand, J.-F. Le Bas, and J. Warnking. Functional imaging of cerebral perfusion. *Diagnostic and interventional imaging*, 94(12):1259–1278, 2013.

K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, and R. Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679, 1992.

L. Chaari, F. Forbes, T. Vincent, and P. Ciuciu. Adaptive hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework. In *15th Proc. MICCAI*, LNCS 7512, (Part III), pages 180–188, Nice, France, Oct. 2012. Springer Verlag.

O. Leontiev and R. B. Buxton. Reproducibility of BOLD, perfusion, and CMRO 2 measurements with calibrated-BOLD fMRI. *Neuroimage*, 35(1):175–184, 2007.

T. T. Liu and G. G. Brown. Measurement of cerebral perfusion with arterial spin labeling: Part 1. Methods. *Journal of the International Neuropsychological Society*, 13:517–525, 2007. ISSN 1469-7661.

T. T. Liu and E. C. Wong. A signal processing model for arterial spin labeling functional MRI. *Neuroimage*, 24(1): 207–215, 2005.

N. K. Logothetis and B. A. Wandell. Interpreting the BOLD signal. *Annu. Rev. Physiol.*, 66:735–769, 2004.

N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150–157, 2001.

S. Lorthois, F. Cassot, and F. Lauwers. Simulation study of brain blood flow regulation by intra-cortical arterioles in an anatomically accurate large human vascular network. Part II: flow variations induced by global or localized modifications of arteriolar diameters. *Neuroimage*, 54(4):2840–2853, 2011.

W.-M. Luh, E. C. Wong, P. A. Bandettini, J. S. Hyde, et al. QUIPSS II with thin-slice TI 1 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling. *Magnetic resonance in medicine*, 41(6):1246–1254, 1999.

W.-M. Luh, E. C. Wong, P. A. Bandettini, B. D. Ward, and J. S. Hyde. Comparison of simultaneously measured perfusion and BOLD signal increases during brain activation with T1-based tissue identification. *Magnetic resonance in medicine*, 44(1):137–143, 2000.

S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline. Joint detection-estimation of brain activity in functional MRI: a multichannel deconvolution solution. *IEEE Transactions on Signal Processing*, 53(9):3488–3502, 2005.

S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3):941–969, 2008.

P. Mesejo, S. Saillet, O. David, C. Bénar, J. M. Warnking, and F. Forbes. Estimating biophysical parameters from BOLD signals through evolutionary-based optimization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 528–535. Springer, 2015.

P. Mesejo, S. Saillet, O. David, C. G. Bénar, J. M. Warnking, and F. Forbes. A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):416–427, 2016.

J. A. Mumford, L. Hernandez-Garcia, G. R. Lee, and T. E. Nichols. Estimation efficiency and statistical power in arterial spin labeling fMRI. *Neuroimage*, 33(1):103–114, 2006.

R. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*, pages 355–368. Springer, 1998.

T. Nichols and S. Hayasaka. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research*, 12(5):419–446, 2003.

T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2002.

L. E. Nordin, T.-Q. Li, J. Brogren, P. Johansson, N. Sjögren, K. Hannesdottir, C. Björk, M. Segerdahl, D. J. Wang, and P. Julin. Cortical responses to amphetamine exposure studied by pCASL MRI and pharmacokinetic/pharmacodynamic dose modeling. *NeuroImage*, 68:75–82, 2013.

U. Nöth, G. E. Meadows, F. Kotajima, R. Deichmann, D. R. Corfield, and R. Turner. Cerebral vascular response to hypercapnia: determination with perfusion MRI at 1.5 and 3.0 Tesla using a pulsed arterial spin labeling technique. *Journal of Magnetic Resonance Imaging*, 24(6):1229–1235, 2006.

T. Obata, T. T. Liu, K. L. Miller, W.-M. Luh, E. C. Wong, L. R. Frank, and R. B. Buxton. Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: application of the balloon model to the interpretation of BOLD transients. *NeuroImage*, 21(1):144–153, 2004.

S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

S. Ogawa, R. Menon, D. Tank, S. Kim, H. Merkle, J. Ellermann, and K. Ugurbil. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. a comparison of signal characteristics with a biophysical model. *Biophysical journal*, 64(3):803, 1993.

P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, Ch. Pallier, and H. Benali. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Trans. Med. Imag.*, 22(10):1235–1251, Oct. 2003.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

W. Penny, S. Kiebel, and K. Friston. Variational Bayesian inference for fMRI time series. *NeuroImage*, 19(3):727–741, 2003.

W. D. Penny and K. J. Friston. Classical and Bayesian inference in fMRI, 2004.

W. D. Penny, N. J. Trujillo-Barreto, and K. J. Friston. Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, 24(2):350–362, 2005.

W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Academic press, 2011.

M. A. Pimentel, P. Vilela, I. Sousa, and P. Figueiredo. Localization of the hand motor area by arterial spin labeling and blood oxygen level-dependent functional magnetic resonance imaging. *Human brain mapping*, 34(1):96–108, 2013.

P. Pinel, B. Thirion, S. Meriaux, A. Jobert, J. Serres, D. Le Bihan, J.-B. Poline, and S. Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC neuroscience*, 8(1):1, 2007.

J. Pollock, A. Deibler, C. Whitlow, H. Tan, R. Kraft, J. Burdette, and J. Maldjian. Hypercapnia-induced cerebral hyperperfusion: an underrecognized clinical entity. *American Journal of Neuroradiology*, 30(2):378–385, 2009.

J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.

H. Raoult, J. Petr, E. Bannier, A. Stamm, J.-Y. Gauvrit, C. Barillot, and J.-C. Ferré. Arterial spin labeling for motor activation mapping at 3T with a 32-channel coil: reproducibility and spatial accuracy in comparison with BOLD fMRI. *Neuroimage*, 58(1):157–167, 2011.

J. Richiardi, A. Altmann, A.-C. Milazzo, C. Chang, M. M. Chakravarty, T. Banaschewski, G. J. Barker, A. L. Bokde, U. Bromberg, C. Büchel, et al. Correlated gene expression supports synchronous activity in brain networks. *Science*, 348(6240):1241–1244, 2015.

C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

G. O. Roberts. Markov chain concepts related to sampling algorithms. *Markov chain Monte Carlo in practice*, 57, 1996.

A. Roche. A four-dimensional registration algorithm with application to joint correction of motion and slice timing in fMRI. *IEEE transactions on medical imaging*, 30(8):1546–1554, 2011.

A. Roche, S. Mériaux, M. Keller, and B. Thirion. Mixed-effect statistics for group analysis in fMRI: a nonparametric maximum likelihood approach. *Neuroimage*, 38(3):501–510, 2007.

V. J. Schmithorst, L. Hernandez-Garcia, J. Vannest, A. Rajagopal, G. Lee, and S. K. Holland. Optimized simultaneous ASL and BOLD functional imaging of the whole brain. *Journal of Magnetic Resonance Imaging*, 39(5):1104–1117, 2014.

G. Schumann, E. Loth, T. Banaschewski, A. Barbot, G. Barker, C. Büchel, P. Conrod, J. Dalley, H. Flor, J. Gallinat, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular psychiatry*, 15(12):1128–1139, 2010.

R. Sladky, K. J. Friston, J. Tröstl, R. Cunnington, E. Moser, and C. Windischberger. Slice-timing effects and their correction in functional MRI. *Neuroimage*, 58(2):588–594, 2011.

K. E. Stephan, N. Weiskopf, P. M. Drysdale, P. A. Robinson, and K. J. Friston. Comparing hemodynamic models with DCM. *Neuroimage*, 38(3):387–401, Nov. 2007.

T. Vincent, P. Ciuciu, and B. Thirion. Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI. In *5th Proc. IEEE ISBI*, pages 568–571, Paris, France, May 2008.

B. Thirion, G. Varoquaux, E. Dohmatob, and J.-B. Poline. Which fMRI clustering gives good brain parcellations? *Frontiers in neuroscience*, 8:167, 2014.

T. Tjandra, J. C. Brooks, P. Figueiredo, R. Wise, P. M. Matthews, and I. Tracey. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. *Neuroimage*, 27(2):393–401, 2005.

D. G. Tzikas, A. C. Likas, and N. P. Galatsanos. The variational approximation for Bayesian inference. *IEEE Signal Processing Magazine*, 25(6):131–146, 2008.

N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 562–573. Springer, 2011.

M. Villien, P. Bouzat, T. Rupp, P. Robach, L. Lamalle, I. Tropres, F. Esteve, A. Krainik, P. Lévy, J. M. Warnking, et al. Changes in cerebral blood flow and vasoreactivity to CO 2 measured by arterial spin labeling after 6days at 4350m. *Neuroimage*, 72:272–279, 2013.

T. Vincent, L. Risser, and P. Ciuciu. Spatially adaptive mixture modeling for analysis of within-subject fMRI time series. *IEEE Trans. on Medical Imaging*, 29(4):1059–1074, April 2010.

T. Vincent, F. Forbes, and P. Ciuciu. Bayesian BOLD and perfusion source separation and deconvolution from functional ASL imaging. In *38th Proc. IEEE ICASSP*, pages 1003–1007, Vancouver, Canada, May 2013a.

T. Vincent, J. Warnking, M. Villien, A. Krainik, P. Ciuciu, and F. Forbes. Bayesian Joint Detection-Estimation of cerebral vasoreactivity from ASL fMRI data. In *16th Proc. MICCAI*, volume 2, pages 616–623, Nagoya, Japan, September 2013b.

D. J. Wang, Y. Chen, M. A. Fernández-Seara, and J. A. Detre. Potentials and challenges for arterial spin labeling in pharmacological magnetic resonance imaging. *Journal of Pharmacology and Experimental Therapeutics*, 337(2): 359–366, 2011a.

J. Wang, G. K. Aguirre, D. Y. Kimberg, A. C. Roc, L. Li, and J. A. Detre. Arterial spin labeling perfusion fMRI with very low task frequency. *Magnetic resonance in medicine*, 49(5):796–802, 2003.

Y. Wang, A. J. Saykin, J. Pfeuffer, C. Lin, K. M. Mosier, L. Shen, S. Kim, and G. D. Hutchins. Regional reproducibility of pulsed arterial spin labeling perfusion imaging at 3T. *NeuroImage*, 54(2):1188–1195, 2011b.

Z. Wang, G. K. Aguirre, H. Rao, J. Wang, M. A. Fernández-Seara, A. R. Childress, and J. A. Detre. Empirical optimization of ASL data analysis using an ASL data processing toolbox: ASLtbx. *Magnetic resonance imaging*, 26 (2):261–269, 2008.

J. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244, 1963.

D. Williams, J. Detre, J. Leigh, and A. Koretsky. Magnetic resonance imaging of perfusion using spin inversion of arterial water. *Proceedings of the National Academy of Sciences*, 89(1):212–216, 1992.

D. A. Wolk and J. A. Detre. Arterial spin labeling MRI: an emerging biomarker for Alzheimer's disease and other neurodegenerative conditions. *Current opinion in neurology*, 25(4):421, 2012.

E. C. Wong, M. Cronin, W.-C. Wu, B. Inglis, L. R. Frank, and T. T. Liu. Velocity-selective arterial spin labeling. *Magnetic Resonance in Medicine*, 55(6):1334–1341, 2006.

M. Woolrich and T. Behrens. Variational Bayes inference of spatial mixture models for segmentation. *IEEE Trans. Med. Imag.*, 25(10):1380–1391, October 2006.

M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith. Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage*, 14(6):1370–1386, 2001.

M. W. Woolrich, M. Jenkinson, J. M. Brady, and S. M. Smith. Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE transactions on medical imaging*, 23(2):213–231, 2004.

K. J. Worsley, A. C. Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918, 1992.

Y. Yang. Perfusion MR imaging with pulsed arterial spin-labeling: Basic principles and applications in functional brain imaging. *Concepts in Magnetic Resonance*, 14(5):347–357, 2002.

V. L. Yarnykh. Actual flip-angle imaging in the pulsed steady state: a method for rapid three-dimensional mapping of the transmitted radiofrequency field. *Magnetic resonance in Medicine*, 57(1):192–200, 2007.