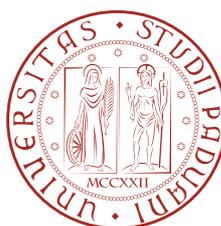


UNIVERSITÀ DEGLI STUDI DI PADOVA
FACOLTÀ DI INGEGNERIA



UNIVERSITÀ DEGLI STUDI DI PADOVA
FACOLTÀ DI INGEGNERIA

—
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

—
TESI DI LAUREA IN BIOINGEGNERIA

Variational Bayesian inference for quantification of brain PET data at the voxel level

Relatore: PROF.SSA ALESSANDRA BERTOLDO

Correlatori: DOTT.SSA GAIA RIZZO

DOTT. MARCO CASTELLARO

Laureando: SIMONE ZANONI

ANNO ACCADEMICO 2014-2015

To my parents

“ ...solo un ingegno infelice imita sempre, senza trarre mai nulla da sé. ”

A. POLIZIANO

Contents

1	Introduction	3
2	Methods	5
2.1	Quantification Methods	5
2.1.1	Variational Bayesian Inference	5
2.1.1.1	Bayesian Inference	6
2.1.1.2	Variational Approach	7
2.1.1.3	Conjugate-Exponential Restriction	9
2.1.1.4	Convergence	9
2.1.1.5	Hyper-parameters Definition	11
2.1.1.6	Update Equations and Free Energy	12
2.1.1.7	Priors Setting	13
2.1.2	Weighted Non-linear Least Squares Method	15
2.1.2.1	Implementation	16
2.2	Compartmental Models and Quantification	16
2.2.1	The 2TCM	17
2.2.2	The 2TCM-1K	18
2.3	Performances Evaluation and Methods Comparison on Real Data	18
2.4	Simulation Studies	19
2.4.1	Sensitivity Test	19
2.4.2	Study on Simulated PET Data	20

3	Datasets	23
3.1	[¹¹ C](<i>R</i>)-rolipram	23
3.2	[¹¹ C]WAY100635	24
3.3	[¹¹ C]PBR28	25
3.4	Measurement Error	26
4	Results	29
4.1	Simulation Studies Results	29
4.1.1	Sensitivity Test Results	29
4.1.2	Study on Simulated PET Data Results	30
4.2	Impact of Precision Priors of the Model Parameters	31
4.3	Real Data Results	32
4.3.1	[¹¹ C](<i>R</i>)-rolipram Results	32
4.3.2	[¹¹ C]WAY100635 Results	35
4.3.3	[¹¹ C]PBR28 Results	38
5	Discussion	43
5.1	Sensitivity Test Discussion	43
5.2	Study on Simulated PET Data Discussion	43
5.3	Impact of Prior Precision of Model Parameters	44
5.4	Quantification Models Discussion	45
5.4.1	The 2TCM Quantification Discussion	45
5.4.2	The 2TCM-1K Quantification Discussion	45
A	Derivation of the Update Equations	47
A.1	Forward Model Parameters	47
A.2	Noise Precision Parameters	49
B	Levenberg-Marquardt Method	51
	Bibliography	53

Chapter 1

Introduction

Positron Emission Tomography (PET) is a functional nuclear medicine imaging technique widely used to study *in vivo* physiological processes in the body. The image acquisition procedure requires a bolus injection of a radioligand into a patient, then pairs of gamma rays generated after annihilation of any emitted positron with any electron of the surrounding material are detected. The four-dimensional images generated by a reconstruction algorithm and corrected for the physical effects involved represent the concentration measurement of the tracer in the districts. This concentration can be related to a specific physiological process targeted by an appropriate tracer, as those used in this study: [^{11}C](*R*)-rolipram, [^{11}C]WAY100635, [^{11}C]PBR28.

The analysis of the data can be applied either at the Region of Interest (ROI) level, where the ROI Time Activity Curve (TAC) is generated by averaging the activity concentration of the specific ROI, or at the voxel level. Since ROI TACs are characterized by a good Signal-to-Noise Ratio (SNR), at the ROI level data are quantified using non-linear estimators, such as Weighted Non-Linear Least Squares (WNLLS). However, the use of ROI TACs causes a loss of spatial resolution. Voxel level analysis produces parametric maps having the same spatial resolution as the original PET image. However the SNR is often so low that the use of non-linear estimators is difficult and unwieldy because of their computational cost. For this reason, alternative quantification approaches are called for.

The most adopted approach for the quantification at the voxel level is the Basis Function Method (BFM) [1]. This method bypasses the problem of non-linearity of the models. However, the setting procedure can heavily penalize the results in terms of bias in the final estimates.

Other available approaches for quantification at the voxel level are the so called “population approaches”. The performances of these methods are sensitive to the goodness of the initial values and to the noise level of data.

Bayesian methods, such as Maximum-a-Posteriori (MAP), are available as well. Their main limitation concerns the computational cost, which is comparable to what required by non-linear estimators.

The purpose of this work is to evaluate, at the voxel level, the performance of a Bayesian method never used before in PET domain that proves powerful performance, even in terms of computation time, in analysing MR data [8]: the Variational Bayes method, which is a Bayesian approach which returns the complete posterior distribution of the parameters. The study includes analysis on both simulated and real data. Therefore, are here presented the equations necessary to attain a weighted estimation without additional computation time and is proposed an algorithm to set the prior parameters.

Chapter 2

Methods

2.1 Quantification Methods

In this section two different estimation methods are presented: Variational Bayes (VB) and Weighted Non-Linear Least Squares (WNLLS). A comparison between them is made over three different datasets, which were made available by Molecular Imaging Branch of National Institute of Mental Health (USA) and by Division of Experimental Medicine of Imperial College of London (UK).

2.1.1 Variational Bayesian Inference

Bayesian methods, which are based on Bayes' theorem, have proved powerful in many applications for the statistical inference of model parameters. However, in practice the computations required are intractable, even for simple cases. Therefore, these methods are either significantly approximated, e.g. Laplace approximation [2, 3, 4], or they achieve samples from exact solution, e.g. Markov Chain Monte Carlo (MCMC) method, which is computationally intense and so not practical for most brain imaging applications [5, 6]. An efficient approximated method has been proposed to facilitate analytical calculations of the posterior distribution over a model: Variational Bayes (VB), also known as Ensemble Learning. This approach takes its name from Feynman's variational free energy method de-

veloped in statistical physics and has been applied in a variety of statistical and signal processing domains. It is now also widely used in neuroimaging, especially fMRI data [7], and in this work, for the first time, in the analysis of PET data. In this Section 2.1.1 are described, first of all the bayesian inference, then the variational approach and the hypothesis that make the problem tractable. Finally, a convergence algorithm is also proposed.

2.1.1.1 Bayesian Inference

As previously said, Bayesian methods are based on Bayes' rule:

$$P(\mathbf{p}|\mathbf{y}, M) = \frac{P(\mathbf{y}, \mathbf{p}|M)}{P(\mathbf{y}|M)} = \frac{P(\mathbf{y}|\mathbf{p}, M)P(\mathbf{p}|M)}{P(\mathbf{y}|M)}, \quad (2.1)$$

where \mathbf{y} denotes a series of measurements to be used to determine the parameters \mathbf{p} of a chosen model M . The Equation 2.1 gives the *posterior* probability of the parameters given the data and the model in terms of: the *likelihood* of the data given the model and its parameters, $P(\mathbf{y}|\mathbf{p}, M)$, the *prior* probability of the parameters given the model, $P(\mathbf{p}|M)$, and the *evidence* for the measurements given the model, $P(\mathbf{y}|M)$. As said, it may not be possible to reach analytically the goal of the Bayesian inference, that is to evaluate the posterior probability distribution of the parameters. Therefore, inferring this value might be a matter of estimation of a simpler form, $Q(\mathbf{p})$, which is the approximation of $P(\mathbf{p}|\mathbf{y}, M)$ [8]. The measure of the information lost in approximating $P(\mathbf{p}|\mathbf{y})$ with $Q(\mathbf{p})$, i.e. a non-symmetric measure of the difference between the two probability distributions, is given by the Kullback-Liebler (KL) divergence, also known as the Relative Entropy [9]:

$$KL [Q(\mathbf{p})||P(\mathbf{p}|\mathbf{y})] = \int Q(\mathbf{p}) \log \frac{Q(\mathbf{p})}{P(\mathbf{p}|\mathbf{y})} d\mathbf{p}. \quad (2.2)$$

Due to Gibb's inequality, this quantity is always assumed greater than or, when the two densities are the same, equal to zero. Thence, the estimation of the correct approximated posterior is achieved by minimising the KL divergence. The log of the evidence, or marginal likelihood, now can be evaluated in terms of both KL

divergence and free energy, F , which is defined as in the following equation, taking the expectation with respect to the density $Q(\mathbf{p})$ [10]:

$$\begin{aligned}
 \log P(\mathbf{y}) &= \log \frac{P(\mathbf{y}, \mathbf{p})}{P(\mathbf{p}|\mathbf{y})} \\
 &= \int Q(\mathbf{p}) \log \frac{P(\mathbf{y}, \mathbf{p})}{P(\mathbf{p}|\mathbf{y})} d\mathbf{p} \\
 &= \int Q(\mathbf{p}) \log \left[\frac{P(\mathbf{y}, \mathbf{p})}{P(\mathbf{p}|\mathbf{y})} \cdot \frac{Q(\mathbf{p})}{Q(\mathbf{p})} \right] d\mathbf{p} \\
 &= \int Q(\mathbf{p}) \log \frac{P(\mathbf{y}, \mathbf{p})}{Q(\mathbf{p})} d\mathbf{p} + \int Q(\mathbf{p}) \log \frac{Q(\mathbf{p})}{P(\mathbf{p}|\mathbf{y})} d\mathbf{p} \\
 &= F + KL.
 \end{aligned} \tag{2.3}$$

Since the log of the evidence is a fixed quantity, minimising the KL divergence is like maximising the free energy, F [8]. In fact, also according to Jensen's inequality, F becomes equal to the model evidence when $Q(\mathbf{p})$ is equal to the true posterior.

2.1.1.2 Variational Approach

To obtain a practical learning algorithm, the integral in F , through which the approximated posterior distribution can be achieved, must be tractable. A procedure for attaining this goal is to assume that the approximating density $Q(\mathbf{p})$ factorizes over groups of parameters. In physics, this is also known as the mean field approximation:

$$Q(\mathbf{p}) = \prod_i Q_{\mathbf{p}_i}(\mathbf{p}_i), \tag{2.4}$$

where \mathbf{p}_i represents the i th group of parameters. Thence the Equation 2.4 can also be written as

$$Q(\mathbf{p}) = Q_{\mathbf{p}_i}(\mathbf{p}_i)Q_{\mathbf{p}_j}(\mathbf{p}_j), \tag{2.5}$$

where $Q_{\mathbf{p}_j}(\mathbf{p}_j)$ denote the approximated distribution for the parameters not in the i th group. The mean field approximation assumes that the parameters in the separate groups are independent, although the complete factorisation of all the individual parameters is generally not required [11] and the choice of the groups is typically made logically, according to their role in the model and the simplicity

of calculus. The approximated distributions $Q(\mathbf{p}_i)$, which maximize the objective function F with respect to a factorised posterior distribution in turn, can be derived from:

$$F = \int Q(\mathbf{p}) \log \frac{P(\mathbf{y}|\mathbf{p})P(\mathbf{p})}{Q(\mathbf{p})} d\mathbf{p}. \quad (2.6)$$

Now, because F is a functional, which is a function of a function, to maximise that objective function is necessary to turn to the calculus of variations, here purposely introduced. Writing the functional in terms of the parameters alone as:

$$F = \int g(\mathbf{p}_i, Q_{\mathbf{p}_i}(\mathbf{p}_i)) d\mathbf{p}_i, \quad (2.7)$$

where

$$g(\mathbf{p}_i, Q_{\mathbf{p}_i}(\mathbf{p}_i)) = \int f(\mathbf{p}, Q(\mathbf{p})) d\mathbf{p}_{i'} \quad (2.8)$$

The maximum of F is the solution of the Euler differential equation:

$$\begin{aligned} & \frac{\partial}{\partial Q_{\mathbf{p}_i}(\mathbf{p}_i)} [g(\mathbf{p}_i, Q(\mathbf{p}_i), Q'(\mathbf{p}_i))] \\ & - \frac{d}{d\mathbf{p}_i} \left\{ \frac{\partial}{\partial Q'_{\mathbf{p}_i}(\mathbf{p}_i)} [g(\mathbf{p}_i, Q(\mathbf{p}_i), Q'(\mathbf{p}_i))] \right\} = 0, \end{aligned} \quad (2.9)$$

where the partial derivative of the second term is zero as g is not dependent upon $Q'_{\mathbf{p}_i}(\mathbf{p}_i)$. Returning to Equation 2.6 and using the logarithms' properties, the previous equation can be written as:

$$\begin{aligned} & \frac{\partial}{\partial Q_{\mathbf{p}_i}(\mathbf{p}_i)} \int Q(\mathbf{p}) \log \frac{P(\mathbf{y}|\mathbf{p})P(\mathbf{p})}{Q(\mathbf{p})} d\mathbf{p}_{i'} \\ & = \int Q_{\mathbf{p}_{i'}}(\mathbf{p}_{i'}) \log P(\mathbf{y}|\mathbf{p})P(\mathbf{p}) d\mathbf{p}_{i'} \\ & - \int Q_{\mathbf{p}_{i'}}(\mathbf{p}_{i'}) \log Q_{\mathbf{p}_{i'}}(\mathbf{p}_{i'}) d\mathbf{p}_{i'} \\ & - \int Q_{\mathbf{p}_{i'}}(\mathbf{p}_{i'}) \log Q_{\mathbf{p}_i}(\mathbf{p}_i) d\mathbf{p}_{i'} = 0. \end{aligned} \quad (2.10)$$

It follows that:

$$\log Q_{\mathbf{p}_i}(\mathbf{p}_i) = \int Q_{\mathbf{p}_{i'}}(\mathbf{p}_{i'}) \log P(\mathbf{y}|\mathbf{p})P(\mathbf{p}) d\mathbf{p}_{i'} + \text{constant}, \quad (2.11)$$

so

$$\log Q_{\mathbf{p}_i}(\mathbf{p}_i) \propto \int Q_{\mathbf{p}_{i'}}(\mathbf{p}_{i'}) \log P(\mathbf{y}|\mathbf{p})P(\mathbf{p}) d\mathbf{p}_{i'} \quad (2.12)$$

2.1.1.3 Conjugate-Exponential Restriction

Typically, rather than assuming a specific parametric form for the posteriors, it is possible to let them fall out of free-form optimisation of the objective function, F . In this connection, Attias [11] propose an iterative algorithm for the VB updating directly analogous to ordinary Expectation-Maximisation (EM) approach. Then again, it is common to work with priors that are conjugate with the complete data likelihood, in which case the posterior has the same parametric form as the prior [5]. In this case, the VB learning becomes a process of updating the posterior hyper-parameters, which is guaranteed by requiring the belonging of the complete data likelihood to exponential family distribution. The list of latent-variables models of practical interest with complete data likelihood in the exponential family is very long, e.g. Gaussian mixture, hidden Markov models and extensions, and all the models combining e.g. Gaussian, gamma, Poisson, Dirichlet distributions [4]. Therefore this choice is not generally so restrictive in the VB domain. Additionally, the advantage of requiring an exponential distribution for the complete data likelihood can be seen by examining Equation 2.12, where the choice naturally leads to an exponential form for the factorised posterior allowing a tractable VB solution.

2.1.1.4 Convergence

In practice, the non-linear models treated are approximated using a Taylor expansion. For the problems tackled in this work, and especially in areas of high non-linearity, the convergence might be no longer guaranteed. A typical consequence is that the VB iterative algorithm cycles through a limited set of solutions without settling on a single set of values.

Objective Function The adopted objective function is the free energy, which is the negative value defined in Equation 2.6. The composition of this quantity

can be deduced from [7] as:

$$\begin{aligned} F &= \int Q(\mathbf{p}) \log P(\mathbf{y}|\mathbf{p}) d\mathbf{p} - KL [Q(\mathbf{p})||P(\mathbf{p})] \\ &= L_{av} - KL [Q(\mathbf{p})||P(\mathbf{p})], \end{aligned} \quad (2.13)$$

where L_{av} is the average likelihood of the data and the second term is the KL divergence defined at Equation 2.2, which grows with the number of model parameters. Thus, F contains both accuracy and complexity terms, reflecting the two conflicting requirements of a good model. Considering the factorisation provided for the mean field approximation for $Q(\mathbf{p})$:

$$F = L_{av} - \sum_i KL [Q(\mathbf{p}_i)||P(\mathbf{p}_i|\mathbf{y})]. \quad (2.14)$$

Convergence Algorithm During the iterative process, the value of F , which must be maximised, may reach a maximum and, in reverse, start to decrease. Considering the non-monotonic trend of the objective function over the iterative process, a suitable convergence algorithm is called for. The approach chosen on this work can be summed in the following pseudo-code [8].

Data: Initial values of the hyper-parameters.

Result: Hyper-parameters corresponding to maximum value of Free Energy.

```

while  $iter < max\ iter$  do
  | if  $F_{new} = F_{old} \pm tol$  then
  | | break;
  | else if  $F_{new} > F_{old}$  then
  | | update;
  | | trials=0;
  | else if  $F_{new} < F_{old} \ \&\& \ trials < max\ trials$  then
  | | update;
  | | trials=trials+1;
  | else
  | | break;
  | end
end

```

Algorithm 1: Standard Trials (ST) method algorithm

In Section 2.4 are analysed two different update approaches by simulation studies: in the first instance, the update equations are derived from the factorized log-posterior and for all the hyper-parameters of the factorized posterior distributions; secondly, the performances of Levenberg-Marquardt method are observed. The best method for ensuring convergence depends upon the particular model used.

2.1.1.5 Hyper-parameters Definition

The model of the measurements \mathbf{y} can be written as:

$$\mathbf{y} = \mathbf{g}(\theta) + \mathbf{v}, \quad (2.15)$$

where $\mathbf{g}(\theta)$ is the non-linear forward model for the measurements, with parameters vector θ . The noise is assumed to be from a Gaussian distribution and can be

written as:

$$\mathbf{v} \sim N(0, \phi^{-1}\Sigma_v), \quad (2.16)$$

where Σ_v is the covariance matrix of the measurement error. Thus, for the N observations, the log-likelihood is:

$$\log P(\mathbf{y}|\Theta) = \frac{N}{2} \log \phi - \frac{1}{2} \phi (\mathbf{y} - \mathbf{g}(\theta))^T \Sigma_v^{-1} (\mathbf{y} - \mathbf{g}(\theta)), \quad (2.17)$$

where $\Theta = \{\theta, \phi\}$ is the set of all model and noise parameters. Hence, the factorised approximation is based on the set of hyper-parameters included in Θ :

$$Q(\Theta|\mathbf{y}) = Q_\theta(\theta|\mathbf{y})Q_\phi(\phi|\mathbf{y}). \quad (2.18)$$

The hyper-parameters here introduced, according to M. Chappell et al. (2009), are described by the following prior distributions:

$$P(\theta) \sim MVN(\theta; \mathbf{m}_0, \Lambda_0^{-1}); \quad (2.19)$$

$$P(\phi) \sim Ga(\phi; s_0, c_0), \quad (2.20)$$

where this gamma distribution is defined as:

$$Ga(\phi; s_0, c_0) = \frac{1}{\Gamma(c_0)} \frac{\phi^{c_0-1}}{s_0^{c_0}} e^{-\frac{\phi}{s_0}}, \quad \phi > 0. \quad (2.21)$$

The factorized posterior distributions are chosen conjugate with the priors as:

$$Q(\theta) \sim MVN(\theta; \mathbf{m}, \Lambda^{-1}); \quad (2.22)$$

$$Q(\phi) \sim Ga(\phi; s, c). \quad (2.23)$$

2.1.1.6 Update Equations and Free Energy

With the update equations proposed in this work and fully deduced in Appendix A, it is possible to achieve a weighted estimation, in the same way as in the WNLLS estimations, without additional computational price:

$$\Lambda = \Lambda_0 + sc\mathbf{J}^T\Sigma_v^{-1}\mathbf{J}; \quad (2.24)$$

$$\Lambda\mathbf{m}_{\text{new}} = \Lambda_0\mathbf{m}_0 + sc\mathbf{J}^T\Sigma_v^{-1}(\mathbf{k} + \mathbf{J}\mathbf{m}_{\text{old}}); \quad (2.25)$$

$$c = \frac{N}{2} + c_0; \quad (2.26)$$

$$\frac{1}{s} = \frac{1}{s_0} + \frac{1}{2}(\mathbf{k}^T \Sigma_v^{-1} + \text{tr}(\Lambda^{-1} \mathbf{J}^T \Sigma_v^{-1} \mathbf{J})). \quad (2.27)$$

Alternatively to these, in adopting Levenberg-Marquardt algorithm, might be used the equations described in details in Appendix B:

$$\mathbf{m}_{\text{new}} = \mathbf{m}_{\text{old}} + (\Lambda + \alpha \cdot \text{diag}(\Lambda))^{-1} \Delta, \quad (2.28)$$

where

$$\Delta = (sc \mathbf{J}^T \Sigma_e^{-1} (\mathbf{k} + \mathbf{J} \mathbf{m}_{\text{old}}) + \Lambda_0 \mathbf{m}_{\text{old}}) - \Lambda \mathbf{m}_{\text{old}}. \quad (2.29)$$

For both approach, the objective function to maximize is the free energy. This quantity can be inferred aptly introducing the covariance matrix, Σ_v , in the equation proposed by M. Chappell et al. (2009), as follow:

$$\begin{aligned} F = & -\frac{sc}{s_0} + \left(\frac{N}{2} + c_0 - 1 \right) [\log s + \psi(c)] \\ & - \frac{1}{2} \{ (\mathbf{m} - \mathbf{m}_0)^T \Lambda_0 (\mathbf{m} - \mathbf{m}_0) \} + \text{tr} \left(\frac{\Lambda_0}{\Lambda} \right) \\ & - \frac{1}{2} \left\{ \mathbf{k}^T \Sigma_v^{-1} \mathbf{k} + \text{tr} \left(\frac{\mathbf{J}^T \Sigma_v^{-1} \mathbf{J}}{\Lambda} \right) \right\} \\ & - c \log s - \log \Gamma(c) - c + (c - 1) [\log s + \psi(c)] \\ & + \frac{1}{2} \log \det(\Lambda) + \text{constant}, \end{aligned} \quad (2.30)$$

where ψ represents the di-gamma function and is defined as:

$$\psi(c) = \frac{d}{dc} \ln \Gamma(c) = \frac{\Gamma'(c)}{\Gamma(c)}. \quad (2.31)$$

2.1.1.7 Priors Setting

Noise Precision Priors Parameters that describe the gamma distribution of the noise precision are chosen to be relatively non-informative, as in [8], using:

$$s_0 = 10^6 \quad (2.32)$$

$$c_0 = 10^{-6} \quad (2.33)$$

Mean Priors of the Model Parameters The mean priors of the model parameters, \mathbf{m}_0 , used for the VB estimation of all the voxels of a ROI are deduced by a WNLLS estimation on the TAC of the same ROI. In fact, since data has an high SNR at the ROI level, WNLLS offers reliable estimation.

Precision Prior of the Model Parameters The priors setting of the precision, Λ_0 , of the model parameters can be deduced in terms of \mathbf{m}_0 and its coefficient of variation, CV , concerning the whole brain. The relationship between Λ_0 , CV and \mathbf{m}_0 is described by the following equation:

$$\Lambda_0 = \text{diag} \left(\frac{1}{(CV \mathbf{m}_0)^2} \right). \quad (2.34)$$

The setting algorithm is here described by the following pseudo-code:

Data: Initial value of the coefficient of variation of the model mean parameters.

Result: Chosen value of the coefficient of variation of the model mean parameters.

while $| \text{outliers}_{s_1} - \text{outliers}_{s_2} | > \text{tol}$ **do**

if $\text{outliers}_{s_1} > \text{outliers}_{s_2}$ **then**

$CV = CV - \Delta_d$;

else if $\text{outliers}_{s_1} < \text{outliers}_{s_2}$ **then**

$CV = CV + \Delta_i$;

else

 break;

end

 quantification;

end

Algorithm 2: Algorithm for the setting of the precision priors of the model parameters

The quantity outliers_{s_1} represents the percentage of outliers found when the precision of estimation is not satisfactory, i.e. when priors is not enough informative.

The value of $outliers_2$ represents the percentage of outliers found when the estimates are too constraint to priors, i.e. when priors are too informative. The parameters used in this algorithm, i.e. tol , the decrease factor Δ_d and the increase factor Δ_i , are empirically set in dependence to the specific needs.

Details concerning the choice of the classification criteria of outliers for real data quantification are described in Section 2.3. Therefore, the specific choice of the parameters tol , Δ_d and Δ_i , the results of the setting procedure for the different datasets and the impact of the precision prior of the model parameters on estimation are described in Section 4.2 and 5.3.

2.1.2 Weighted Non-linear Least Squares Method

Non-linear least squares problems arise when fitting a parametrized function, which is non-linear in parameters, to a set of N measured data points by minimizing the sum of the weighted square errors between the function and the measured data points. In general, the weights value might be arbitrary. However, when the variance of the measurement error is known, setting weights as the reciprocal of it is highly recommended. Consequently, smaller weights correspond to noisier measurements. Hence, the objective function is:

$$WRSS = \sum_{i=1}^N \frac{\mathbf{r}_i^2}{\sigma_i^2} = \sum_{i=1}^N \frac{(\mathbf{y}_i - \mathbf{g}_i(\mathbf{p}))^2}{\sigma_i^2}, \quad (2.35)$$

where \mathbf{y}_i is the i th measured data point, g represents the function which is non-linear in the parameters, \mathbf{p} , and σ_i is the standard deviation of the measurement error correspondent to i th sample. Introducing the covariance matrix of the measurement error, Σ_v , the objective function can be written in the matrix form as follow:

$$WRSS = \|\mathbf{r}^2\|_{\Sigma_v^{-1}} = \mathbf{r}^T \Sigma_v^{-1} \mathbf{r} = (\mathbf{y} - \mathbf{G}(\mathbf{p}))^T \Sigma_v^{-1} (\mathbf{y} - \mathbf{G}(\mathbf{p})) \quad (2.36)$$

and

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \{(\mathbf{y} - \mathbf{G}(\mathbf{p}))^T \Sigma_v^{-1} (\mathbf{y} - \mathbf{G}(\mathbf{p}))\}, \quad (2.37)$$

where $\hat{\mathbf{p}}$ is the correct estimated parameters vector. Since only in the case of linearity of the models the problem become analytically tractable, the estimations shown in this work need a numerical approach, e.g., the Levenberg-Marquardt (LM) method.

2.1.2.1 Implementation

Each result concerning the WNLLS estimations presented in the following sections is obtained using the MATLAB[®] (2012a) (The MathWorks Inc., Natick, MA, USA) function `lsqnonlin`, in which are implemented and alternatively used either LM method [14] or, by default, Trust-Region-Reflective (TRR) optimization approach [15]. A limitation of LM method is that it does not handle bound constraints. However, all model parameters estimated in this work have to satisfy, among them, the non-negativity constraint. For this reason TRR optimisation approach is adopted.

2.2 Compartmental Models and Quantification

In this section two different compartmental models used in this study to quantify the kinetic of the different tracers are described. In addition to this, there are some aspect of implementation that cannot be neglected, in particular in the matter of features of Variational Bayesian inference.

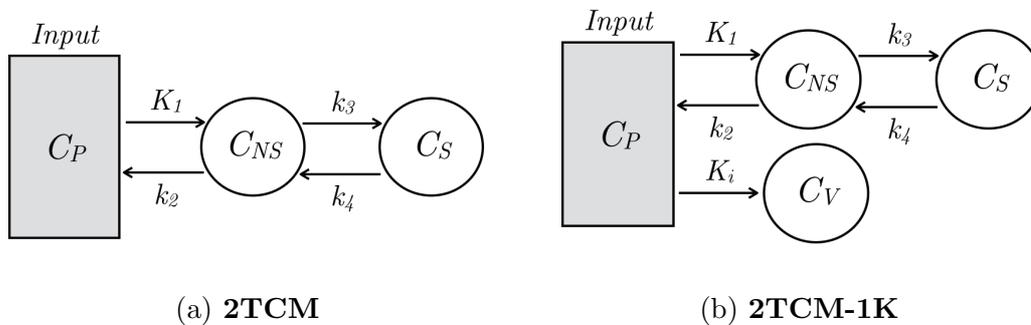


Figure 2.1: Compartmental models used to quantify the kinetics of the different tracers

2.2.1 The 2TCM

For both [^{11}C](*R*)-rolipram and [^{11}C]WAY100635 datasets the optimal model to describe the kinetic in the brain is the 2-tissue compartmental model (2TCM) [17, 20]. As shown in Figure 2.1, the model includes an arterial plasma fictitious compartment, C_P , and two tissue compartments: C_{NS} , which represents the non-displaceable component, and C_S , which describes the specific binding. The kinetic of the radioligands is described by the following first-order differential equations:

$$\begin{cases} \dot{C}_{NS}(t) = K_1 C_P(t) - (k_2 + k_3) C_{NS}(t) + k_4 C_S(t) & C_{NS}(0) = 0 \\ \dot{C}_S(t) = k_3 C_{NS}(t) - k_4 C_S(t) & C_S(0) = 0, \end{cases} \quad (2.38)$$

where K_1 ([ml/cm³/min]), k_2 ([min⁻¹]), k_3 ([min⁻¹]) and k_4 ([min⁻¹]) are the micro-parameters that respectively represent the rate constants for tracer transport from plasma to tissue and back and from the non-displaceable to the specific compartment and back [17]. Furthermore, the measurement equation in a given volume is:

$$C_{measured}(t) = (1 - V_b)[C_{NS}(t) + C_S(t)] + V_b C_b(t), \quad (2.39)$$

where V_b is a unitless quantity that represents the fraction of blood volume and C_b is the concentration of the radioligand in the whole blood including metabolites ([kBq·ml⁻¹]). The estimated measurement of concentration, $C_{measured}$, can also be written as in [20]:

$$C_{measured}(t) = (1 - V_b) (\alpha_1 e^{\beta_1 t} + \alpha_2 e^{\beta_2 t}) \otimes C_P + V_b C_b, \quad (2.40)$$

with

$$\beta_{1,2} = \frac{(k_2 + k_3 + k_4) \mp \sqrt{(k_2 + k_3 + k_4)^2 - 4k_2 k_4}}{2}, \quad (2.41)$$

$$\alpha_1 = \frac{K_1(\beta_1 - k_3 - k_4)}{\beta_1 - \beta_2}; \quad (2.42)$$

$$\alpha_2 = \frac{K_1(k_3 - k_4 - \beta_2)}{\beta_1 - \beta_2}. \quad (2.43)$$

In practice, the principal macro-parameter of interest is the total volume of distribution V_T ([ml/cm³]), which is calculated as:

$$V_T = \frac{K_1}{k_2} \left(1 + \frac{k_3}{k_4} \right). \quad (2.44)$$

2.2.2 The 2TCM-1K

[¹¹C]PBR28 binding is quantified using the 2TCM-1K and an arterial input function i.e. a 2-tissue compartment model with irreversible vascular trapping. The introduction of an irreversible tracer retention compartment is based on biological knowledge. The model is described by the following equations:

$$\begin{cases} \dot{C}_{NS}(t) = K_1 C_P(t) - (k_2 + k_3) C_{NS}(t) + k_4 C_S(t) & C_{NS}(0) = 0 \\ \dot{C}_S(t) = k_3 C_{NS}(t) - k_4 C_S(t) & C_S(0) = 0 \\ \dot{C}_V(t) = K_i C_P(t) & C_V(0) = 0, \end{cases} \quad (2.45)$$

where the K_i ($[\text{min}^{-1}]$) is the rate constant for tracer binding to the vascular receptors. The measurement equation is:

$$C_{measured}(t) = (1 - V_b)[C_{NS}(t) + C_S(t)] + V_b(C_b(t) + C_V(t)) \quad (2.46)$$

where the K_i ($[\text{min}^{-1}]$) is the rate constant for tracer binding to the vascular receptors. The macro-parameter of interest, V_T is still calculated according to equation 2.44.

2.3 Performances Evaluation and Methods Comparison on Real Data

The comparison between VB and the reference method, WNLLS, is made in terms of outliers percentage and the computational time required. For both [¹¹C](R)-rolipram and [¹¹C]WAY100635 the outliers classification criteria is based on estimation coefficient of variation of V_T ($threshold = 100\%$), when outliers percentage in other sections is also called *outliers*₁, and, only for VB, difference between estimated and prior parameters ($threshold = 5\%$), when outliers percentage in other sections is also called *outliers*₂. For [¹¹C]PBR28 data are also excluded estimations where $CV_{K_i} > 250\%$. In addition to that, V_T estimates > 10 always represents non-physiological values and are not included in results.

The correlation among the results obtained with the two quantification methods is evaluated with a regression analysis: the slope (m) and the intercept (q) of the fitted regression line with WNLLS as independent variable and VB as dependent variable is calculated for each subject. Pearson's R^2 values are reported as the correlation measure. Finally the percentage Mean Relative Difference, MRD , is calculated between the estimates obtained with WNLLS and VB for each subject:

$$MRD = \frac{1}{N} \sum_{i=1}^N \frac{\frac{1}{M} \sum_{j=1}^M V_{TVB}^{i,j} - \frac{1}{M} \sum_{j=1}^M V_{TWNLLS}^{i,j}}{\frac{1}{M} \sum_{j=1}^M V_{TWNLLS}^{i,j}} \cdot 100, \quad (2.47)$$

where N is the number of ROIs, M is the number of voxels of the ROI, $V_{TVB}^{i,j}$ is the value of the VB estimate in the j th voxel for the i th ROI and $V_{TWNLLS}^{i,j}$ is the value of the WNLLS estimate in the j th voxel for the i th ROI.

2.4 Simulation Studies

Since there aren't any results available on the theme, before starting the analysis of the PET datasets with the VB approach, simulation studies are necessary. The aims of these investigations mainly concern the choice of the convergence algorithm and the performance of the quantification methods. For these purposes, in particular, two simulation studies are proposed:

- Sensitivity test;
- Study on simulated PET data.

2.4.1 Sensitivity Test

The sensitivity test shows the performance of the method depending on parameters prior setting. In details, fixing a prior precision vector, Λ_0 , and varying some of the elements of the mean prior vector of the forward model, \mathbf{m}_0 , over an established values range, it's possible to test the behaviour of the chosen algorithm. Another aim of this test is the comparison between the Standard Trials (Algorithm 1) and the Levenberg-Marquardt approach in order to chose the best of

them for this study. A thousand noisy simulated curves were generated from a noiseless curve described as:

$$C_{true} = \alpha_1 e^{-\beta_1} + \alpha_2 e^{-\beta_2}, \quad (2.48)$$

where $\alpha_1 = 10$, $\beta_1 = 0.1$, $\alpha_2 = 4$ and $\beta_2 = 0.01$, which are also define as true priors, and applying each time a noise vector of pseudo-random values drawn from the standard normal distribution $N(0, 1)$. The log-spaced 25-element prior ranges, concerning only β_1 and β_2 parameters, are:

$$\beta_1 = 0.01 \div 1; \quad (2.49)$$

$$\beta_2 = 0.001 \div 0.1. \quad (2.50)$$

Thence, the prior matrix can be defined as:

$$\begin{pmatrix} (\beta_2(1), \beta_1(1)) & (\beta_2(1), \beta_1(2)) & \dots & (\beta_2(1), \beta_1(25)) \\ (\beta_2(2), \beta_1(1)) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ (\beta_2(25), \beta_1(1)) & \dots & \dots & (\beta_2(25), \beta_1(25)) \end{pmatrix}$$

For every couple of priors of this matrix $(\beta_2(i), \beta_1(j))$ is made an estimation. Finally, The results obtained for each simulated curve are then averaged. The classification criteria for the successful estimation is based on percentage bias (*threshold* = 5%) between $\hat{\mathbf{p}}_{i,j}$ and \mathbf{p}_{true} , which is the estimated parameters vector deduced from true priors.

2.4.2 Study on Simulated PET Data

The second simulation study includes a comparison between VB and WNLLS in which the performances of quantification methods are investigated in terms of outliers. The simulated voxels are generated as described in Figure 2.2. Measurement data voxels are taken from a cluster of a $[^{11}\text{C}](R)$ -rolipram dataset. In details, the frontal pole cluster is chosen as representative for the results. The outliers exclusion is based on estimation coefficients of variation (*threshold* = 100%)

and bias between $\hat{\mathbf{p}}$ and \mathbf{p}_{true} (*threshold* = 5%). The behaviour of the VB algorithm chosen from previous simulation study is tested in dependence of noise standard deviation, where $SD = 1 \div 5$.

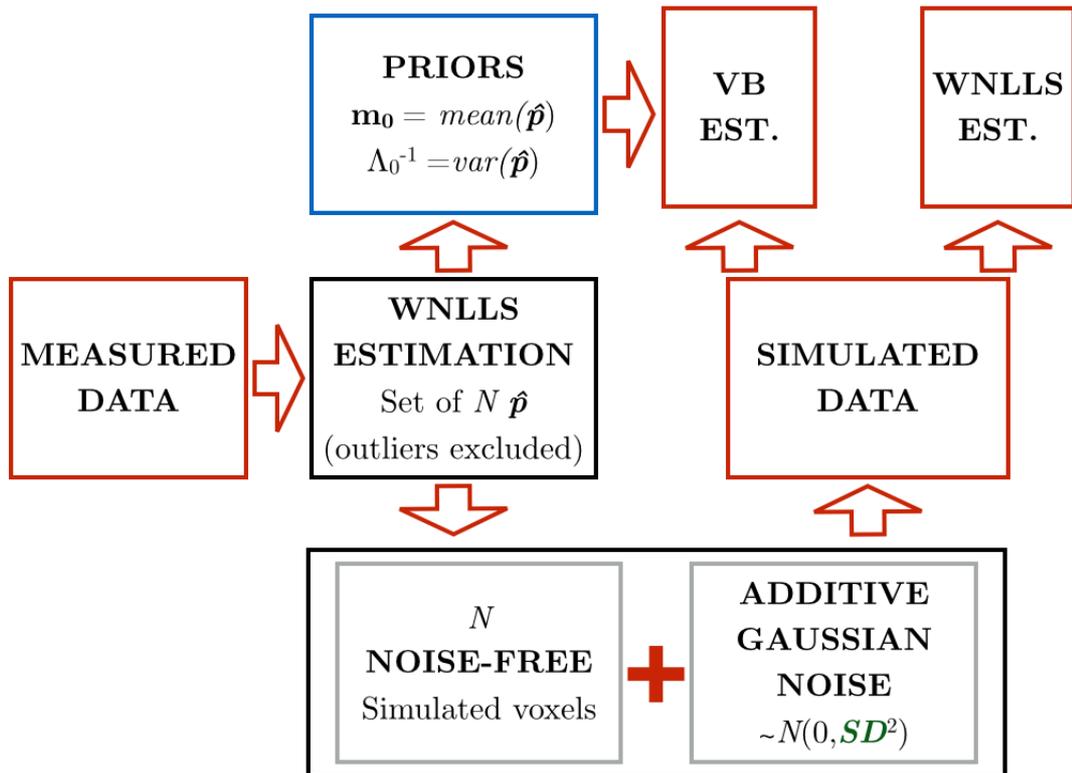


Figure 2.2: Stimulated voxels generation

Chapter 3

Datasets

The datasets analysed in this study include three different widely used radioligands: [^{11}C](*R*)-rolipram, [^{11}C]WAY100635 and [^{11}C]PBR28. Due to distinct quantification methods required for the analysis of these tracers, these datasets was chosen to evaluate the Variatioal Bayes performances.

3.1 [^{11}C](*R*)-rolipram

[^{11}C](*R*)-rolipram is a PET radioligand for the *in vivo* quantification of phosphodiesterase type IV (PDE4), whereof the Rolipram is a selective inhibitor. The phosphodiesterase IV is an enzyme that metabolizes 3',5'-cyclic adenosine monophosphate. The cyclic adenosine monophosphate (cAMP) cascade plays an important role in major depressive disorder (MDD) and it is a potential target for drugs development [17].

Image Acquisition Data were obtained from 3 healthy subject and were made available by National Institute of Mental Health. PET data were acquired using Advance tomograph (GE Medical System, Waukesha, WI, USA) after a bolus injection of 695 ± 152 MBq (range, 727 to 756) of [^{11}C](*R*)-rolipram. An 8-minute ^{68}Ge transmission scan was obtained before injection of the tracer for attenuation correction. Dynamic image data were acquired in 3D mode for 90 minutes. The

dynamic scan comprised 27 frames (6×30 , then 30×6 , 2×120 and 16×300 seconds). PET data were reconstructed on a 128×128 matrix with a pixel size of $2.0 \times 2.2 \times 4.25$ mm³. Informations about protocol approval and informed consent are described in [17].

Blood Data Acquisition Blood samples were drawn during the acquisition from the radial artery at 15-second intervals until 150 seconds, followed by 3 ml samples at 3, 4, 6, 8, 10, 15, 20, 30, 40 and 50 minutes, 4.5 ml at 60, 75 and 90 minutes. The decay-corrected whole blood activity, the fraction of unchanged radioligand in plasma and the plasma and whole blood ratio were calculated as previously described [18, 19].

ROIs definition Each subject underwent a high resolution 3D T1-weighted MRI scan. These MRI acquisitions were used, by adopting a maximum probability brain atlas, to derive the anatomic information necessary to define the 58 ROIs. The 9 regions chosen as significant for the results are: frontal pole, thalamus, caudate nucleus, putamen, pallidum, brainstem, hippocampus, amygdala and cerebellum.

3.2 [¹¹C]WAY100635

[¹¹C]WAY100635, is an high affinity radioligand and selective serotonin 5-hydroxytryptamine-1A (5HT_{1A}) receptor antagonist suitable for quantitative determination of 5HT_{1A} receptors in the human brain [20]. This receptors are of interest in the pathophysiology of several neuropsychiatric disorders, such as anxiety, depression and schizophrenia [21].

Image Acquisition A dataset of 3 healthy male subjects was made available from the Imperial College. Each subject underwent a 90-minute dynamic PET study in a high-sensitivity ECAT EXACT3D (Siemens/CTI, Knoxville, TN,

USA) scanner after a bolus injection of [^{11}C]WAY100635. Acquisition was performed in a list mode (event by event) and scans comprised 23 time frames of an increasing duration (two variable length background frames, 3×5 , 2×15 , 4×60 , 7×300 and 5×600 seconds). Data were reconstructed on a 128×128 matrix with a pixel size of $2.0 \times 2.1 \times 2.4 \text{ mm}^3$.

Blood Data Acquisition The radioactivity concentration in the blood was measured continuously and in addition serial discrete blood samples were taken at increasing time intervals throughout the study for the measurement of the radioactivity in the blood and plasma. Eight and nine of these samples respectively were also used for the quantification of the fraction of radioactivity attributable to the unmetabolized parent radiotracer, generating the metabolite-corrected arterial plasma input function for all subjects.

ROIs definition Each subject also had high resolution 3D T1-weighted MRI scans. These datasets were used only for ROIs definition, by using a maximum probability brain atlas subdivided in 73 regions, among them only 9 is here considered for the results: hippocampus, amygdala, cerebellum, thalamus, brainstem, putamen, caudate nucleus, pallidum, insula.

3.3 [^{11}C]PBR28

[^{11}C]PBR28 is a radioligand who targets the 18 kDa translocator protein (TSPO), formerly known as the peripheral benzodiazepine receptor (PBR). This protein is located on the outer mitochondrial membrane and is part of the mitochondrial permeability transition pore [22]. In particular, because TSPO is a putative biomarker of inflammation, the use of [^{11}C]PBR28 in brain PET data shows the neuroinflammation. TSPO proteins are also localized in the endothelium of brain vessels, such as venous sinuses and arteries. However, the impact of this component on quantification has to be investigated [23].

Image Acquisition The analysis of [^{11}C]PBR28 has been limited to a dataset of a single subject, that was made available by National Institute of Mental Health, previously analysed in [23], as proof of concept. The PET data were acquired on an Advance Nxi tomograph (GE Medical System, Waukesha, WI, USA), after a bolus injection of 690 ± 13 MBq of [^{11}C]PBR28. An 8-minute ^{68}Ge transmission scan of the brain was acquired for subsequent attenuation correction. The dynamic scans comprised 33 frames (6 frames of 30 seconds each, then 3×60 , 2×120 , 22×300 seconds) for a 120 minutes experiment duration. PET data were reconstructed on a 128×128 matrix with a pixel size of $2.0 \times 2.0 \times 4.25$ mm³. Informations concerning protocol approval and informed consent are included in [23].

Blood Data Acquisition During the acquisition, blood samples (1.0 ml each) were drawn from the radial artery at 15-second intervals until 150 seconds, followed by 3 to 4.5 ml samples at 3, 4, 6, 8, 10, 15, 20, 30, 40, 50, 60, 75, 90, 120 minutes. The plasma time activity curve was corrected with the fraction of uncharged radioligand, as previously described [19].

ROIs definition The chosen subject underwent 3 T clinical brain T1-weighted MRI. These images were used to derive the anatomical information necessary to define the 10 ROIs, similarly as done for previous tracers. The regions selected to show the results are: frontal cortex, parietal cortex, striatum, temporal pole, cingulum, insula, thalamus and cerebellum. In addition to these also white matter is considered.

3.4 Measurement Error

The data are affected by a measurement error, \mathbf{v} , which is assumed to be additive, uncorrelated and from a Gaussian distribution with zero mean. The model of measurements \mathbf{y} is:

$$\mathbf{y} = \mathbf{g}(\mathbf{p}) + \mathbf{v}. \quad (3.1)$$

The corresponding covariance matrix is assumed to be diagonal with the variance of the i -th element, $Var(C_i)$, define as ([24]):

$$Var(C_i) = \gamma \frac{\bar{C}_i}{\Delta_{t_i}}, \quad (3.2)$$

where C_i is the intensity of one reconstructed voxel C of i th frame, \bar{C}_i is the average whole brain concentration in the i th frame and Δ_{t_i} the duration of the same frame. As said before, the covariance matrix might be reasonably used to calculate the weights. Furthermore, the proportionally constant γ is an unknown scale factor and it is estimated *a posteriori* as in A. Bertoldo et al. (1998) for the WNLLS estimation. For the VB inference, the utilisation of covariance matrix, Σ_v , as weights matrix are described in Section 2.1.1.6 and γ must be inferred according to noise precision hyper-parameter distribution. In particular, the mean of the posterior gamma distribution is equal to $s \cdot c$. Therefore, as can be deduced from Equation 2.16, the scale factor can be calculated as:

$$\gamma = \frac{1}{sc}. \quad (3.3)$$

Chapter 4

Results

4.1 Simulation Studies Results

4.1.1 Sensitivity Test Results

The sensitivity test described in Section 2.4.1 gives the results shown in the following figure.

Performance of VB on Simulated Dataset: Impact of Prior Definition

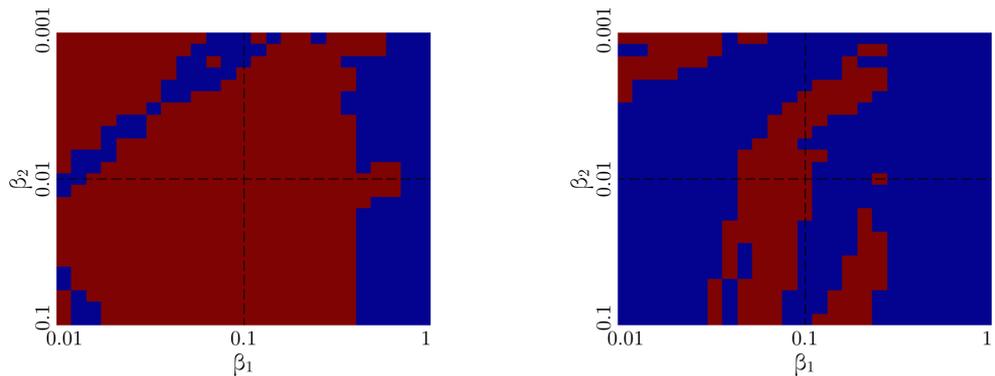


Figure 4.1: Sensitivity test results. Comparison between standard trials (left) and LM algorithm (right). Blue represents a failure in estimation, red shows a success

The standard trials (ST) shows great tolerance to priors definition. In most cases,

it finds a successful estimation, even for high differences between true and tested priors. In details, concerning borderline cases, both algorithms exhibit better performance when priors are set to a smaller value than true. To note that, for the model tested, LM has a higher percentage of failure. For this reason, ST is hereon chosen as the convergence method.

4.1.2 Study on Simulated PET Data Results

The performances in terms of outliers percentage of VB, implemented using ST algorithm, and WNLLS are compared for various level of noise variance. In particular, as shown in Figure 4.2, setting the noise standard deviation over a discrete range ($SD = 1 \div 5$), the two approaches exhibit different performances.

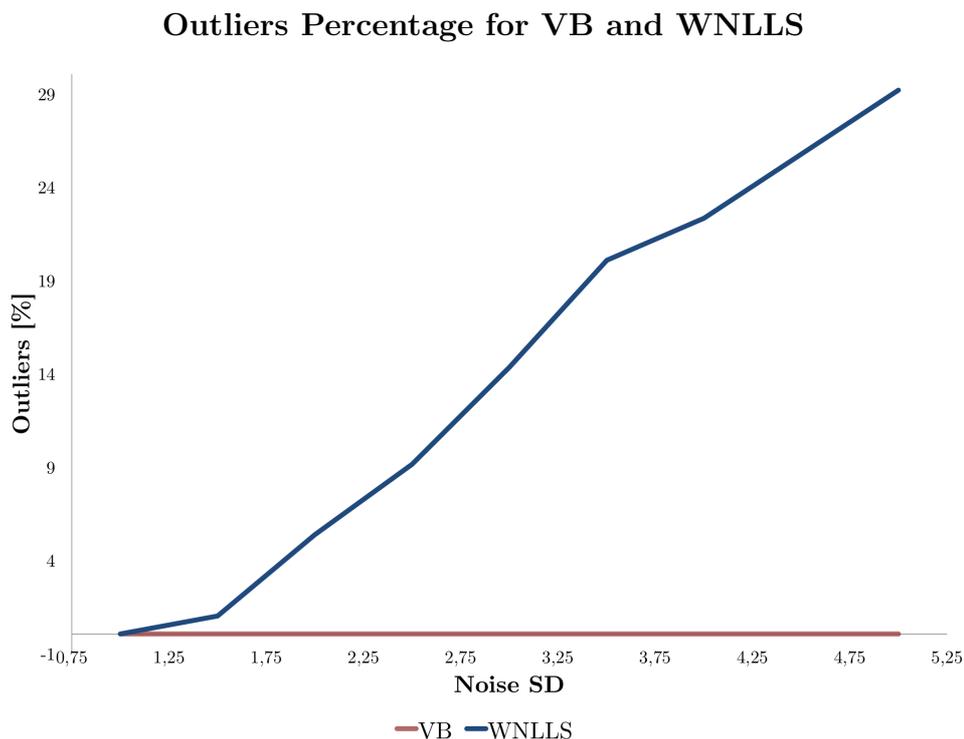


Figure 4.2: Noise variance impact test: a comparison between VB and WNLLS approach

At each noise variance level, VB exhibits better performance, i.e. smaller per-

centage of outliers, on these simulated data. In particular, in adopting WNLLS estimation method, the percentage of outliers constantly increases with the noise variance. Differently, for VB approach, this percentage appears to be almost constant ($\simeq 0.5\%$).

4.2 Impact of Precision Priors of the Model Parameters

The impact of model precision parameters prior on VB performance is shown in the following figure, where the frontal pole region and the temporal pole region are chosen as representative.

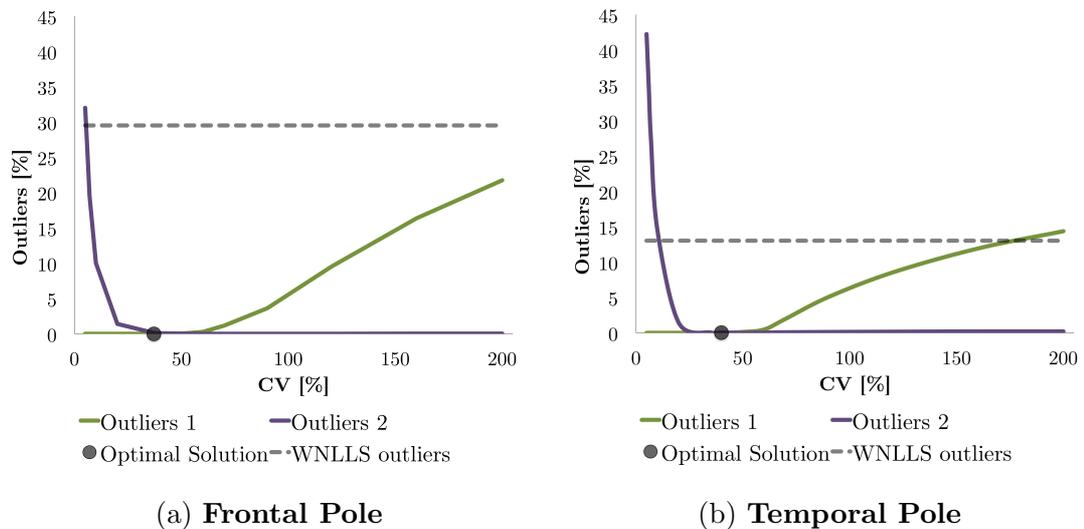


Figure 4.3: Impact of precision priors of model parameters on VB performance. Frontal pole (a) and temporal pole (b) ROIs are chosen as representative and significant for the results

When CV gives the optimal solution, the VB estimations are almost always successful (outliers percentage $\simeq 0\%$). When the prior precision becomes too small (i.e. when mean priors are too informative), or too high (i.e. when mean priors are not enough informative) the percentage of outliers increases. Therefore,

comparing VB and WNLLS, the first method exhibit better performance for a wide range of CV values, in which the true CV is obviously included. Results shown in Figure 4.3 are assumed to be representative of all ROIs.

4.3 Real Data Results

4.3.1 $[^{11}\text{C}](R)$ -rolipram Results

<i>interp. ratio = 560</i>		VB			WNLLS	
ROI	voxels	out. [%]	iter.	c. time [min]	out. [%]	c. time [min]
Fr Pole	5980	0.05	14	2.97	29.13	3.62
Thal	1420	0.00	13	0.65	45.08	1.04
Caud	520	0.39	13	0.25	37.16	0.41
Put	821	0.49	14	0.40	38.37	0.61
Pall	271	0.74	13	0.12	39.11	0.19
Br Stem	1967	0.25	13	0.91	54.55	1.24
Hipp	781	0.38	13	0.37	44.17	0.47
Amyg	366	0.55	15	0.19	46.72	0.26
Cer	2918	0.21	14	1.45	64.87	1.76

Table 4.1: Performances table. Subject 25 taken from $[^{11}\text{C}](R)$ -rolipram dataset

The parametric maps of distribution volume V_T , which is the macro-parameter of interest, included in Figure 4.4, clearly show the difference between VB and WNLLS. In details, for each subject, VB achieves successful estimations in $\simeq 100\%$ of voxels. Differently, WNLLS finds a significant percentage of outliers (in most cases $> 25\%$). Also, VB appears to slightly underestimate V_T in comparison with WNLLS. Figure 4.5 shows the inter-subject regional values of V_T obtained with VB and WNLLS, with VB underestimation. Also $MRD = -2.8\% \pm 0.5\%$ and the regression line ($m = 1.2 \pm 0.03$), with a considerable correlation between the methods ($R^2 = 0.77 \pm 0.03$), reports that result. Finally, Table 4.1 presents

the performances of the two methods for one representative subject.

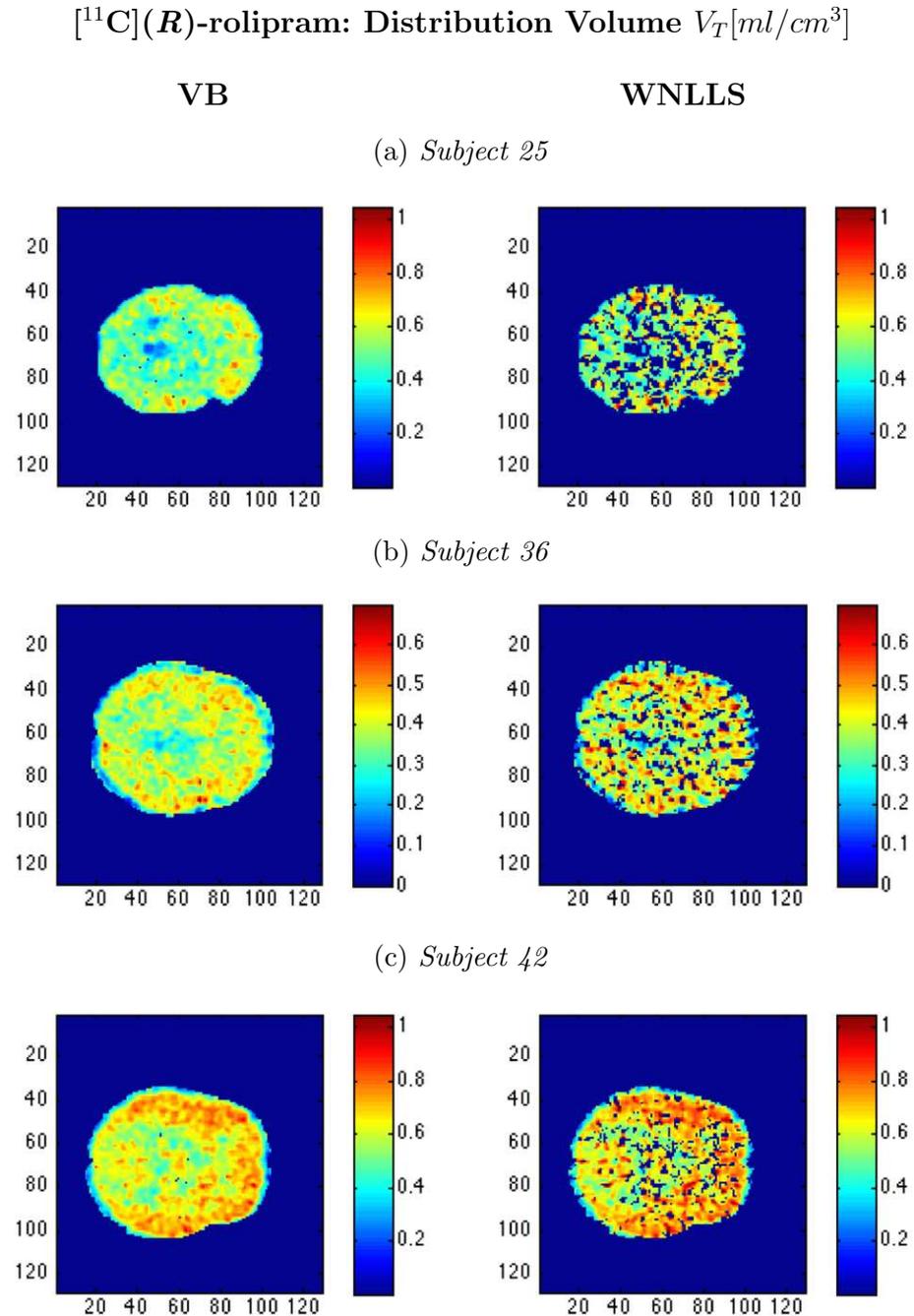


Figure 4.4: Parametric maps (V_T) of one representative slice, for each quantification method and for each subject. $[^{11}\text{C}](R)\text{-rolipram}$ datasets

The computation time is always in relationship with the interpolation ratio, which

is equal to the ratio of length of time vector referred to C_p and C_b to length of time vector relative to PET images acquisition. In fact, the interpolations required during the quantification process significantly weights on computations, iteration by iteration.

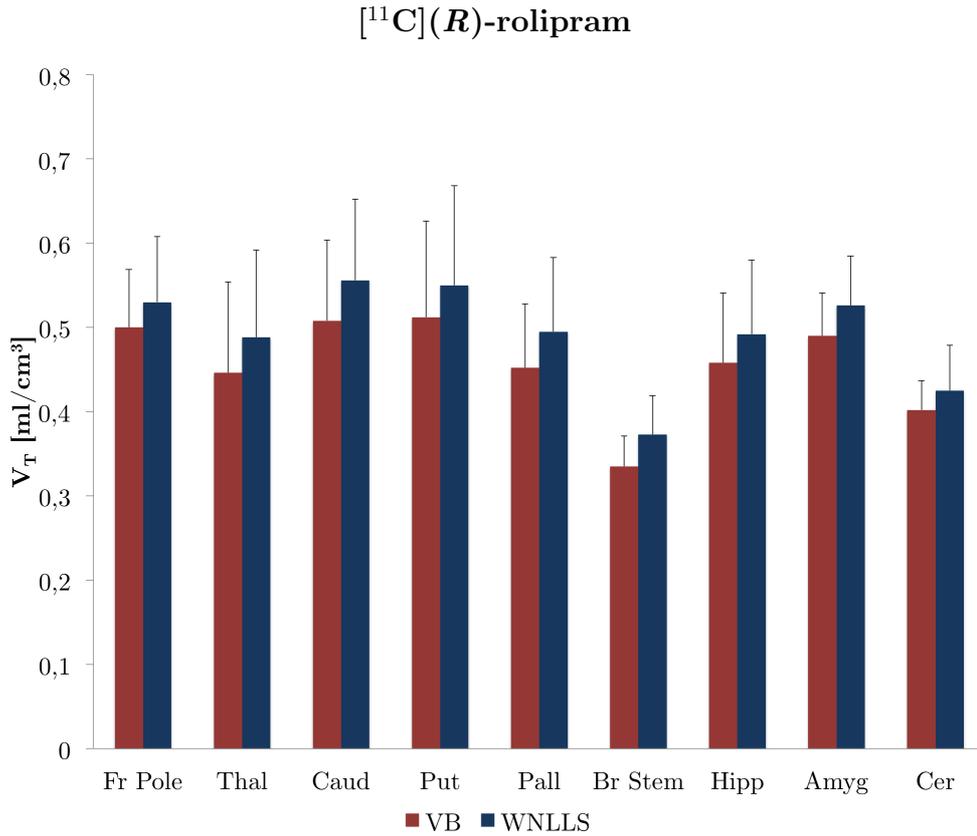


Figure 4.5: Inter-subject regional values of V_T obtained with VB (red) and WNLLS (blue). The average estimated precision is expressed in terms of between-subject SDs. The chosen representative ROI are: Fr Pole = Frontal Pole, Thal = Thalamus, Caud = Caudate, Put = Putamen, Pall = Pallidum, Br Stem = Brainstem, Hipp = Hippocampus, Amyg = Amygdala, Cer = Cerebellum. [^{11}C](R)-rolipram datasets

Also the table demonstrates what previously said about outliers percentage. The computation time required by VB is considerably shorter than that necessary for

WNLLS estimation. On equal terms, at whole brain level and concerning same previously introduced representative subject, the difference between methods is clear: 39 minutes for VB, 52 minutes for WNLLS.

4.3.2 $[^{11}\text{C}]\text{WAY100635}$ Results

Many results concerning $[^{11}\text{C}]\text{WAY100635}$ datasets, such as the results of the qualitative evaluation of the parametric maps included in Figure 4.6, are comparable to those presented in previous section. For this reason, particular attention is paid to the distinctive feature for these datasets.

<i>interp. ratio = 356</i>		VB			WNLLS	
ROI	voxels	out. [%]	iter.	c. time [min]	out. [%]	c. time [min]
Ins	3602	2.00	13	0.94	44.03	1.56
Thal	1865	3.64	13	0.5	78.98	0.64
Caud	1099	6.00	13	0.3	83.08	0.39
Put	1061	3.11	13	0.29	60.34	0.38
Pall	294	3.74	13	0.08	76.78	0.10
Br Stem	2911	4.98	13	0.79	82.58	1.04
Hipp	593	4.59	12	0.13	67.38	0.23
Amyg	352	4.55	12	0.09	65.45	0.15
Cer	16130	4.95	13	4.27	72.65	5.55

Table 4.2: Performances table. Subject 4028 taken from $[^{11}\text{C}]\text{WAY100635}$ dataset

Inter-subject values of V_T of representative chosen ROIs included in Figure 4.7 exhibit comparable results for the two estimation methods for small values of V_T . Differently, for higher values of V_T , the difference between VB and WNLLS increases.

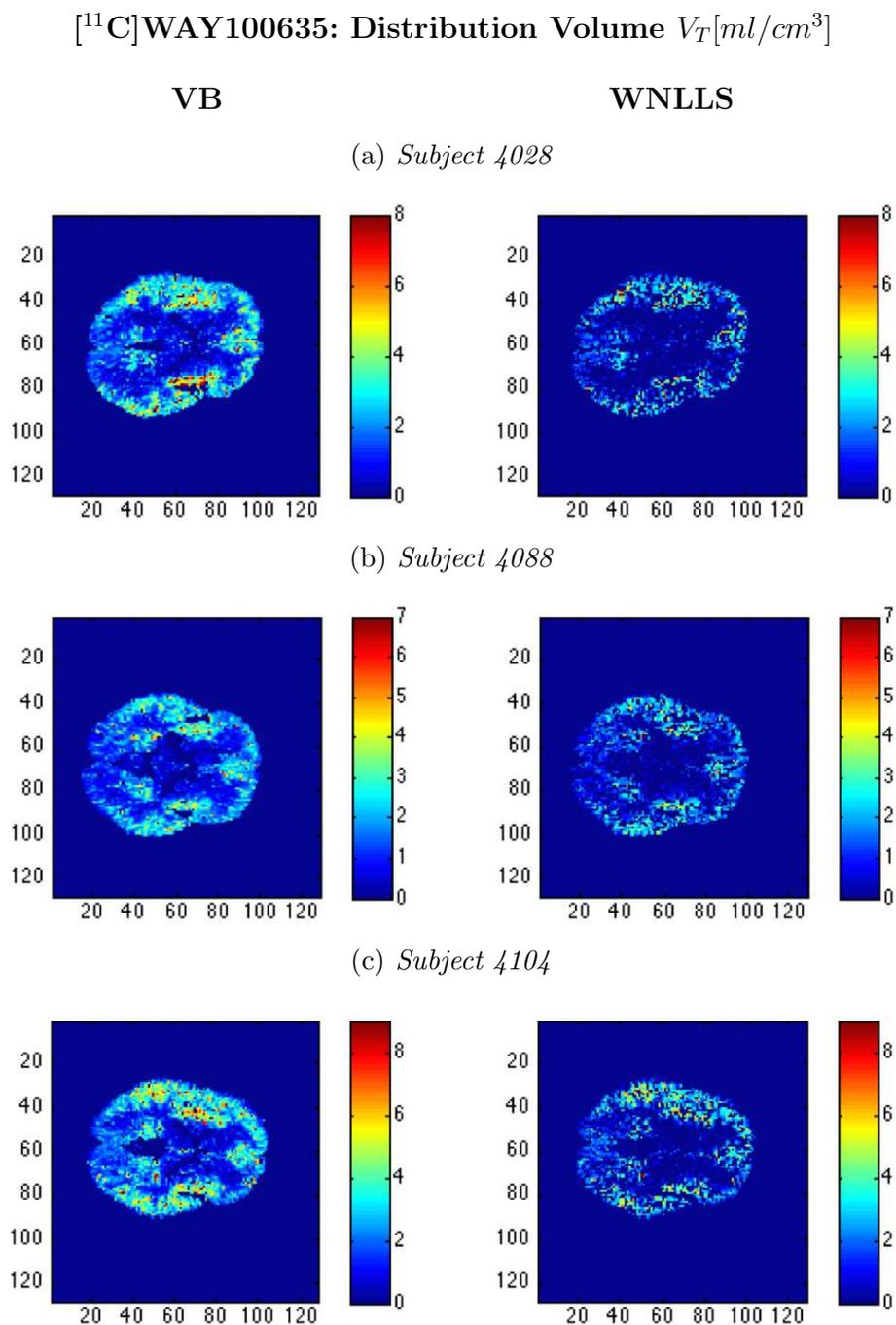


Figure 4.6: Parametric maps (V_T) of one representative slice, for each quantification method and for each subject. [¹¹C]WAY100635 datasets

The [¹¹C]WAY100635 measurements appear to be noisier than previous data. In consequence of this, Table 4.2 reports VB results (with an outliers percentage

< 6%) and WNLLS performance (that presents always a > 44% percentage of outliers).

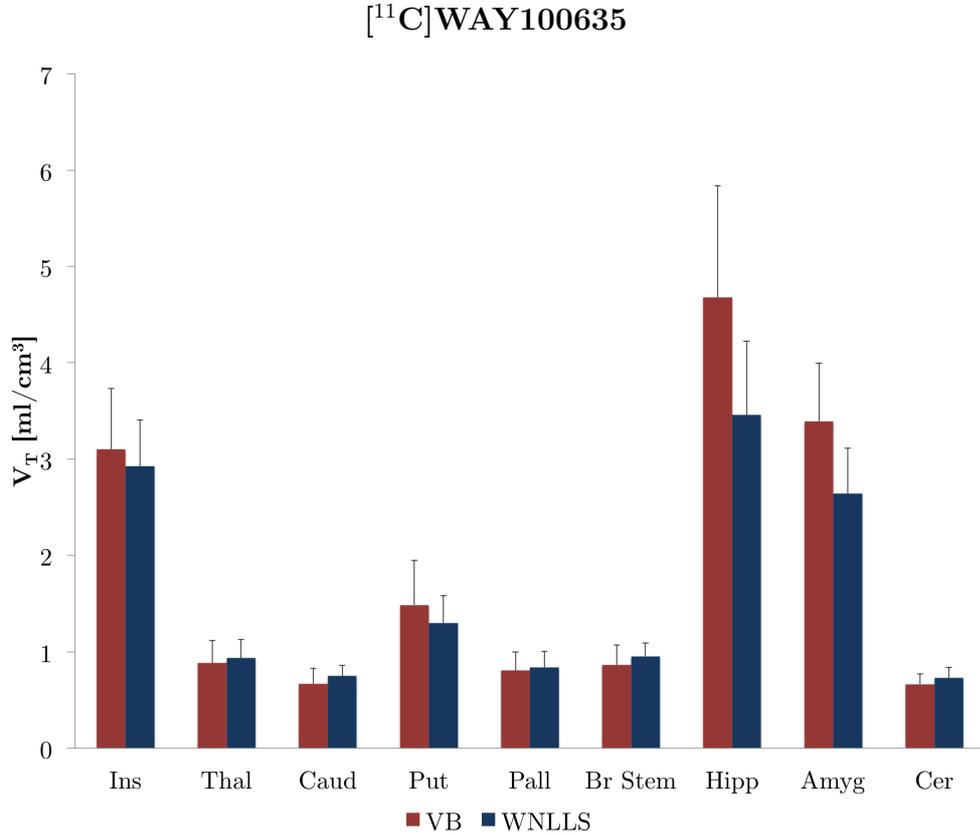


Figure 4.7: Inter-subject regional values of V_T obtained with VB (red) and WNLLS (blue). The average estimated precision is expressed in terms of between-subject SDs. The chosen representative ROI are: Ins = Insula, Thal = Thalamus, Caud = Caudate, Put = Putamen, Pall = Pallidum, Br Stem = Brainstem, Hipp = Hippocampus, Amyg = Amygdala, Cer = Cerebellum. [¹¹C]WAY100635 datasets

However the methods have a very great agreement with $MRD = 2.9 \pm 2.2$, $R^2 = 0.91 \pm 0.02$ and a regression line very similar to the bisector ($m = 1.05 \pm 0.06$, $q = -0.09 \pm 0.06$). Computation time is still significantly shorter for VB and, at the whole brain level, it requires 34 minutes. 12 minutes less than WNLLS.

Another important result is shown in Figure 4.8, where the average weighted residuals are shown for both VB and WNLLS. VB presents on average zero-mean weighted residuals, in accordance with the hypothesis in the error variance.

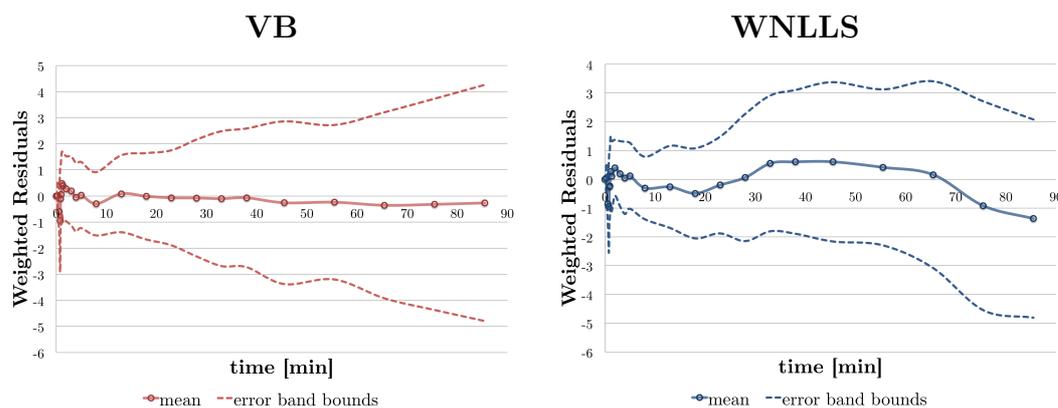


Figure 4.8: Average weighted residuals achieved from voxelwise quantification on a representative ROI (left thalamus) with VB (left) and WNLLS (right)

4.3.3 $[^{11}\text{C}]\text{PBR28}$ Results

$[^{11}\text{C}]\text{PBR28}$: Distribution Volume $V_T[\text{ml}/\text{cm}^3]$

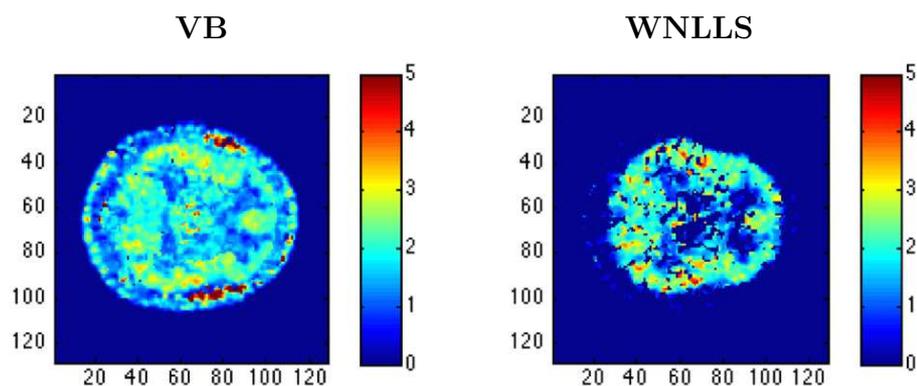


Figure 4.9: Parametric maps (V_T) of one representative slice, for each quantification method and for each subject. $[^{11}\text{C}]\text{PBR28}$ datasets

The results shown in this section reveal a great variability for WNLLS performance between ROIs. In details, this approach fails almost always in the white

matter, which is a very noisy region. On the other side, the maps included in Figure 4.9 exhibit almost constant and satisfying performances for VB.

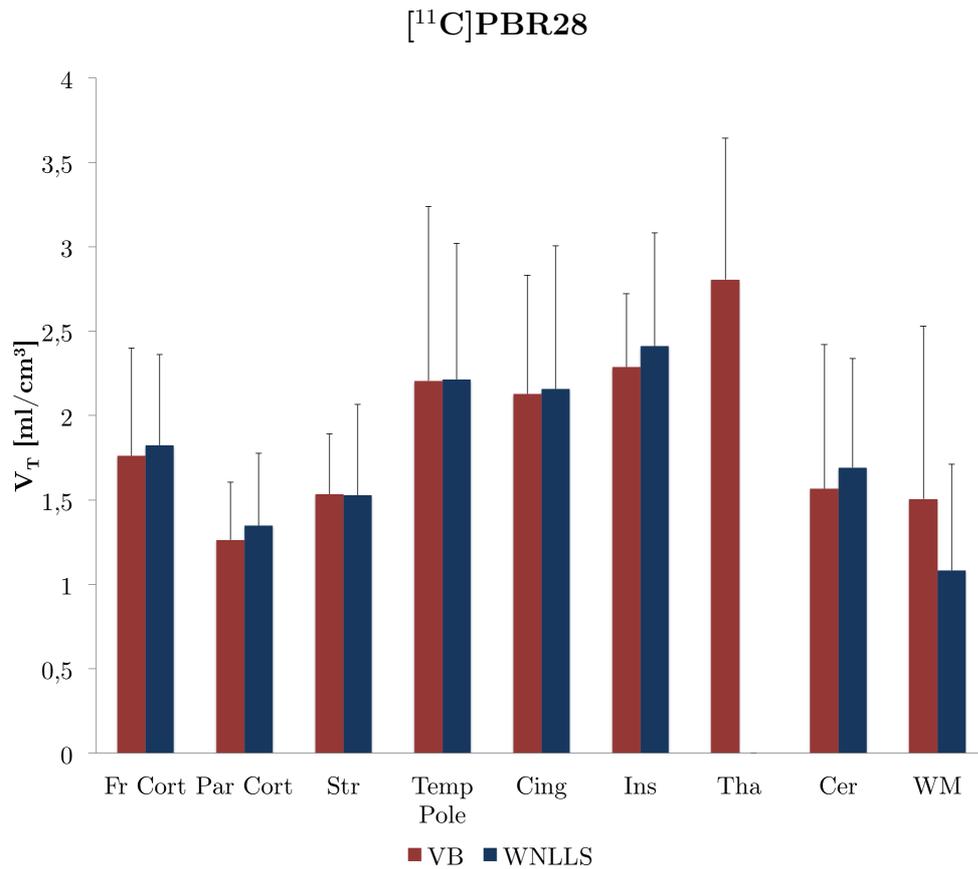


Figure 4.10: Subject regional values of V_T obtained with VB (red) and WNLLS (blue). The estimated precision is expressed in terms of SD. The chosen representative ROI are: Fr Cort = Frontal Cortex, Par Cort = Parietal Cortex, Str = Striatum, Temp Pole = Temporal Pole, Cing = Cingulate, Ins = Insula, The = Thalamus, Cer = Cerebellum, WM = White Matter. [¹¹C]PBR28 datasets

In Table 4.3 the computation time aren't included because, differently from 2TCM model, the solution of the 2TCM-1K differential equations is numerically achieved. As consequence of this, the computation time comparison is distorted

and comparable to the WNLLS one. Thalamus region appears to be very difficult to quantify. Indeed, WNLLS always fails and VB requires a number of iteration notably greater than the average value. Nevertheless, VB percentage outliers is still satisfying for this ROI. Differently, satisfactory performance is found only in frontal cortex and insula region for WNLLS.

<i>interp. ratio = 480</i>		VB		WNLLS
ROI	voxels	out. [%]	iter.	out. [%]
Fr Cort	22080	0.52	13	8.65
Par Cort	9223	6.12	12	39.65
Str	1904	0.11	14	29.04
Temp Pole	12681	1.81	14	27.39
Cing	3076	0.23	13	37.71
Ins	1421	0.42	14	15.41
Tha	847	2.72	19	100.00
Cer	9860	3.36	14	26.00
WM	133503	8.06	15	85.25

Table 4.3: Performances table concerning the subject taken from [¹¹C]PBR28 dataset

White matter represents also the region where VB exhibits its worst performance mainly because V_T reaches non-physiological values ($V_T > 10$). In the same region, as shown in Figure 4.10, the greatest difference between VB and WNLLS quantification is found. Therefore, lowest levels of estimation precision are especially found in small and noisy, i.e. thalamus and white matter. If thalamus region, in with WNLLS does not achieve any estimation, is excluded in comparing the two quantification methods, MRD is very small ($= 0.7\%$) and the regression line is

similar to the bisector.

Chapter 5

Discussion

5.1 Sensitivity Test Discussion

As previously said, the performance of both ST and LM method is variable in dependence to the particular quantification model [8]. Hence, before starting an analysis on real data, a sensitivity test is highly recommended. In this work, the best method for ensuring convergence appears to be the ST. Indeed, it shows so great tolerance on priors definition that, in most cases, it achieves successful estimation according to classification criteria described in Section 2.4.1. In analysing real data, the priors \mathbf{m}_0 used for the voxelwise estimation in a particular ROI are attained by WNLLS quantification on the related ROI TAC. However, these priors rarely represent the best choice to achieve a satisfactory voxelwise estimation. For this reason, the features here exhibited by ST are undoubtedly precious. Thence, the convergence method adopted in this work is the ST.

5.2 Study on Simulated PET Data Discussion

The influence of noise level on estimation process can be deduced from both simulated studies and real data analysis. First of all, as shown in Figure 4.2, VB exhibits almost constant and satisfactory performance for all noise level here tested on simulated voxels. However, a clear dependence between the signal to

noise ratio (SNR) and the percentage of outliers on real data are expected for VB as well. Indeed, different performances, in terms of outliers percentage, emerge in analysing different datasets. In details, the estimations on $[^{11}\text{C}]\text{WAY100635}$ data, which are the noisiest of the datasets analysed in this work, find higher outliers percentages than those obtained on $[^{11}\text{C}](R)\text{-rolipram}$ datasets, on equal quantification terms. In most cases, differently from WNLLS, VB find outliers percentage smaller than 5%, even when the SNR became small. These performances appears clear also for $[^{11}\text{C}]\text{PBR28}$ data, which are quantified at the voxel level using the 2TCM-1K model. In particular, the estimation on white matter ROI, which is usually a very noisy region, is highly problematic for WNLLS and, in reverse, successful for VB. In conclusion, the dependence between estimation performance and noise level, which is less evident for VB, can be even problematic for WNLLS and it can prevent the quantification process.

5.3 Impact of Prior Precision of Model Parameters

As previously said, the parameters used in the Algorithm 2, i.e. tol , the decrease factor Δ_d and the increase factor Δ_i , can be empirically set in dependence to the specific needs. For example, in this work the parameters can be set as $tol = 3$, $\Delta_d = 0.025$ and $\Delta_i = 0.05$, where the increase and the decrease factors are not equal because the slopes of the different outlier curves shown in Figure 4.3 are clearly distinct. Few iterations are usually required to achieve the CV value. Therefore, for each tracer, the setting procedure can be done only on a representative cluster. This process guarantees almost constant performance in the whole brain and for all the different subjects of the dataset. In conclusion, the approximated values of the coefficients of variation found for the datasets described in Chapter 3 are respectively: $CV = 40\%$ for $[^{11}\text{C}](R)\text{-rolipram}$ datasets and $CV = 45\%$ for $[^{11}\text{C}]\text{WAY100635}$ and $[^{11}\text{C}]\text{PBR28}$ data.

5.4 Quantification Models Discussion

The discussion about the real data quantification must be done in dependence to the different models described in Chapter 3. Particular attention is paid to the 2TCM-1K because of its use on $[^{11}\text{C}]\text{PBR28}$ dataset and the first attempt of voxelwise quantification using this model.

5.4.1 The 2TCM Quantification Discussion

The 2TCM is used to quantify datasets relative to differently noisy radioligands. $[^{11}\text{C}](R)\text{-rolipram}$ and $[^{11}\text{C}]\text{WAY100635}$ are also characterized by different binding affinity levels and targeting properties. Despite that and differently from WNLLS, VB method exhibit successful and promising performance at the voxel level for all ROIs of the brain of all subjects. For all these reasons, satisfactory performance of VB method are expected also for datasets concerning other radioligands quantified using 2TCM. Another important aspect must be discussed: the computation time. Indeed, VB method achieves more reliable estimation than WNLLS in a 25% (or more) computational time shorter. The mean of *MRD* reports a small difference between methods for both $[^{11}\text{C}](R)\text{-rolipram}$ and, with an higher *SD*, $[^{11}\text{C}]\text{WAY100635}$. The regression lines are always similar to the bisector with a stronger correlation between methods when the binding affinity is higher, i.e. for $[^{11}\text{C}]\text{WAY100635}$. In any case, VB gives reliable estimation according to what obtained with WNLLS.

5.4.2 The 2TCM-1K Quantification Discussion

The use of this model, which explicitly accounts for endothelial TSPO binding, is fundamental for a precise and accurate quantification of $[^{11}\text{C}]\text{PBR28}$ brain PET data [23]. However, previous studies do not investigated this quantification at the voxel level. The two estimation methods used in this work exhibit very different behaviours. First of all, WNLLS shows so heterogeneous performance on different ROIs that it completely fails on quantifying thalamus region, in which is observed

the highest binding level, as it's shown in Figure 4.10 and as it's reported by G. Rizzo et al. (2014). However a strict dependence between the binding level and the WNLLS performance is not observed. A considerably correlation is here observed ($R^2 = 80$) with a regression line similar to the bisector. Also *MRD* value reports a small difference between methods. On another side, the white matter region represents an exception for both VB and WNNLS, as described in Sections 5.2 and 4.3.3. The segmentation process classifies the white matter voxels into a very numerous cluster that concerns also other anatomical regions. As said before, the mean priors are given by a WNLLS estimation on the ROI TAC and, in this case, they might not be sufficiently accurate. Thence, a more specific segmentation might facilitate the quantification process on this ROI for both methods. In any case, VB exhibits promising performance with a outliers percentage, in most cases, smaller than 5%. Nevertheless, the quantification on a single subject is not sufficient to evaluate a method, but the results shown in this work represent a promising starting point.

Appendix A

Derivation of the Update Equations

A.1 Forward Model Parameters

Equation 2.12 represents the condition of maximization of the objective function F . In details that equation can be rewritten as:

$$\log Q(\theta|\mathbf{y}) \propto \int LQ(\phi|\mathbf{y})d\phi, \quad (\text{A.1})$$

where

$$\begin{aligned} L = & -\frac{1}{2}\phi(\mathbf{y} - \mathbf{g}(\theta))^T \Sigma_v^{-1}(\mathbf{y} - \mathbf{g}(\theta)) + \frac{N}{2} \log \phi - \frac{1}{2}(\theta - \mathbf{m}_0)^T \\ & \times \Lambda_0(\theta - \mathbf{m}_0)(c_0 - 1) \log \phi - \frac{1}{s_0} \phi + \text{const} \{ \theta, \phi \} \end{aligned} \quad (\text{A.2})$$

is the log-posterior. The log of the multivariate normal distribution (MVN) of the forward model parameters gives:

$$\log Q(\theta|\mathbf{y}) = -\frac{1}{2}\theta^T \Lambda \theta + \theta^T \Lambda \mathbf{m} + \text{const} \{ \theta \}. \quad (\text{A.3})$$

Taking no notice of terms dependent to ϕ of the log posterior, also the second term of A.1 can be rewritten as:

$$\begin{aligned}
\int LQ(\phi|\mathbf{y})d\phi &= \int \left(-\frac{1}{2}\phi(\mathbf{y} - \mathbf{g}(\theta))^T \Sigma_v^{-1}(\mathbf{y} - \mathbf{g}(\theta)) \right) \times Ga(\phi; s, c)d\phi \\
&\quad + \int \left(-\frac{1}{2}(\theta - \mathbf{m}_0)^T \Lambda_0(\theta - \mathbf{m}_0) \right) \times Ga(\phi; s, c)d\phi \\
&\quad + \int \text{const}\{\theta\} \times Ga(\phi; s, c)d\phi \\
&= -\frac{1}{2}(\theta - \mathbf{m}_0)^T \Lambda_0(\theta - \mathbf{m}_0) \\
&\quad - \frac{1}{2}\phi(\mathbf{y} - \mathbf{g}(\theta))^T \Sigma_v^{-1}(\mathbf{y} - \mathbf{g}(\theta)) \int \phi Ga(\phi; s, c)d\phi \\
&\quad + \text{const}\{\theta\} \\
&= -\frac{1}{2}(\theta - \mathbf{m}_0)^T \Lambda_0(\theta - \mathbf{m}_0) \\
&\quad - \frac{1}{2}sc\phi(\mathbf{y} - \mathbf{g}(\theta))^T \Sigma_v^{-1}(\mathbf{y} - \mathbf{g}(\theta)) + \text{const}\{\theta\}.
\end{aligned} \tag{A.4}$$

Using the linearisation $\mathbf{g}(\theta) \approx \mathbf{g}(\mathbf{m}) + \mathbf{J}(\theta - \mathbf{m})$, where \mathbf{J} is the Jacobian

$$\mathbf{J}_{j,k} = \frac{d(\mathbf{g}(\theta)_j)}{d\theta_k} \tag{A.5}$$

computed in $\theta = \mathbf{m}$:

$$\mathbf{y} - \mathbf{g}(\theta) = \mathbf{y} - (\mathbf{g}(\mathbf{m}) + \mathbf{J}(\theta - \mathbf{m})) = \mathbf{k} - \mathbf{J}(\theta - \mathbf{m}), \tag{A.6}$$

where $\mathbf{k} = \mathbf{y} - \mathbf{g}(\mathbf{m})$. Proceeding with the calculus:

$$\begin{aligned}
\int LQ(\phi|\mathbf{y})d\phi &= -\frac{1}{2}(\theta - \mathbf{m}_0)^T \Lambda_0(\theta - \mathbf{m}_0) \\
&\quad - \frac{1}{2}sc\phi(\mathbf{y} - \mathbf{g}(\theta))^T \Sigma_v^{-1}(\mathbf{y} - \mathbf{g}(\theta)) + \text{const}\{\theta\} \\
&= -\frac{1}{2}(\theta^T \Lambda_0 \theta - \theta^T \mathbf{m}_0 - \mathbf{m}_0^T \Lambda_0 \theta) - \frac{1}{2}sc(-\mathbf{k}^T \Sigma_v^{-1} \mathbf{J} \theta \\
&\quad - \theta^T \mathbf{J}^T \Sigma_v^{-1} \mathbf{k} + \theta^T \mathbf{J}^T \Sigma_v^{-1} \mathbf{J} \theta - \theta^T \mathbf{J}^T \Sigma_v^{-1} \mathbf{J} \mathbf{m} \\
&\quad - \mathbf{m}^T \mathbf{J}^T \Sigma_v^{-1} \mathbf{J} \theta) + \text{const}\{\theta\} \\
&= -\frac{1}{2}\theta^T (\Lambda_0 + sc\mathbf{J} \Sigma_v^{-1} \mathbf{J}) \theta - \frac{1}{2}\theta^T (\Lambda_0 \mathbf{m}_0 + sc\mathbf{J}^T \Sigma_v^{-1} (\mathbf{k} + \mathbf{J} \mathbf{m})) \\
&\quad - \frac{1}{2}(\Lambda_0 \mathbf{m}_0 + sc\mathbf{J}^T \Sigma_v^{-1} (\mathbf{k} + \mathbf{J} \mathbf{m}))^T \theta.
\end{aligned} \tag{A.7}$$

Comparing this solution with the second term in A.3 can be deduced the update equations:

$$\Lambda = \Lambda_0 + sc\mathbf{J}^T\Sigma_v^{-1}\mathbf{J}; \quad (\text{A.8})$$

$$\Lambda\mathbf{m}_{\text{new}} = \Lambda_0\mathbf{m}_0 + sc\mathbf{J}^T\Sigma_v^{-1}(\mathbf{k} + \mathbf{J}\mathbf{m}_{\text{old}}). \quad (\text{A.9})$$

A.2 Noise Precision Parameters

In the same way as in the previous section, the noise precision hyper-parameters can be inferred from equation:

$$\log Q(\phi|\mathbf{y}) = \int LQ(\theta|\mathbf{y})d\theta. \quad (\text{A.10})$$

The log posterior is given by noise precision distribution:

$$\log Q(\phi|\mathbf{y}) = (c - 1) \log \phi - \frac{\phi}{s} + \text{const}\phi. \quad (\text{A.11})$$

Rewriting now the integral of the equation A.10 setting L as in A.2:

$$\begin{aligned} \int LQ(\theta|\mathbf{y})d\theta &= \int \left\{ -\frac{1}{2}\phi(\mathbf{y} - \mathbf{g}(\theta))^T\Sigma_v^{-1}(\mathbf{y} - \mathbf{g}(\theta)) + \frac{N}{2} \log \phi \right\} Q(\theta)d\theta \\ &\quad + \int \left\{ (c_0 - 1) \log \phi - \frac{1}{s_0}\phi + \text{const} \{\phi\} \right\} Q(\theta)d\theta \\ &= \left\{ \left(\frac{N}{2} + c_0 - 1 \right) \log \phi - \frac{1}{s_0}\phi - \frac{\phi}{2} \right\} \\ &\quad \times \int (\mathbf{y} - \mathbf{g}(\theta))^T\Sigma_v^{-1}(\mathbf{y} - \mathbf{g}(\theta))MVN(\theta; \mathbf{m}, \Lambda^{-1})d\theta, \end{aligned} \quad (\text{A.12})$$

and using the linearisation as in A.6 this last integral can be written as:

$$\begin{aligned} &= \int \left\{ \mathbf{k}^T\Sigma_v^{-1}\mathbf{k} - 2(\theta - \mathbf{m})^T\mathbf{J}^T\Sigma_v^{-1}\mathbf{k} \right\} MVN(\theta; \mathbf{m}, \Lambda^{-1})d\theta \\ &\quad + \int (\theta - \mathbf{m})^T\mathbf{J}^T\Sigma_v^{-1}\mathbf{J}(\theta - \mathbf{m})d\theta \\ &= \mathbf{k}^T\Sigma_v^{-1}\mathbf{k} + \text{tr}(\Lambda^{-1}\mathbf{J}^T\Sigma_v^{-1}\mathbf{J}). \end{aligned} \quad (\text{A.13})$$

Hence integral in A.10 becomes

$$= \left(\frac{N}{2} + c_0 - 1 \right) \log \phi - \frac{\phi}{s_0} - \frac{1}{2}\phi \left\{ \mathbf{k}^T\Sigma_v^{-1}\mathbf{k} + \text{tr}(\Lambda^{-1}\mathbf{J}^T\Sigma_v^{-1}\mathbf{J}) \right\} \quad (\text{A.14})$$

and the update equations are easily deduced:

$$c = \frac{N}{2} + c_0; \tag{A.15}$$

$$\frac{1}{s} = \frac{1}{s_0} + \frac{1}{2}(\mathbf{k}^T \Sigma_v^{-1} \mathbf{k} + \text{tr}(\Lambda^{-1} \mathbf{J}^T \Sigma_v^{-1} \mathbf{J})). \tag{A.16}$$

Appendix B

Levenberg-Marquardt Method

In the context of this work, the quantification method with whom the Variational Bayesian inference procedure is compared is the Weighted Nonlinear Least Squares (WNLLS) approach, analyzed in Section ???. A standard technique proposed by K. Levenberg before (1944) and D. Marquardt then (1963) to solve nonlinear least squares problems is the damped least squares method, better known as the Levenberg-Marquardt method. This approach is a combination of two minimization processes: the gradient descent method, in which the sum of the squared errors is reduced by updating the parameters in the direction of the greatest reduction of the least squares objective, and the Gauss-Newton method, in which the sum of the squared errors is reduced by assuming the least squares function as locally quadratic, and finding the minimum of the quadratic [12]. Thence, the L-M method acts more like a gradient descent method when the parameters are far from their optimal estimation and, differently, plays more like Gauss-Newton method when parameters are close to the optimal value. In details, the L-M approach takes the G-N equation [8],

$$\gamma_{new} = \gamma_{old} + H^{-1}\delta, \tag{B.1}$$

and introduces the scalar, α , initialized to a small value as

$$\gamma_{new} = \gamma_{old} + (H + \alpha \cdot \text{diag}(H))^{-1} \delta, \tag{B.2}$$

where H represents the Hessian matrix and δ is the increment from the initial value of the parameter. Availing of what obtained in the Appendix A, the Levenberg-Marquardt method can be implemented in the VB on the update for the means of the forward model parameters as follow:

$$\mathbf{m}_{\text{new}} = \mathbf{m}_{\text{old}} + (\Lambda + \alpha \cdot \text{diag}(\Lambda))^{-1} \Delta, \quad (\text{B.3})$$

where

$$\Delta = (sc\mathbf{J}^T \Sigma_e^{-1}(\mathbf{k} + \mathbf{J}\mathbf{m}_{\text{old}}) + \Lambda_0 \mathbf{m}_{\text{old}}) - \Lambda \mathbf{m}_{\text{old}}. \quad (\text{B.4})$$

Furthermore, α represents also the behaviour of the algorithm between the gradient descent update, when the scalar become large, and the G-N update, when it's small [12]. Therefore, in practice, if the convergence measure falls, i.e., F takes a backward step, then an update according to B.3 is attempt with $\alpha = 0.01$. If it follows a reduction in the free energy then VB update proceed, otherwise α is increased by a factor of 10 and the process repeted until F increases. Finally, if no improvement in the convergence measure can be found, which mean that *alpha* reaches a large value at which no significant change in \mathbf{m} is called for and the matrix to be inverted become diagonally dominant, then halt. The Levenberg-Marquardt procedure in the Variational Bayesian inference context concerns only the mean of the forward model parameter and the algorithm explained above takes place of “update” procedure in the pseudocode of the Algorithm 1 [8]. The stability of the method is guaranteed as long as the matrix to be inverted is not ill conditioned [13].

Bibliography

- [1] R. N. Gunn, A. A. Lammertsma, S. P. Hume, V. J. Cunningham. Parametric Imaging og Ligand-receptor Binding in PET Using A Simplified Reference Region Model. *NeuroImage*, vol. 6, pp. 279-287, 1997.
- [2] L. Tierney, J. B. Kadane. Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of American Statistical Association*, vol. 81, pp. 82-86, 1986.
- [3] T. Leonard, J. S. J. Hsu, K. W. Tsui. Bayesian Marginal Inference. *Journal of American Statistical Association*, vol. 84, pp. 1051-1058, 1989.
- [4] M. J. Beal. Variational Algorithms for Approximate Bayesian Inference. *PhD Thesis*, Gatsby Computational Neuroscience Unit, University College London, London, 2003.
- [5] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin. *Bayesian Data Analysis*, 2nd ed. London, U.K.: Chapman, Hall/CRC, 2003.
- [6] W. R. Gilks, S. Richardson, D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. London, U.K.: Chapman, Hall/CRC, 1996.
- [7] W. Penny, S. Kiebel, K. Friston. Variational Bayes. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Elsevier, London, 2006.
- [8] M. A. Chappell, A. R. Groves, B. Whitcher, M. W. Woolrich. Variational Bayesian Inference for a Nonlinear Forward Model. *IEEE Transactions on Signal Processing*, vol. 57, no. 1, 2009.

-
- [9] S. Kullback, R. A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics*, vol. 22, pp. 79-86, 1951.
- [10] M. A. Chappell, A. R. Groves, B. Whitcher, M. W. Woolrich. The FMRIB Variational Bayes Tutorial: Variational Bayesian Inference for a Non-Linear Forward Model. FMRIB Centre, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU.
- [11] H. Attias. A Variational Bayesian Framework for Graphical Models. *Advances in Neural Information Processing System*, MIT press, Cambridge, MA, 2000.
- [12] H. P. Gavin. The Levenberg-Marquardt Method for Nonlinear Least Squares Curve-fitting Problems. Departement of Civil and Environmental Engineering, Duke University, Durham, 2013.
- [13] G. A. F. Seber, C. J. Wild. *Nonlinear Regression*, Wiley, New York, 1989.
- [14] J. J. Moré. The Levenberg-Marquardt Algorithm: Implementation and Theory. *Numerical Analysis*, G. A. Watson, pp. 105-116, 1977.
- [15] T. F. Coleman, Y. Li. An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM Journal on Optimization*, vol. 6, pp. 418-445, 1996.
- [16] C. L. Lawson, R. J. Hanson. *Solving Least Squares Problems*, Prentice-Hall, chap. 23, p. 161, 1974.
- [17] G. Rizzo, M. Veronese, P. Zanotti-Fregonara, A. Bertoldo. Voxelwise Quantification of $[^{11}\text{C}](R)$ -rolipram PET Data: A Comparison Between Model-based and Data-driven Methods. *Journal of Cerebral Blood Flow & Metabolism*, vol 33, pp. 1032-1040, 2013.
- [18] P. Zanotti-Fregonara, S. S. Zoghbi, J. S. Liow, E. Luong, R. Boellaard, R. L. Gladding. Kinetic Analysis in Human Brain of $[^{11}\text{C}](R)$ -rolipram, A Positron

- Emission Tomography Radioligand to Image Phosphodiesterase 4: A Retest Study and Use of An Image-derived Input Function. *NeuroImage*, vol. 54, pp. 1903-1909, 2011.
- [19] S. S. Zoghbi, H. U. Shetty, M. Ichise, M. Fujita, M. Imaizumi, J. S. Liow. PET Imaging of the Dopamine Transporter with ^{18}F -FECNT: A Polar Radiometabolite Confounds Brain Radioligand Measurements. *Journal of Nuclear Medicine*, vol. 47, pp. 520-527, 2006.
- [20] G. Rizzo, F. E. Turkheimer, A. Bertoldo. Multi-scale Hierarchical Approach for Parametric Mapping: Assesment on Multi-compartmental Models. *NeuroImage*, vol. 67, pp. 344-353, 2013.
- [21] H. Ito, C. Halldin, L. Farde. Localization of 5-Ht_{1A} Receptors in the Living Human Brain Using [carbonyl- ^{11}C]WAY-100635: PET With Anatomic Standardization Technique. *Journal of Nuclear Medicine*, vol 40, pp. 102-109, 1999.
- [22] J. Gatliff, M. Campanella. The 18 kDa Translocator Protein (TSPO): A New Perspective in Mitochondrial Biology. *Current Molecular Medicine*, vol. 12, pp. 356-368, 2012.
- [23] G. Rizzo, M. Veronese, M. Tonietto, P. Zanotti-Fregonara, F. E. Turkheimer, A. Bertoldo. Kinetic Modeling Without Accounting for the Vascular Component Impairs the Quantification of [^{11}C]PBR28 Brain PET Data. *Journal of Cerebral Blood Flow & Metabolism*, vol. 34, pp. 1060-1069, 2014.
- [24] B. M. Mazoyer, R. H. Huesman, T. F. Budinger, B. L. Knittel. Dinamic PET Data Analysis. *Journal of Computer Assisted Tomography*, vol. 10, pp. 645-653, 1986.
- [25] A. Bertoldo, P. Vicini, G. Sambuceti, A. A. Lammertsma, O. Parodi, C. Cobelli. Evalutation of compartmental and spectral analysis model of [^{18}F]FDG

kinetics for heart and brain studies with PET. *IEEE Transactions On Biomedical Engineering*, vol. 45, pp. 1429-1448, 1998.

Acknowledgements

Firstly, I would like to thank Dr. Alessandra Bertoldo, who allowed me to work on such an ambitious and engaging project. I also would like to thank her, the co-supervisors Dr. Gaia Rizzo and Dr. Marco Castellaro and, last but not least, Mr. Matteo Tonietto for guiding me in this work and for all their valuable advices. It has been a great honour and pleasure to work with them.

I would like to thank my parents, Paola and Luigi, for always supporting me and helping me in finding my own way. I would like to dedicate this work to them, who make it possible.

I am also grateful to my brother, Manuele, for his advices.

Finally, I would like to thank my fellow students Alberto, Lorenzo, Massimo, Matteo and Riccardo, who make these years at university unforgettable.