

# Contents

<b>Riassunto della tesi</b>	<b>4</b>
<b>Introduction</b>	<b>10</b>
<b>1 Background on Bayesian estimators</b>	<b>14</b>
1.1 Introduction . . . . .	14
1.2 Mathematical formulation of Bayesian approaches . . . . .	15
1.2.1 WNLLS estimate . . . . .	15
1.2.2 Bayesian estimate . . . . .	16
1.3 A simple example of a Bayesian approach . . . . .	18
1.4 Discussion . . . . .	21
<b>2 Ridge regression: theory and validation on simulated data</b>	<b>22</b>
2.1 Linear ridge regression . . . . .	22
2.2 Evaluation of linear ridge regression . . . . .	24
2.2.1 Indices for performance evaluation . . . . .	24
2.2.2 The tracer $[^{11}\text{C}]\text{MP4A}$ . . . . .	24
2.2.3 Generation of simulated $[^{11}\text{C}]\text{MP4A}$ data . . . . .	26
2.2.4 Results . . . . .	27
2.3 Non-linear ridge regression . . . . .	30
2.3.1 Prior based on WNLLS: NLRR1 . . . . .	31
2.3.2 Priors based on clusters: NLRR2 . . . . .	32
2.3.3 Elimination of non linear minimizations: NLRR3 . . . . .	34
2.4 Evaluation of non-linear ridge regression . . . . .	36
2.4.1 The tracer $[^{11}\text{C}]\text{WAY100,635}$ . . . . .	36
2.4.2 Results . . . . .	37
2.4.3 Comments on computational cost . . . . .	40
2.4.4 Discussion on NLRR . . . . .	40
2.5 Parametric imaging of $[^{18}\text{F}]\text{FDG}$ in skeletal muscle . . . . .	40
2.6 Discussion . . . . .	44
<b>3 Population approaches: theory and validation on simulated data</b>	<b>45</b>
3.1 A few general ideas . . . . .	45
3.2 The two-stages methods . . . . .	46
3.2.1 TS . . . . .	46
3.2.2 ITS . . . . .	47
3.2.3 GTS . . . . .	47
3.2.4 Comparison between ITS and GTS . . . . .	48

3.3	Important remarks on ITS and GTS . . . . .	49
3.3.1	Termination criteria . . . . .	49
3.3.2	Segmentation into sub-populations . . . . .	49
3.4	Comparison between ridge regression and Two-Stages algorithms . . . . .	50
3.4.1	Analogies . . . . .	50
3.4.2	Differences . . . . .	51
3.5	The tracer $[^{18}\text{F}]\text{FDG}$ . . . . .	51
3.5.1	Patlak plot . . . . .	52
3.6	Performances of Two-Stages algorithms . . . . .	53
3.6.1	Generation of the basic simulated data set . . . . .	53
3.6.2	Results: different noise levels . . . . .	55
3.6.3	Results: population homogeneity . . . . .	55
3.6.4	Results: number of subjects forming the population . . . . .	57
3.6.5	Results: amount of data available for each subject . . . . .	57
3.6.6	Brief summary of the main outcomes . . . . .	59
3.7	Pixel level simulations: linear case . . . . .	59
3.7.1	Simulated data set . . . . .	60
3.7.2	Segmentation . . . . .	60
3.7.3	Termination criteria . . . . .	61
3.7.4	Results . . . . .	62
3.7.5	Comparison with linear ridge regression . . . . .	63
3.8	Pixel level simulations: nonlinear case . . . . .	65
3.8.1	Results . . . . .	67
3.8.2	Comparison with non linear ridge regression . . . . .	68
3.9	Recovery of information in unreliable scans . . . . .	70
3.9.1	Adaptation of GTS to the problem . . . . .	72
3.9.2	Adaptation of ridge regression to the problem . . . . .	73
3.9.3	Results . . . . .	73
3.10	Discussion . . . . .	75
<b>4</b>	<b>Applications to a clinical dataset:<math>[^{11}\text{C}]</math>-(R)-PK11195</b>	<b>77</b>
4.1	Introduction to $[^{11}\text{C}]$ -(R)-PK11195 . . . . .	77
4.2	Inclusion of the vascular components into the SRTM . . . . .	79
4.2.1	Standard SRTM equation . . . . .	79
4.2.2	Modifications of the standard SRTM equation . . . . .	80
4.2.3	Estimation of the whole blood TAC . . . . .	81
4.3	Application of ridge regression . . . . .	82
4.3.1	Non linear version . . . . .	82
4.3.2	Linear version . . . . .	84
4.4	Application of GTS . . . . .	85
4.5	Computational cost . . . . .	86
4.6	Examples of parametric maps . . . . .	86
4.7	Correlation between ROI and pixel estimates . . . . .	92
4.8	Additional analysis on binding potentials . . . . .	93
4.8.1	Correlation with SRTMV BPs . . . . .	95
4.8.2	BP variability within a certain ROI . . . . .	95
4.9	Differentiation between healthy and HD subjects . . . . .	97
4.9.1	Statistical tests employed . . . . .	97

4.9.2	Sensitivity and specificity . . . . .	99
4.9.3	Results: sensitivity . . . . .	100
4.9.4	Results: specificity . . . . .	100
4.10	Summary of main results on [ <sup>11</sup> C]-(R)-PK11195 . . . . .	103
4.11	Discussion . . . . .	104
<b>Conclusion</b>		<b>105</b>
<b>Bibliography</b>		<b>106</b>

# RIASSUNTO DELLA TESI

## INTRODUZIONE

La PET (dall'inglese Positron Emission Tomography, ovvero tomografia ad emissione di positroni) è una tecnica sviluppatasi a partire dagli anni '70 per l'analisi in vivo di importanti processi biologici. Essa permette la quantificazione di parametri fisiologici significativi quali, ad esempio, consumo locale di glucosio e flusso sanguigno. Nella PET un tracciante radioattivo che decade per emissione positronica viene sintetizzato tramite opportune procedure chimiche e iniettato nel soggetto sotto esame, in dosi tali da garantire la non tossicità. Il cuore della complessa strumentazione utilizzata per la misurazione è costituito da cristalli scintillatori in grado di rilevare i due fotoni che vengono emessi quando un positrone, emesso dal tracciante, collide con un elettrone della materia del tessuto circostante; un processo di annichilazione positrone-elettrone ha così luogo e le due particelle si trasformano in pura energia, dando luogo a due fotoni. Da queste misurazioni, attraverso sofisticati algoritmi di ricostruzione che tengono conto di fenomeni complessi quali quelli di attenuazione, scattering e dead time, si riesce a localizzare con precisione la posizione in cui l'evento di emissione del positrone ha avuto luogo, e, in definitiva, si riesce a ricostruire la distribuzione nel tempo del tracciante nell'area di interesse.

Dopo la ricostruzione l'immagine è pronta per l'analisi quantitativa, comunemente detta kinetic modelling, essenziale perchè da essa si ricavano i parametri di interesse fisiologico, detti parametri cinetici. Questa analisi può essere effettuata a livello di ROI (Region Of Interest, cioè regioni anatomicamente omogenee quali ad esempio cervelletto, talamo, ippocampo) o a livello di pixel. In quest'ultimo contesto i parametri cinetici vengono calcolati per ciascuno delle centinaia di migliaia (talvolta milioni) di pixel che costituiscono l'immagine 3D e vengono generate le cosiddette mappe parametriche. L'analisi a livello di pixel è caratterizzata da problemi dovuti al fatto che la TAC (Time Activity Curve, cioè l'andamento della concentrazione del tracciante in funzione del tempo) di un generico pixel è tipicamente affetta da un significativa quantità di rumore, e questo può provocare la non convergenza degli algoritmi usati per stimare i parametri cinetici, o può dar luogo a valori non fisiologici degli stessi. A livello di ROI, d'altro lato, il rumore è trascurabile poichè quando si calcola la TAC di una ROI come media delle TACs dei pixel di quella ROI si ha un effetto di regolarizzazione dovuto al fatto che l'operazione di media ha un chiaro impatto di attenuazione del rumore. Le mappe parametriche, tuttavia, hanno un'importanza notevole poichè sono caratterizzate da un'alta risoluzione spaziale e quindi consentono l'individuazione di aree patologiche, quali una lesione localizzata, un'occlusione, la presenza di una piccola massa tumorale, che potrebbero non essere rilevabili a livello di ROI.

Scopo della presente tesi è stato quello di analizzare metodiche per l'analisi a livello di pixel che consentano una generazione di mappe parametriche più affidabili, condizione importante specie per il loro utilizzo in ambito clinico. In particolare è stata analizzata e ulteriormente sviluppata una metodica, già proposta in letteratura, denominata ridge regression (RR) ed è stato inoltre proposto e testato un metodo matematico, completamente nuovo in ambito PET, denominato GTS (Global-Two-Stage) appartenente al gruppo degli approcci di popolazione. L'idea di fondo di queste metodiche, che le fa rientrare nella famiglia degli approcci Bayesiani, è, semplificando, quella di non utilizzare nella stima dei parametri di un dato pixel solo e soltanto la TAC del pixel, ma di incorpo-

rare in qualche modo informazioni derivanti dagli altri pixel al fine di ottenere un globale effetto di regolarizzazione penalizzando, ad esempio, le TACs particolarmente rumorose. Le procedure Bayesiane richiedono di calcolare un valore atteso  $\mu$ , detto prior, per il vettore dei parametri, assieme alla matrice di covarianza  $\Delta$  che è indice dell'affidabilità del prior; prior e covarianza associata verranno poi utilizzati nella stima finale dei parametri del modello analizzato, che, nel caso generale, si otterranno minimizzando una funzione del tipo

$$[y - h(p)]'\Sigma^{-1}[y - h(p)] + (p - \mu)'\Delta^{-1}(p - \mu)$$

in cui  $y$  indica il vettore dei dati misurati,  $\Sigma$  la matrice di covarianza dell'errore di misura e  $h(p)$  la funzione non lineare che esprime i dati in funzione dei parametri incogniti. Mentre nell'ottica del tradizionale WNLLS (Weighted NonLinear Least Squares), il vettore dei parametri  $p$  è quello che minimizza

$$[y - h(p)]'\Sigma^{-1}[y - h(p)]$$

in ottica Bayesiana, quindi, un termine aggiuntivo che incorpora l'informazione a priori è presente.

## MATERIALI E METODI

L'analisi delle metodologie considerate (RR, sia nella sua versione lineare che in quella non-lineare, e GTS) è stata svolta in due fasi. E' stata inizialmente effettuata un'analisi su dati simulati; questo si è rivelato necessario poichè, per calcolare indici che quantificano la correttezza delle stime finali quali BIAS e Root Mean Square Error (RMSE), la conoscenza dei "veri" parametri è necessaria e per far ciò i dati devono necessariamente essere simulati, poichè, nei dataset reali, il "vero" valore dei parametri non è noto. Questa analisi è l'oggetto dei capitoli 2 (RR) e 3 (GTS). I traccianti utilizzati per le simulazioni, [ $^{11}\text{C}$ ]MP4A (RR lineare), [ $^{18}\text{F}$ ]FDG (GTS nel caso lineare) e [ $^{11}\text{C}$ ]WAY100,635 (RR non-lineare e GTS nel caso non-lineare) sono fra i più diffusi nella PET; in particolare [ $^{18}\text{F}$ ]FDG è molto usato in quanto permette di stimare il consumo locale di glucosio a livello tessutale e fornisce quindi significative indicazioni sul metabolismo cellulare in vivo. Nella generazione dei dati simulati notevole attenzione è stata prestata nel creare mappe con una distribuzione dei parametri il più possibile identica a quella reale, e nel generare pixel TACs con un livello di rumore paragonabile a quello che sperimentalmente si riscontra. Le performances delle varie metodiche proposte sono state poi confrontate con quelle dell'appropriato "gold standard", cioè del metodo comunemente più usato per la stima parametrica per il tracciate sotto esame (ad esempio WNLLS o metodo di Patlak). Nel capitolo 4 l'interesse è stato rivolto ad un ricco dataset reale del tracciante [ $^{11}\text{C}$ ](R)-PK11195, molto usato nello studio di patologie quali Alzheimer, Huntington ed altre, in quanto esso è legato al livello di attivazione delle cellule microgliali cerebrali, e di conseguenza al grado di infiammazione cerebrale. Il metodo standard di confronto, in questo contesto, è stato il SRTMV (Simplified Reference Tissue Model, modificato per includere le componenti vascolari) e il confronto fra le performance dei vari algoritmi è stato effettuato sulla base della valutazione visiva della mappe risultanti, sulla correlazione fra i parametri stimati dalle TACs di alcune ROIs e quelli ottenuti come media dai pixel delle stesse ROIs, e sulla capacità di differenziare fra sani e malati poichè il dataset comprendeva scan sia di soggetti affetti dalla malattia di Huntington e che di soggetti sani. Di particolare rilievo è l'ultimo criterio, in quanto una migliore

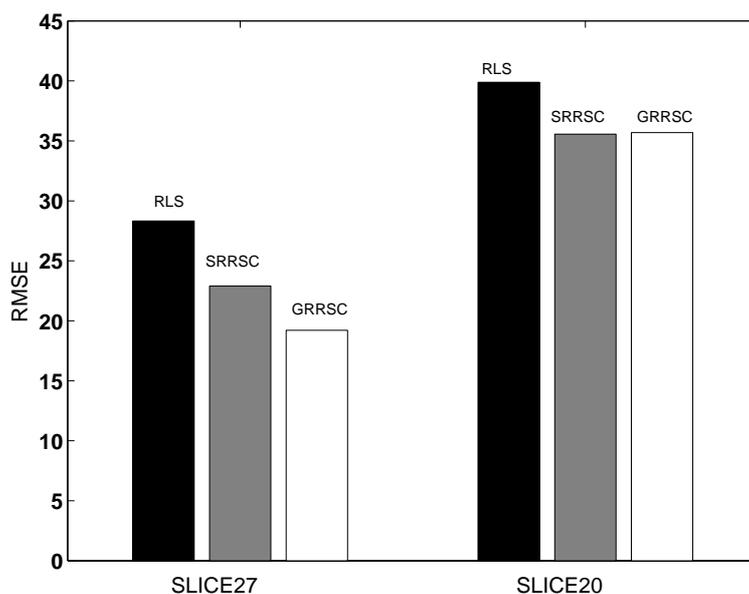


Figure 1: *Il RMSE per il parametro  $k_3$ , mediato su tutti i pixel, ottenuto con RLS, SRRSC e GRRSC è rappresentato per le fette n°27(sinistra) e n°20(destra).*

differenziazione sano-malato, effettuata nel caso del  $[^{11}\text{C}]$ -(R)-PK11195 sulla base di un parametro detto Binding Potential (BP) che quantifica il livello di legame del tracciante, è chiaramente un fatto positivo in quanto garantisce diagnosi più affidabili e meno affette da errori.

## RISULTATI

L'analisi di dati simulati ha indicato che RR e GTS provocano sempre la diminuzione del RMSE, lasciando il BIAS sostanzialmente invariato o persino riducendolo. La figura 1, ad esempio, illustra la riduzione del RMSE per il parametro  $k_3$  per le simulazioni relative al tracciante  $[^{11}\text{C}]$ MP4A. Il "gold standard" per  $[^{11}\text{C}]$ MP4A, RLS, è confrontato con due versioni della linear ridge regression, SRRSC and GRRSC per simulazioni relative a due diverse fette dell'immagine (n°21 e n°27). Chiaramente, in generale, i miglioramenti dipendono dal tracciante, dal livello di rumore e del parametro cinetico analizzato.

L'analisi del ricco data set di  $[^{11}\text{C}]$ -(R)-PK11195 ha mostrato come RR e GTS rendano le mappe parametriche molto più regolari rispetto al SRTMV. La figura 2 mostra, ad esempio, mappe del parametro BP ottenute con SRTMV (sinistra) e GTS (destra): si noti come nella mappa GTS regioni anatomicamente omogenee (ad esempio il cervello) si riescano nettamente a distinguere all'interno della mappa stessa.

I metodi proposti si sono inoltre rivelati particolarmente efficaci nell'incrementare la differenziazione fra sani e malati, effettuata a livello di pixel a partire dalle mappe parametriche. Questo fatto è illustrato nella figura 3, che rappresenta la percentuale di pixel riconosciuti come "malati" nel confronto fra HD e soggetti sani, per lo standard SRTMV, per due diverse versioni di non linear ridge regression (NLRR3a e NLRR3b), per il GTS e per LRR.

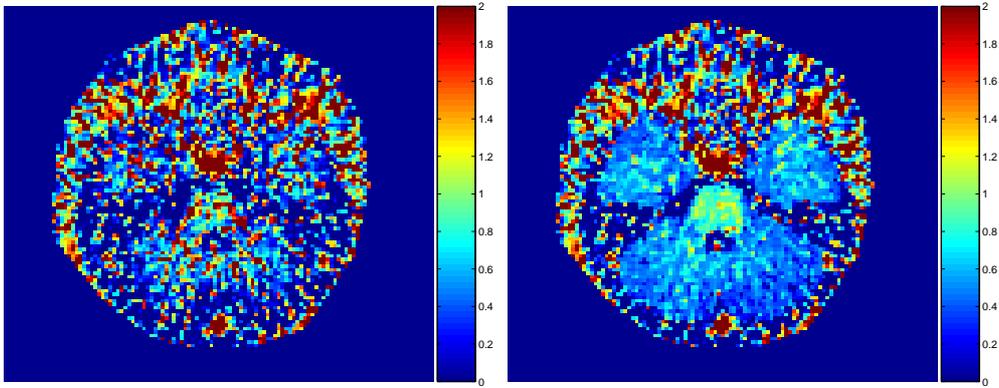


Figure 2: Mappe di BP per la fetta n°21 di un soggetto HD. La mappa ottenuta con SRTMV (sinistra) è confrontata con quella ottenuta con GTS (destra).

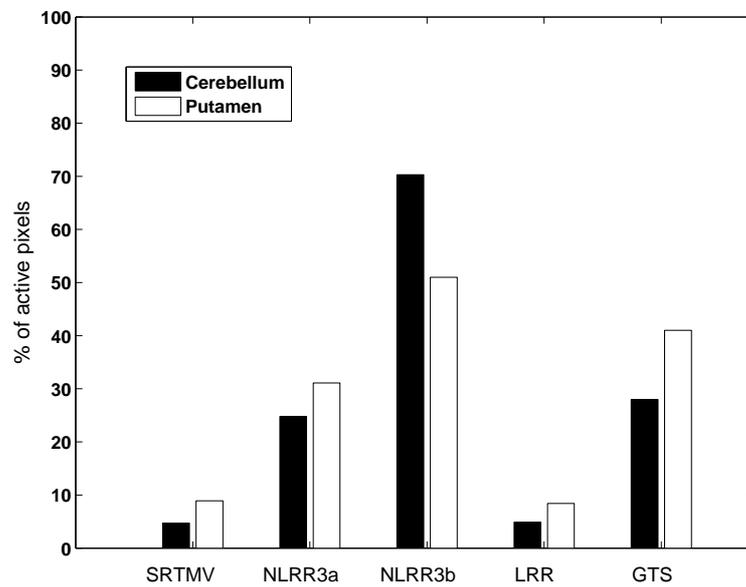


Figure 3: La percentuale di pixel "malati" di cervelletto e putamen ( $p$ -value minore di 0.01) è raffigurata per SRTMV e per 4 approcci Bayesiani. Un one-tailed Wilcoxon rank-sum test è stato qui usato per calcolare i  $p$ -values.

## DISCUSSIONE

I metodi bayesiani sono un utile strumento per migliorare la qualità delle mappe parametriche. La ridge regression (RR) è stata finora l'unico metodo proposto nella PET per affrontare il problema della generazione di mappe più affidabili in ottica bayesiana. In questa tesi un confronto sistematico fra i metodi di ridge regression, che non era mai stato fatto nel passato, è stato realizzato. Alcune nuove procedure per l'applicazione della RR a problemi non lineari basate sulla linearizzazione dell'equazione non lineare e/o sull'uso della cosiddetta cluster analysis sono state inoltre proposte e confrontate. In aggiunta, la metodica del global-two-stages GTS, nuova in campo PET, è stata per la prima volta applicata per la formazione di immagini parametriche.

I capitoli 2 e 3 hanno indicato, attraverso l'analisi di dati simulati, che RR e GTS provocano sempre la diminuzione del RMSE, lasciando il BIAS sostanzialmente invariato o persino riducendolo. I miglioramenti possono variare da poco significativi -se le mappe originali sono già di buona qualità- a notevoli; in ogni caso, comunque, nessuna diminuzione delle prestazioni è stata mai notata in termini di BIAS e RMSE. Contrariamente alla ridge regression lineare, differenti versioni di ridge regression non lineare sono state esaminate. Il metodo che è stato selezionato come ottimale richiede, per calcolare i priors a livello di pixel, di eseguire il fit del modello appropriato al tracciante di interesse sulle TACs determinate attraverso cluster analysis. Nell'applicazione ad immagini reali tridimensionali, il calcolo del prior può essere effettuato tramite l'uso di atlanti anatomici che vanno impiegati per associare ad un generico pixel i parametri della ROI alla quale quel pixel appartiene.

Il GTS ha due impostazioni che sono in qualche modo a discrezione dell'utente. Le prestazioni del GTS non hanno mai mostrato di dipendere significativamente dal metodo impiegato per arrestare le iterazioni; in ogni caso, il criterio di terminazione consistente nell'arrestare l'algoritmo quando la differenza relativa del parametro di interesse fra l'iterazione corrente e la precedente è più bassa dell'1% per almeno il 90% degli individui nella popolazione considerata si è rivelato essere robusto in ogni situazione esaminata. I risultati, d'altro lato, hanno dimostrato una significativa dipendenza dalla tecnica di segmentazione impiegata per suddividere il gruppo di pixel in sottopopolazioni omogenee, alle quali applicare poi separatamente l'algoritmo. Se l'immagine è normalizzata, gli atlanti stereotassici possono essere impiegati per generare le sottopopolazioni costituite dai pixel che appartengono alla stessa ROI. A volte, tuttavia, la normalizzazione non può essere realizzata perché non c'è nessuna risonanza magnetica disponibile, o perché essa causerebbe una perdita indesiderata di risoluzione spaziale se l'immagine è stata acquisita su di uno scanner ad alta risoluzione. In questo caso una segmentazione di ogni fetta dell'immagine in un piccolo numero di sottopopolazioni, effettuata in base al parametro di interesse, è una buona soluzione alternativa che consente di effettuare il GTS su popolazioni relativamente omogenee.

L'analisi del ricco data set di [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 ha mostrato come RR e GTS rendano le mappe parametriche molto più regolari rispetto al SRTMV: ciò era del tutto previsto poiché i metodi bayesiani hanno un effetto di regolarizzazione. Inoltre, i metodi proposti si sono rivelati particolarmente efficaci nell'incrementare la differenziazione fra sani e malati, effettuata a livello di pixel a partire dalle mappe parametriche, e questo è notevolmente significativo in ottica diagnostica, in quanto elevate sensibilità e specificità sono requisiti essenziali per applicazioni in ambito clinico.

La ridge regression e il global-two-stage, in conclusione, sono metodi preziosi per il

miglioramento delle mappe parametriche nella PET. Entrambe le metodiche possono essere impiegate con pressochè ogni tipo di tracciante e modello, ed hanno quindi una vasta gamma di applicabilità. I metodi proposti, inoltre, hanno un costo computazionale basso che li rende adatti alla formazione di immagine parametriche, anche con i recenti scanner che forniscono immagini ad elevatissima risoluzione caratterizzate quindi da un altissimo numero di pixel (anche un milione).

# INTRODUCTION

PET (Positron Emission Tomography) is a widely used technique for the in-vivo analysis of important biological processes in both animals and humans. In PET, radioactive tracers which decay by positron emission are injected -or, less frequently, inhaled- and specific scintillator crystals are employed to detect the two 511 Kev photons which are generated when any emitted positron collides with any electron of the surrounding matter, thus annihilating. The distribution of the injected tracer over time in the area of interest is then derived from these measurements through sophisticated reconstruction algorithms which keep into account, and correct for, several complex phenomena such as scattering, dead time, tissue attenuation.

After the process of reconstruction the PET image is ready for quantitative analysis that allows to compute from the image parameters with a physiological meaning, such as blood flow, rate of glucose consumption, amount of binding of the tracer to its specific receptors. The basic principles of PET quantitative analysis, which is called kinetic modelling, were established in a series of papers dating back to the early 80s and extending to the mid 90s. PET kinetic modelling, besides the image itself, requires the knowledge of the so-called input function: this can be either the time course of the tracer concentration in plasma or the time course of the tracer concentration in an appropriate region, called reference region, devoid of receptors specific for the tracer under examination. The basic dogma of PET kinetic analysis states that the known input function and the image are to be analyzed through an appropriate model to extract the corresponding so-called *kinetic parameters*, whose number and biological meaning depend on the method and model applied. Most of these parameters have a biological meaning (e.g uptake of the tracer into the tissue, clearance of the tracer) and their accurate quantification is essential for the application of PET in clinical practise.

The methods developed to perform kinetic modelling may require an assumption on the compartmental structure describing the tracer under investigation ("compartmental analysis") or may not make specific assumptions on this structure ("data-driven" approaches like Patlak plot, Logan plot, spectral analysis, basis pursuit). In addition, the quantitative determination of kinetic parameters can be made either at ROI level (Region-Of-Interest, e.g. cerebellum, thalamus, putamen, basal ganglia) or at pixel level. Pixels are used here as synonym of voxels -the three-dimensional (3D) elements of the reconstructed image. Kinetic modelling -i.e. the estimation of kinetic parameters- at pixel level is performed using the values of radioactive concentration of the tracer as a function of the scan time (time activity curve or TAC) of each pixel separately. The TAC of a given ROI, computed as average of the pixels belonging to that region, is instead used in ROI analysis, which is characterized by more robust results since the average of the pixels TACs in the region of interest is employed, with a dramatic noise reduction. Pixel-by-pixel maps, however, are of paramount importance since they are characterized by a much higher spatial resolution. A phenomenon such as a lesion in a subsection of a cerebral structure may be invisible if only ROI analysis is performed whereas it may become evident at visual inspection if a pixel-wise analysis is carried out. In addition pixel-by-pixel analysis is essential, for instance, to compare 3D parametric maps of healthy subjects to the corresponding maps of ill subjects in order to determine which areas are those most affected by a specific disease, or which tracers are the best in differentiating between healthy and ill subjects in the process of diagnosis of a given

disease. The drawback of pixel-by-pixel analysis is the increased amount of noise of pixel TACs with respect to ROI TACs; this fact may lead to problems in the numerical identification of the kinetic parameters under investigation (e.g. no numerical convergence of the algorithm used or non-physiological values of the results). Another practical aspect that should be addressed carefully is computational time. If a ROI analysis typically involves a dozen of regions for each subject at most, with the current scanner resolution the number of pixels of a human brain may reach the number of 500.000 and the issue of the time required to perform kinetic analysis at each pixel becomes relevant.

To address the issue of the generation of parametric maps -i.e. maps that associate to each pixel of the image a specific kinetic parameter- Bayesian techniques have been taken into consideration in this thesis. A Bayesian approach, loosely speaking, requires to determine an appropriate expected value for the parameters to be identified (the so-called prior) with a corresponding covariance matrix which is an index of the reliability of the prior. In the process of parameter estimation, then, to the term which in the usual Weighted Non Linear Least Squares (WNLLS) keeps into account measured data, a term containing the prior and its covariance is added. The parameter vector is then calculated to minimize the sum-of-squares "distance" not only from the data -as in conventional weighted least squares- but also from the prior, the latter suitably weighted through the covariance matrix associated to the prior. In this thesis, the mathematical theory of WNLLS and Bayesian methods will be presented in Chapter 1, which was consequently named "Background" chapter. A first simple example of Bayesian method is also presented in order to illustrate the potentialities of this class of methods.

Few Bayesian approaches are currently used in PET : one remarkable example is given by *ridge regression*. Ridge regression techniques are analyzed in Chapter 2, where the theory behind these methods is presented and examples of applications on simulated data are given. In the so-called "linear ridge regression" the kinetic model is fitted to each pixel TAC separately, and then the prior is computed by averaging the parameter estimates of the pixels located in the neighbourhood of the pixel of interest. Simulated data from the tracers  $[^{11}\text{C}]\text{MP4A}$  and  $[^{18}\text{F}]\text{FDG}$  were employed to test the algorithm. In the "non-linear" version of ridge regression the issue of computational time prevents from applying the same principles of linear ridge regression. A feasible approach to overcome this problem is based on the preliminary use of clustering algorithms, which yield clusters whose TACs are fitted to the model of interest and whose resulting kinetic parameters are used as priors for the analysis at pixel level. A first-order linearization of the non-linear model around an appropriate preliminary estimate of the parameter vector, besides, may be used to transform the original problem into a linear problem, thus reducing computational time. These two principles can be applied together, or separately, and the three resulting versions of non-linear ridge regression were tested on simulated  $[^{11}\text{C}]\text{WAY100,635}$  data. In the end an example of the potentiality of non-linear ridge regression on a real  $[^{18}\text{F}]\text{FDG}$  data set is presented, as ridge regression allowed to estimate parameters at pixel level even for the 5k model employed to describe kinetics of  $[^{18}\text{F}]\text{FDG}$  in human skeletal muscle.

Another possible approach to improve the quality of parametric maps is based on techniques originally developed in the pharmacokinetic/pharmacodynamic sector and globally referred to as "population" approaches, which are new to the field of PET. These techniques are described in Chapter 3. With these methods all the homogenous subjects which underwent the same experimental protocol are studied together, with the aim of identifying, in each subject, the individual kinetic model parameters. The

intersubject homogeneity is then exploited to improve the parameter estimates for those subjects for which standard WNLLS yielded unreliable estimates or failed to converge. The population parameter estimation paradigm -i.e. poor individual data sets borrow strength from the others if there is some homogeneity- can be clearly applied in a number of situations different from the one just described. In the case considered here, the "subjects" are pixels instead on individuals, and the "population" can be considered as made of those pixels belonging to a same ROI, or slice, or showing some homogeneity according to a given criterion. The attention was focused on Two-Stages iterative algorithms Global-Two-Stages (GTS) and ,to a minor extent due to computational reasons, Iterative-Two-Stages (ITS). It is a matter of controversy whether GTS can be considered or not a Bayesian approach. These methods update at every iteration the estimates of each element of the population deriving the prior as an average from the population, and the covariance of the prior from the population parameter variability and reliability.

As in Chapter 2 for ridge regression, performances of GTS were assessed by means of simulations on synthetic data sets also in Chapter 3, comparing the proposed approaches to the appropriate "gold standard" for the analysis at pixel level and using Bias and Root Mean Square Error as indices to evaluate the algorithms. Even though there is no consensus on the ideal method for generating realistic simulated data, the use of such data was unavoidable; in fact, the need to know "real" parameters was central to understand if, and to which extent, the use of the proposed approaches lead to more correct and reliable parameter estimates.

After testing their performances on simulated data, ridge regression and GTS were applied to *real* 3D [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 images and Chapter 4 contains the corresponding results. One of the most used approach in kinetic modelling, the simplified reference tissue model SRTM, in a different version keeping into account also the vascular volume component (SRTMV), was used as method of comparison. In the large data set analyzed the use of simple visual inspection of parametric maps and of correlation between parameters estimated at ROI and pixel level were employed as criteria to evaluate the algorithms. The correlation between Binding Potential (BP), one of the most important kinetic parameters, estimated through SRTMV, and the same parameter computed with the analyzed Bayesian approaches was also considered together with the variability of this kinetic parameter within a given ROI. In addition, the ability to efficiently differentiate between healthy subjects and patients (in the specific case patients suffering from Huntington disease) was also taken into account.

Before concluding it is worth to shortly summarize the main purposes PET is used for. Many neurodegenerative diseases (e.g. Alzheimer's disease, AD), nowadays, are diagnosed on the basis of psychological tests; it has been shown, however, that for many of them PET guarantees a better specificity -ability to correctly diagnose an ill subject- and sensitivity -ability to correctly diagnose an healthy subject. PET, besides, allows in some cases to differentiate between similar diseases, diagnosis that in some cases couldn't be possible otherwise. This is possible thanks to the fact that, for the most diffused neurodegenerative diseases, specific tracers with different uptake in presence or absence of the neuropathological condition have been developed, allowing to distinguish between ill and healthy subjects. In addition, there is accumulating evidence supporting the idea that PET could be used to get a reliable diagnosis before clinical symptoms manifest, which is particularly interesting in that early diagnosis allows an early treatment. It is quite obvious how these potentialities can also be used, for instance, to evaluate the effectiveness a newly developed drug or treatment, and how, therefore, the impact of

PET in the world of health care is central.

In conclusion, the problem of a fast and reliable estimation of kinetic parameters at pixel level is central for every tracer used in PET. Bayesian approaches, through the use of a-priori information which can reduce the impact of noise, can be successfully employed to obtain more reliable parametric maps. It is important to point out that both ridge regression based algorithms and population approaches lend themselves to be used with virtually every tracer and model, even though performances clearly depend on several factors such as level of noise and model complexity. Given the high number of tracers developed to test the various neurotransmitter systems (e.g [<sup>11</sup>C]raclopride and [<sup>18</sup>F]fallypride for the dopaminergic system, [<sup>11</sup>C]DASB and [<sup>11</sup>C]WAY100,635 for the serotonergic and so on) the possibility to generate more reliable parametric maps has an important impact in daily clinical practise for diagnosis, for the assessment of efficacy of newly developed drugs, for the possibility to predict the future appearance of a given disease with the consequent possibility of a preventive and therefore more effective treatment.

# Chapter 1

## Background on Bayesian estimators

### 1.1 Introduction

As already stated in the introductory part, a Bayesian approach requires to integrate some kind of a-priori information into the model of interest to get in the end more reliable or more physiological values for the parameters of interest. A question naturally arises: how do we get the a-priori information which is used in the process of parameter estimation?

A first strategy makes use of reasonable assumptions which do not rely on the measured data; for instance one can postulate that the parameter vector has a Gaussian distribution, with known expected value and covariance matrix. This approach has the evident drawback that the assumptions, albeit in general reasonable, may not be appropriate for the data set analyzed and may lead to erroneous results.

A second possible approach requires to use first the measured data alone and then to derive, from the first estimation process, appropriate parameter and/or covariance estimates that will be then used in the subsequent final estimation. Basically the a-priori knowledge is derived from the data itself; both ridge regression and population approaches are examples in this sense. This simple yet effective concept will be now briefly illustrated with one example derived from [1], whereas the mathematical details of Bayesian approaches, and of the maximum-a-posteriori (MAP) estimator in particular, will be described in the following paragraph.

The simplified reference tissue model (SRTM), proposed for the first time in [2], is one of the most diffused approaches for kinetic analysis in brain studies. The basic assumption of the SRTM is that there exists a region, called reference region, devoid of receptors specific for the tracer under examination, so that the compartment associated to the specific binding can be considered absent in the reference region; all other cerebral regions are referred to as target regions. The parameters that can be estimated are  $R$  [unitless],  $k_{2,TAR}$  [ $min^{-1}$ ] and  $BP$  [unitless], where  $R$  is the relative delivery, which is the ratio of the influx constant between the target region and the reference region, and  $k_{2,TAR}$  and  $BP$  indicate respectively the outflux rate and the binding potential of the target region. Having denoted with  $K_1$  the influx rate from plasma to tissue, another assumption of SRTM is

$$\frac{K_{1,TAR}}{k_{2,TAR}} = \frac{K_{1,REF}}{k_{2,REF}}$$

which allows to write

$$k_{2,TAR} = k_{2,REF} R$$

with

$$R = \frac{K_{1,TAR}}{K_{1,REF}}$$

as explained before. Therefore the model can be re-parameterized, using  $R$ ,  $k_{2,REF}$  and  $BP$  (SRTM2) by substituting to  $k_{2,TAR}$  ( $k_{2,REF} R$ ). The application of SRTM2 to a given set of ROIs or pixels should in theory provide the same value of  $k_{2,TAR}$  for each ROI or pixel whereas, due to noise and to the approximations inherent into the model applied, this doesn't happen. In [1], through the use of both simulated and real data sets of [ $^{18}F$ ]FCWAY, [ $^{11}C$ ]flumazenil, and [ $^{11}C$ ]raclopride, SRTM2 was first applied to all the pixels of the brain. Then  $k_{2,REF}$  was fixed to the global median computed considering all pixels but those of the reference region; the median was chosen as it was less biased than the mean being less sensitive to outliers. SRTM2 was then applied again, this time with  $k_{2,REF}$  fixed, and relevant noise reduction in R and BP parametric maps were thus obtained.

This is a very interesting example of Bayesian approach in which the a-priori information -the value of  $k_{2,REF}$ - is derived from the data. Especially in presence of noisy kinetics and therefore especially at pixel level, as a matter of fact, reducing the number of parameters to be estimated by fixing one of them normally gives rise to more reliable estimates of the others. It is probably straightforward to point out that fixing the parameter to an incorrect value may lead to completely wrong results so each assumption should be checked carefully.

## 1.2 Mathematical formulation of Bayesian approaches

### 1.2.1 WNLLS estimate

Consider a generic dynamic model with  $m$  unknown parameters  $p_1, p_2, \dots, p_m$  and assume that  $N$  samples collected at times  $t_1, t_2, \dots, t_N$  of a signal which brings information on the parameter vector are available. Assume also that the following equation, written in vectorial form as

$$y = h(p) + e \quad (1.1)$$

holds, where  $e$  is assumed to be a zero mean Gaussian random vector made of random variables  $e_1, e_2, \dots, e_N$  describing the additive errors affecting the measurements, and  $y$  is a  $N$ -dimensional column vector containing the measured data. For the sake of simplicity the explicit dependence on time will be omitted : in other words instead of indicating the function of  $m$  parameters  $p_i$  as  $h(p, t_i)$ ,  $i=1, 2, \dots, N$ , or  $h(p, t)$ , the simpler  $h(p)$  will be employed. Indicated with  $\Sigma$  the covariance matrix of the measurement error, assumed to be known at least up to a proportionality constant, the Weighted NonLinear Least Squares (WNLLS) estimate of the vector  $p$  is the one which minimizes the cost function:

$$[y - h(p)]' \Sigma^{-1} [y - h(p)] \quad (1.2)$$

which is the weighted distance between the observed data  $y$  and fitted data  $h(p)$ . In general  $h$  is a non-linear function of  $p$  and the minimization problem (1.2) has to be solved through iterative and computationally expensive algorithms such as Gauss-Newton's, or

better, through modified versions of it. If  $h$  is a linear function of  $p$  -i.e.  $h(p)=Xp$  holds, where  $X$  is an appropriate  $N \times m$  matrix- the well-known expression

$$p = (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}y) \quad (1.3)$$

provides the desired solution allowing to compute  $p$  quickly. In PET  $\Sigma$  is known up to a proportionality factor, i.e.

$$\Sigma = \sigma^2 W^{-1} \quad (1.4)$$

with  $W$  known and  $\sigma^2$  unknown. Inserting this expression into (1.2) and (1.3) and neglecting the constant  $\sigma^2$ , which is irrelevant in the determination of the point of minimum, one obtains

$$[y - h(p)]'W[y - h(p)] \quad (1.5)$$

$$p = (X'WX)^{-1}(X'Wy) \quad (1.6)$$

which represent respectively the function to be minimized in PET and the WNLLS estimate of the parameter vector when the model is linear. Under the previous set of hypothesis on  $e -e_1, e_2, \dots, e_N$  independent from each other, Gaussian distributed with zero mean- the WNLLS estimate of  $p$  coincides with its maximum-likelihood (ML) estimate.

$\sigma^2$ , whose knowledge is necessary for the computation of the covariance matrix of  $p$ , can be estimated a-posteriori evaluating (1.5) in correspondence of its minimum. This value is called WRSS (Weighted Residuals Sum of Squares) and an unbiased estimator of  $\sigma^2$  is given by

$$\frac{WRSS}{N - m} \quad (1.7)$$

As to the uncertainty of estimates, the covariance matrix of the vector  $p_{WLS}$  estimated according to the WNLLS criterion can be approximated as:

$$(S'\Sigma^{-1}S)^{-1} \quad (1.8)$$

where  $S$  is the so-called sensitivity matrix, i.e. the  $N \times m$  matrix whose element in position  $(i, j)$  is the derivative of  $h(p, t)$  with respect to  $t_i$  and  $p_j$  evaluated in correspondence of the final estimate of  $p$ . This expression for the covariance matrix is an underestimation of the real, though unknown, covariance, whereas it becomes an *exact* expression in the case of linear models for which, besides,  $S$  becomes equal to  $X$ .

### 1.2.2 Bayesian estimate

A different way to address the estimation of  $p$  is to exploit not only the measured data  $y$  but also other information that is assumed to be known on  $p$ . This kind of approaches in which an assumption on the probability distribution of  $p$  is made are called Bayesian. More generally, the expression "Bayesian methods" is used when any kind of a-priori information on the parameter vector is assumed known and used for the estimation; in this paragraph, however, the word Bayesian will be used in its narrower meaning, and the probability distribution of  $p$  will be assumed known. This assumption allows

to write the probability density function of the parameter vector given the data (the so-called posterior distribution)

$$f(p|y) \tag{1.9}$$

that, according to Bayes's rule, is proportional to

$$f(y|p)f(p) \tag{1.10}$$

with the proportionality constant independent from  $p$ . The first factor in (1.10), as a consequence of the hypothesis on  $e$ , is normally distributed with mean  $h(p)$  and covariance  $\Sigma$ , whereas the second factor depends on the assumption on the distribution of  $h(p)$ .

The most common way to estimate  $p$  is to select the one which maximizes (1.10). This kind of estimator is called maximum-a-posteriori (MAP) estimator in that it is the one which maximizes the posterior probability density function. A closed-form solution for the MAP estimator does not exist since the structure of equation (1.10) depends on the probability distribution of  $p$ .

Another possible way to exploit (1.10) in the estimation process consists in setting  $p$  equal to the mean of the posterior distribution, i.e. to  $\int_{R^n} f(p|py) p dp$ . This approach, however, is typically too expensive computationally and seldom used; from now on the MAP estimator only will be considered.

Under the common additional assumption that  $p$  is a multivariate Gaussian random vector, with mean  $\mu$  and covariance matrix  $\Delta$ , the vector which maximizes (1.10), i.e. the MAP estimator of  $p$ , is the one which minimizes :

$$[y - h(p)]'\Sigma^{-1}[y - h(p)] + (p - \mu)'\Delta^{-1}(p - \mu) \tag{1.11}$$

where this expression is simply obtained by taking the natural logarithm of (1.10) and omitting the terms independent on  $p$ . In (1.11), the cost function consists of two terms. The first equals the cost function of the WNLLS estimator; the second term weights the adherence of the estimate to the available a-priori knowledge. If the elements of the covariance matrix  $\Delta$  are "small", i.e. the prior is reliable, the elements of  $\Delta^{-1}$  are "large", and the second term in (1.11) becomes the most important: the final estimate of  $p$  will be then close to  $\mu$  as any variation of  $p_{MAP}$  from  $\mu$ , being multiplied by  $\Delta^{-1}$ , would give rise to a high increase of the cost function. If, on the other side, the elements of  $\Delta$  are "large" the final estimate  $p_{MAP}$  will be determined mainly by the first term of (1.11), and the MAP estimator of  $p$  will be consequently close to the WLS estimator of  $p$ .

Being (1.11) a non-linear function of the parameter vector, a closed-form solutions of (1.11) does not exist and iterative algorithms, such as Gauss-Newton's, need to be employed. If, however, the original model is linear  $-h(p) = Xp$ - the MAP estimator of  $p$  becomes:

$$(X'\Sigma^{-1}X + \Delta^{-1})^{-1}(X'\Sigma^{-1}y + \Delta^{-1}\mu) \tag{1.12}$$

The existence of a closed-form solution significantly reduces the time required to estimate  $p$ , and this is particularly important when generating parametric maps since the estimation needs to be performed at each and every pixel and the issue of computational time therefore becomes central. It is important to point put again that the closed-form solution (1.6) for the WNLLS problem exists only if the model is linear and, analogously,

the closed-form solution (1.12) for the MAP estimator holds only if  $p$  is assumed to be Gaussian distributed and the model is linear.

As to the uncertainty of estimates, the MAP estimator is characterized by the additional presence of the covariance matrix of the prior, if compared to the WNLLS estimator. In fact, the approximate expression for  $\Sigma_{MAP}$  is

$$(S'\Sigma^{-1}S + \Delta^{-1})^{-1} \quad (1.13)$$

As before, (1.13) becomes the exact expression when the model is linear. The final estimate of  $p_{MAP}$  has a "smaller" covariance matrix with respect to  $p_{WLS}$  as a consequence of the introduction of the a-priori information which reduces the uncertainty of estimates: it is sufficient to compare (1.8) and (1.13) considering the case in which  $p$  is scalar to realize this simple fact.

Always referring to the case of PET in which  $\Sigma$  is known up to a proportionality factor according to (1.4), the value of  $\sigma^2$  needs to be estimated also in the Bayesian framework because the exact expression of  $\Sigma$  is required for the computation of the final covariance matrix (1.13). To solve this problem it is sufficient to perform a preliminary WNLLS fit and to compute an estimate of  $\sigma^2$  according to (1.7):  $\Sigma$  can be then easily derived and (1.13) used.

### 1.3 A simple example of a Bayesian approach

Assume that one of the  $m$  parameters of the vector  $p$  (the  $j$ -th) is fixed to  $p_j^0$ , thus leaving  $(m-1)$  parameters to be estimated. This fact can be seen in the Bayesian framework in the following way. Let's get back to the basic equation (1.11) and set the  $j$ -th element of the prior  $u$  to the desired value  $p_j^0$  and the others to arbitrary values. Let's set also all the elements of  $\Delta^{-1}$  to 0, except for the element in position  $(j, j)$  which we set to a "very big" value  $M$ : loosely speaking, this is equivalent to assume a null variance for the  $j$ -th element of the prior and an infinite variance for the others. With this choice the Bayesian term of (1.11) -the second term of (1.11)- becomes  $M(p_j - p_j^0)^2$ . In conclusion, while the estimate of all parameters but the  $j$ -th will be free and determined only by the data, the final estimate of  $p_j$  will be necessarily equal to the prior  $p_j^0$  because any difference of  $p_j$  from this value would determine a huge increase in the cost function due to the presence of the multiplicative constant  $M$ .

Having shown that fixing a parameter can be seen as a simple particular case of Bayesian inference, we give here a simple example of the benefits that, in certain situations, can be obtained when one of the kinetic parameters is fixed. To illustrate this aspect the tracer N-methyl-4-piperidin acetate ( $[^{11}\text{C}]$ MP4A) was considered ([3], [4]). This tracer will be described in depth in the following chapter, and, for now, it is sufficient to know that its kinetics is accurately described by a two-compartment three-constant irreversible model with  $K_1$  [ml/ml/min] representing the rate of tracer influx from plasma to tissue,  $k_2$  [ $\text{min}^{-1}$ ] the rate of outflux from tissue to plasma, and  $k_3$  [ $\text{min}^{-1}$ ] the rate of tracer hydrolysis performed by the enzyme acetylcholinesterase (AChE).  $k_3$  is the most important parameter because it allows to differentiate between healthy subjects and subjects suffering from pathologies as Alzheimer's disease.

Indicated with  $C_{REF}$  the time course of the cerebellum, used as reference region, and with  $R$  [unitless] the ratio between the target and reference region  $K_1$  values, it can

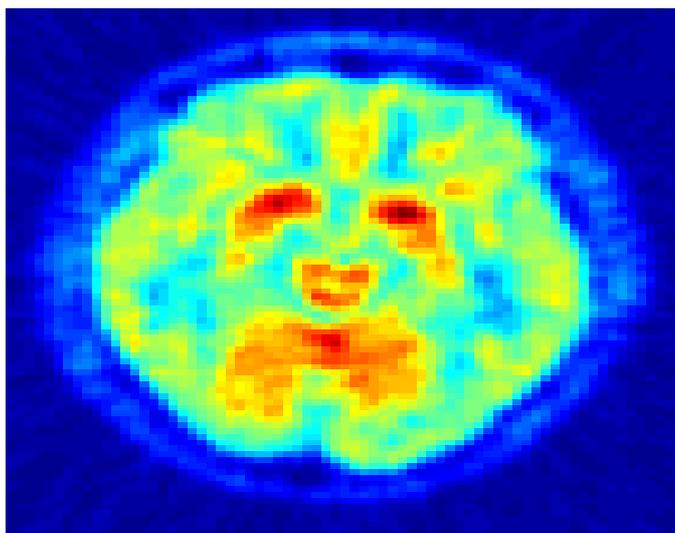


Figure 1.1: *The summed image for an healthy subject for a slice with high AChE activity is displayed.*

be shown ([5]) that the equation

$$C(t) = RC_{REF}(t) + Rk_3 \int_0^t C_{REF}(x)dx - (k_2 + k_3) \int_0^t C(x)dx \quad (1.14)$$

holds, with  $C(t)$  indicating the time course of the region or pixel of interest. Using weighted linear regression, one can easily estimate  $R$ ,  $k_2$  and  $k_3$ .

For this tracer the cerebellum is a reference region with a meaning very different from the usual, in that the uptake of the tracer in the reference, instead of being lower than in the other regions due to the absence of specific binding as it is usually, it is so high that the model in the cerebellum can be approximated as a one compartment irreversible model with only one rate constant  $K_1$ . (1.14) is based on this assumption.

It is known, however, that (1.14) heavily underestimates  $k_3$  in regions with high AChE activity, that is in regions with high  $k_3$ ; this is something which can create problems due to the importance of this parameter for  $[^{11}C]MP4A$ . Figure (1.1) displays the summed image for an healthy subject for a slice with high AChE activity (slice n°16 out of the 35). The expression "summed" image stands for the image in which for each pixel the values of the TACs are summed in order to get an idea of the areas with high and low uptake. Figure (1.2) shows the  $k_3$  parametric image obtained using (1.14): while there should be agreement between high AChE activity -that is high values of the summed image- and high  $k_3$  values, not only this agreement is totally absent but areas with high AChE activity are characterized by low or even negative  $k_3$  estimates.

Let's now fix, for each pixel,  $k_2$  to the median (or mean, the difference is negligible) of the values obtained in the whole slice from the preliminary application of (1.14). (1.14) can be rearranged as follows

$$\begin{aligned} C(t) + k_2 \int_0^t C(x)dx &= RC_{REF}(t) \\ + Rk_3 \int_0^t C_{REF}(x)dx + k_3 \int_0^t C(x)dx & \end{aligned} \quad (1.15)$$

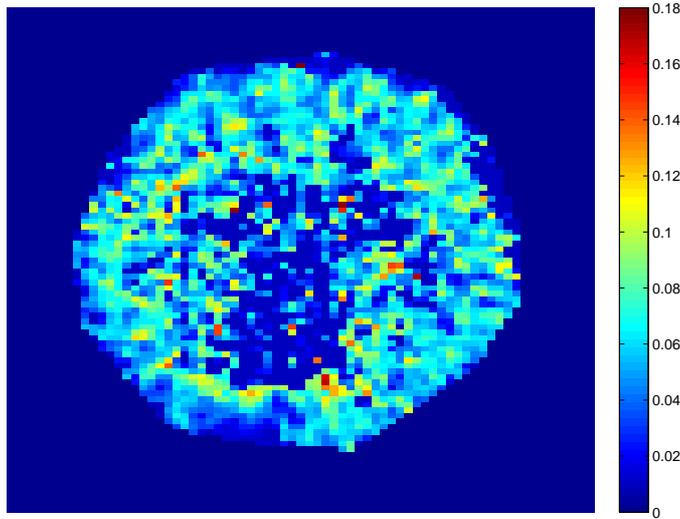


Figure 1.2:  $k_3$  parametric map obtained using (1.14) is displayed.

The left side of (1.15) is now known, as  $k_2$  is fixed. To estimate  $R$  and  $k_3$  from (1.15) one could think to apply a simple linear regression as the model is linear. This is not possible as one would obtain three parameters, whereas there are only two free parameters ( $R$  and  $k_3$ ).

A basis function approach, on the other hand, can be successfully employed. If  $k_3$  is fixed to a value  $k_3^*$ , (1.15) becomes an estimation problem linear in the parameter  $R$ , as expressed by

$$C(t) + (k_3^* + k_2) \int_0^t C(x)dx = R(C_{REF}(t) + k_3^* \int_0^t C_{REF}(x)dx) \quad (1.16)$$

with the left side known and the right side known, apart from  $R$ .

As  $k_3$  is clearly unknown, the estimation must be performed for each value belonging to a predefined grid covering the physiological range for  $k_3$ , and in the end the value of  $R$  which gave rise to the smallest weighted sum of residuals is kept, together with the corresponding  $k_3$ . A grid comprised of 30 values for  $k_3$  was employed here. This technique was applied to every pixel and, due to its linear nature, it allowed to quickly compute new parametric maps for  $k_3$ . Figure (1.3) displays the  $k_3$  parametric image obtained using this approach: it is evident how it is much more resembling the summed image (1.1) than the preliminary parametric map (1.2).

The most natural idea, probably, would have been to fix  $k_2$ , for each pixel, not the global median but to the specific  $k_2$  value obtained for that pixel from the application of (1.14); this approach, however, didn't yield noticeable improvements with respect to RLS. No differences, besides, were seen between the proposed and standard approaches when regions with low AChE activity were considered because (1.14) provided in that case estimates of good quality.

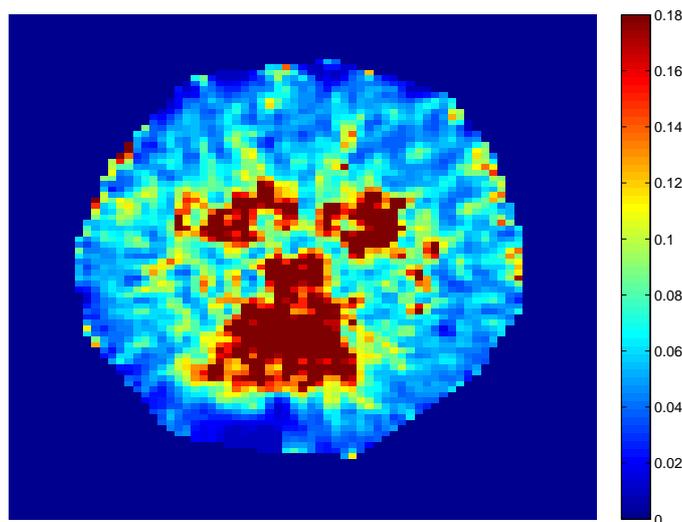


Figure 1.3:  $k_3$  parametric map obtained using (1.15) is displayed.

## 1.4 Discussion

In this chapter, after presenting the basic principles of WLS, the basic idea of Bayesian approaches -employ for the estimation of the parameters of interest some kind of a-priori information- was introduced and mathematically formalized. The use of such approaches may lead to dramatic improvements of final estimates, especially when measured data are characterized by high noise. In the case of PET, therefore, these approaches are more suitable for pixel analysis due to the higher noise of pixel TACs with respect to ROI TACs.

Few Bayesian approaches have been employed in the case of PET in the past. There is only one remarkable example: ridge regression, introduced in [6], and then brilliantly developed in [7], [8] and [9]. Other examples of Bayesian approaches, developed by Kimura and colleagues, are described in [10], [11], [12], but these techniques have a narrower range of application because their use is confined to specific combination of compartmental models and tracers.

In addition it is important to recall here that the methodologies that will be presented in the following chapters -ridge regression and population approaches- are much more general and powerful than the simple strategy of fixing a value for an element of the parameter vector. In fact, these methods address all the parameters -not only one- eliminating the arbitrariness in the choice of the parameter to be fixed, and the weight of the prior is not arbitrarily a-priori specified as here, but suitably computed.

## Chapter 2

# Ridge regression: theory and validation on simulated data

### 2.1 Linear ridge regression

The Ridge Regression with Spatial Constraint method (RRSC) was originally applied to PET in [6] by O'Sullivan and Saha and then developed by Zhou and co-workers in both its linear ([7],[8]) and non-linear ([9]) versions. The main idea behind these methods is to improve parametric images by penalizing both local spatial variation of parameters and noisy pixel kinetics.

In linear ridge regression, which will be considered in this paragraph and in the following one, parameters are estimated for each pixel first using conventional LLS (Linear Least Squares) minimizing the usual function

$$(y - Xp)'W(y - Xp) \quad (2.1)$$

where  $W$  denotes a  $N \times N$  square matrix containing appropriate weights. The parametric maps thus obtained are then smoothed using for each pixel a window of size  $5 \times 5$  or  $3 \times 3$  with equal weights in it, centered in the pixel, so that a prior  $p_0$  of the parameter vector  $p$  is obtained for every pixel; an equivalent way to indicate this smoothing process is to say that the maps are filtered with a median filter. Parameter estimation is then performed again, but this time a Bayesian term derived from the previous preliminary estimation process is added. The function to minimize becomes:

$$(y - Xp)'W(y - Xp) + (p - p_0)'H(p - p_0) \quad (2.2)$$

with  $H$  determining the weight of the Bayesian term.

Two versions of linear ridge regression have been proposed: they are simplified ridge regression with spatial constraints (SRRSC) and generalized ridge regression with spatial constraints (GRRSC). If we let  $m$  be the number of elements of the parameter vector, the diagonal ridge matrix  $H$  is calculated for SRRSC as

$$h_i = \frac{m\sigma^2}{(p - p_0)'(p - p_0)} \quad \forall i \quad (2.3)$$

and for GRRSC as

$$h_i = \frac{\sigma^2}{(p_i - p_{i0})^2} \quad i = 1, 2, \dots, m \quad (2.4)$$

where the noise variance of the data  $\sigma^2$  is estimated for each pixel from the residuals of the preliminary weighted regression as in (1.7). Before the application of (2.2) the elements of  $H$  are spatially smoothed with the same median filter employed to compute  $p_0$ . This second smoothing is not strictly necessary for the application of the algorithm, but final parametric maps become more regular when the median filter is applied also to the elements of  $H$ .

The difference between the two methods is that GRRSC makes use of different weights for each parameter while SRRSC does not. The ridge matrix  $H$  is therefore a generic diagonal matrix for GRSSC while it is a multiple of the identity matrix for SRRSC.

According to the previous equations, the higher the estimated variance is, the higher the values of  $H$  are, and the more important the Bayesian term which appears in (2.18) is in the computation of  $p$ ; this is totally coherent with the idea that the higher the variance for a given pixel is, the greater the importance that should be given to the prior is, as measured data are not very reliable in the computation of  $p$  for the specific pixel due to the initial poor fit.

Similar considerations hold for the denominator of (2.3) and (2.4). Small local spatial variations of the estimates of  $p$ , i.e. a small denominator in (2.3) and (2.4), give rise to high values of the elements of  $H$ ; the Bayesian term acquires in this case more weight consistently with the idea that small local spatial variations of the estimates are an indicator of reliability of the prior.

The assumption on the linearity of the original problem is not essential. In fact, "linear" ridge regression can be applied to non-linear models as well. The only issue that limits the applicability of this procedure for non-linear problems is computational, since a non-linear fit is to be performed twice at each pixel to solve first (2.1) and then (2.2), with  $h(p)$  instead of  $Xp$ . Computational time, which is already high for non-linear procedures, is therefore doubled making the procedure unfeasible. In the linear case, on the other side, the total computational effort is negligible because closed form solutions for (2.1) and (2.2) do exist. They are respectively

$$p = (X^T W X)^{-1} (X^T W Y) \quad (2.5)$$

identical to (1.6), and

$$p_{ridge} = (X^T W X + H)^{-1} (X^T W Y + H p_0) \quad (2.6)$$

The last equation has a structure very similar to (1.12) which gives the MAP estimate in the linear case when the prior is assumed normally distributed. The only differences is that  $W$  in (2.6) is not the inverse of the covariance matrix of the error  $\Sigma$  but just proportional to it, and that  $H$ , to correctly balance the weight of data and prior, is computed to keep this fact into account. Apart from this difference in the structure of the covariance matrices, the identical structure of relation between (1.12) and (2.6) allows to interpret the parameter estimate of linear ridge regression as a MAP estimator with a normal prior.

## 2.2 Evaluation of linear ridge regression

### 2.2.1 Indices for performance evaluation

To be able to quantitatively compare two or more different estimation methods the "correct" values of the parameters to be estimated should be known in advance so that indices which evaluate their performances can be computed. When real data sets are used this is not possible since the "correct" values of the parameters of interest are unknown. It is often necessary, therefore, to generate simulated data which resemble as much as possible real data, with the only difference that the correct parameter values underlying the simulated data are known.

Let  $p_0$  be the true value of a parameter of interest and assume that this parameter is estimated from simulated data  $N$  times, with different noise conditions. Suppose that the probability distribution of the noise added to the simulated data is the same among the  $N$  repetitions, but that the actual values change between repetitions, leading to different final estimates for  $p$ .

The two most important indices for the evaluation of performances are percentage BIAS%

$$BIAS\% = 100 \frac{1}{p_0} \left\{ \frac{\sum_{i=1}^N (p_i - p_0)}{N} \right\} \quad (2.7)$$

which keeps into account the possible systematic over(under)estimation of the given parameter (the lower its absolute value is the better the estimation method is) and Root Mean Square Error RMSE%

$$RMSE\% = 100 \frac{1}{p_0} \left\{ \frac{1}{N} \sqrt{(\sum_{i=1}^N (p_i - p_0)^2)} \right\} \quad (2.8)$$

which quantifies the average difference between the estimate  $p_i$  and the correct value  $p_0$ , irrespectively of its sign. Clearly the lower the RMSE% is, the better the given algorithm is.

Sometimes BIAS% and RMSE% are considered not in percentage form, i.e.  $p_0$  does not appear at the denominator of either (2.7) or (2.8). The percentage form is however more meaningful: a bias of  $0.1 \text{ min}^{-1}$  for a given kinetic parameter  $k_0$  for instance, has clearly a different impact if the real value of  $k_0$  is  $0.2 \text{ min}^{-1}$  or if it is  $1 \text{ min}^{-1}$ . From now on the percentage forms only will be used and referred to simply as BIAS and RMSE; for the sake of space the percentage sign will be omitted.

Another important index is the Coefficient of Variation (CV) which quantifies the reliability of a parameter independently on its correctness (i.e. independently on  $p_0$ ). Let  $\sigma^2$  be the variance associated to a estimated parameter  $\hat{p}_i$  computed from the covariance matrix of the parameter vector. This variance is, in general, not only data-dependent but also model-dependent, and the CV for the given parameter is defined as the ratio between the standard deviation and the parameter estimate

$$100 \frac{\sigma}{\hat{p}_i} \quad (2.9)$$

### 2.2.2 The tracer [ $^{11}\text{C}$ ]MP4A

To test linear ridge regression N-methyl-4-piperidin acetate ([ $^{11}\text{C}$ ]MP4A) was employed. [ $^{11}\text{C}$ ]MP4A is a tracer that is used with PET for diagnosis and assessment of progression

of Alzheimer's disease and other brain pathologies such as dementia with Lewy bodies. As specified in the previous chapter, the different steps of  $[^{11}\text{C}]\text{MP4A}$  kinetics can be adequately quantified by using a two-compartment three-rate constant model ([3]), with  $K_1[\text{ml/ml/min}]$  representing the rate of tracer influx into the brain,  $k_2[1/\text{min}]$  the rate of outflux from the brain, and  $k_3[1/\text{min}]$  the rate of tracer hydrolysis performed by the enzyme acetylcholinesterase (AChE). As for most tracers, the gold-standard method to estimate kinetic parameters of  $[^{11}\text{C}]\text{MP4A}$  is WNLLS, with the plasmatic tracer concentration as input function to the model.

It can be shown, see for instance [13], that , the concentration  $C(t)$  in a given ROI or pixel of a tracer like  $[^{11}\text{C}]\text{MP4A}$  described by an irreversible two-compartment three-constant model (2T-3K), assuming the blood component  $V_b$  equal to 0, can be expressed as:

$$C(t) = \frac{K_1 k_2}{k_2 + k_3} \int_0^t C_{pl}(x) e^{-(k_2+k_3)(t-x)} dx + \frac{K_1 k_3}{k_2 + k_3} \int_0^t C_{pl}(x) dx \quad (2.10)$$

where  $C_{pl}$  denotes the measured plasmatic concentration of the tracer. An alternative expression of (3.15) proposed by Blomquist ([14]), which is based on the integration of the differential equations leading to (3.15), is

$$C(t) = K_1 \int_0^t C_{pl}(x) dx + K_1 k_3 \int_0^t \int_0^\phi C_{pl}(x) dx d\phi - (k_2 + k_3) \int_0^t C(x) dx \quad (2.11)$$

which allows to estimate kinetic parameters by means of weighted *linear* least squares.

The use of the plasmatic tracer concentration, however, requires measurements of the tracer concentration and of its metabolites. These measurements are invasive to patients and require also considerable technical expertise. The quest for alternative but equally reliable methods for parameter estimation has led to the development of a reference-based method. Reference Least Squares (RLS) is a method developed for  $[^{11}\text{C}]\text{MP4A}$  and based on (2.11) to avoid the use of  $C_{pl}$  in (2.11).  $[^{11}\text{C}]\text{MP4A}$  molecules, after entering the cerebellum, are rapidly transformed into the hydrophilic metabolite  $[^{11}\text{C}]\text{MP4OH}$  because AChE activity is very high: this results in a high  $k_3$  in that region. If  $k_3$  is high enough, the 2T-3K model can be adequately approximated by a one-compartment one-rate constant irreversible model described by

$$\frac{C_{REF}(t)}{dt} = K_{1,REF} C_{pl}(t) \quad (2.12)$$

where  $C_{REF}$  denotes cerebellar activity as, for this tracer, cerebellum is named reference region, and  $K_{1,REF}$  is the value of  $K_1$  in this region.

In addition, cerebellar TAC  $C_{REF}$  shows a rapid increase after the tracer injection and a plateau level which begins when plasma  $[^{11}\text{C}]\text{MP4A}$  radioactivity is almost zero, suggesting that  $C_{REF}$  could be used to approximate the time integral of the arterial input function according to:

$$C_{REF}(t) = K_{1,REF} \int_0^t C_{pl}(x) dx \quad (2.13)$$

which is nothing but the integrated form of (2.12) and which can be easily written as

$$\int_0^t C_{pl}(x) dx = \frac{C_{REF}(t)}{K_{1,REF}} \quad (2.14)$$

Sometimes the striatum is used as reference region for  $[^{11}\text{C}]\text{MP4A}$ . Cerebellum is in general preferred for its higher volume, which results in smoother TAC, and for its typically slightly higher  $k_3$ .

Inserting (2.14) into (2.11) one gets:

$$C(t) = p_1 C_{REF}(t) + p_2 \int_0^t C_{REF}(x) dx + p_3 \int_0^t C(x) dx \quad (2.15)$$

with  $p_1 = K_1 / K_{1,REF} = R$ ,  $p_2 = R k_3$  and  $p_3 = -(k_2 + k_3)$ . Using weighted linear regression one can estimate  $p_1$ ,  $p_2$  and  $p_3$  without the need for blood sampling and then evaluate  $R$ ,  $k_2$ ,  $k_3$  from  $R = p_1$ ,  $k_3 = p_2 / p_1$ ,  $k_2 = p_3 - k_3$ .

### 2.2.3 Generation of simulated $[^{11}\text{C}]\text{MP4A}$ data

In order to apply the principles of both ridge regression and population methodologies it was necessary to generate simulated slices of a PET scan. It is very important that the simulated data reproduce accurately real data in terms of amount of noise of pixel kinetics and in terms of the spatial distribution of kinetic parameters, because, otherwise, results are likely to be misleading. Several procedures have been used in literature to create synthetic data set, some of them making use of ad-hoc simulation softwares. An approach robust in terms of both criteria just cited -appropriate noise and appropriate parameter distribution- is described here.

Simulated  $[^{11}\text{C}]\text{MP4A}$  dynamic images were created to reproduce the time activity curves of an healthy subject at pixel level. Parametric maps of  $K_1$ ,  $k_2$ ,  $k_3$  were first computed by quantifying the real  $[^{11}\text{C}]\text{MP4A}$  image of one normal subject using his plasmatic input function and the gold standard WNLLS with weights equal to

$$\frac{\Delta_i}{C_i} \quad (2.16)$$

with  $\Delta_i$  indicating the duration of the  $i$ -th frame and  $C_i$  the measured decay-corrected concentration at frame  $i$  for the pixel under examination. These are the standard weights employed when the plasmatic input function is used in the process of parameter estimation. These maps were then slightly smoothed by a 2D-Gaussian filter to get more regular parametric images: the resulting maps were regarded as the "real" parametric maps to which compare results to.

Two whole simulated slices were generated: slice n°27 out of the 35 of the image, in the upper part of the brain, was chosen as representative "average" slice containing a large amount of cerebral cortex, whereas slice n°20 was selected because of the presence of regions with high AChE activity (i.e. thalamus). Error-free continuous time activity curves were created for each pixel using these maps and the measured plasmatic input function according to the standard equation (3.15). These curves were then sampled at the midtimes of the frames of the real scanning protocol used for the  $[^{11}\text{C}]\text{MP4A}$  dataset: the scan durations were 6x0.5 min, 2x1 min, 2x2.5 min, 10x5 min, for a total of 20 frames and 60 minutes.

For each pixel, 100 noisy TACs were then generated adding, to the error-free TACs, Gaussian noise with zero mean and variance equal to

$$\alpha \frac{C_i}{\Delta_i} \quad (2.17)$$

with  $C_i$  and  $\Delta_i$  as before. The variance was clearly set proportional to the inverse of the weights used in the process of parameter estimation. To choose reliable values for  $\alpha$  - constant for each pixel but varying from pixel to pixel due to differences in the spatial position-  $\alpha$ -maps were generated during the quantification of the real [ $^{11}\text{C}$ ]MP4A image by computing for each pixel  $\alpha = \text{WRSS}/(N-m)$  as in (1.7), where WRSS is the value of the weighted cost function evaluated at its minimum,  $N$  the number of frames and  $m$  the number of model parameters (20 and 3, respectively, in our case). These maps were then used in the process of noise generation: in this way simulated activity curves resembled real ones not only as to the amount of noise but also as to the differences in noise due to pixel position.

A critic to the procedure employed could be made: why add to the noise-free TACs *independent* (among pixels) random samples, if it is known that the reconstruction algorithms create a spatial correlation between noise of adjacent pixels? The fact is that no simple but effective method to model noise correlation among pixels is available. Even though the statistics of PET noise is well known at sinograms level -i.e. at the level of raw data before the application of reconstruction algorithm- no mathematical model of noise spatial correlation is available which starts from noiseless pixel TACs.

## 2.2.4 Results

Results are presented graphically in figure (2.1), which displays the average BIAS and figure (2.2), which shows the average RMSE for both slices. Emphasis was put on  $k_3$ , which, as explained, is the most important parameter for [ $^{11}\text{C}$ ]MP4A as it allows to differentiate between healthy and ill subjects.

GRRSC always outperformed SRRSC, as expected, as it is a more flexible methodology in that it makes use of different weights for the different parameters; the amount of improvement varied from slight to moderate.

The second and more important fact was that GRRSC provided noticeable improvements with respect to standard RLS both in terms of BIAS and RMSE reduction. The increase of computational cost, as explained previously, was negligible. With the performances of the computer employed in the simulations it took roughly 15 seconds to generate results for a single slice with RLS, and therefore roughly 30 seconds with GRRSC or with SRRSC (15 of which for the preliminary RLS estimation). The slight BIAS of slice n°27, for instance, was completely annulled by GRRSC and the heavier negative BIAS of slice n°20 was halved. Results presented in [15], besides, showed that among all the several methodologies available for the quantification of [ $^{11}\text{C}$ ]MP4A without arterial blood sampling, the application of GRRSC to RLS provides an optimal trade-off between performances and speed. Figure (2.3) displays the parametric map of the parameter  $k_3$  of one repetition of the simulated slice n°27 for RLS, SRRSC and GRRSC. It is evident how GRRSC provided much smoother and regular parametric maps with respect to RLS and, to a minor extent, also when compared to SRRSC. This holds true also for the other kinetic parameters: figure (2.4), for instance, compares parametric maps of the relative delivery  $R$  obtained with RLS and GRRSC for slice n°27.

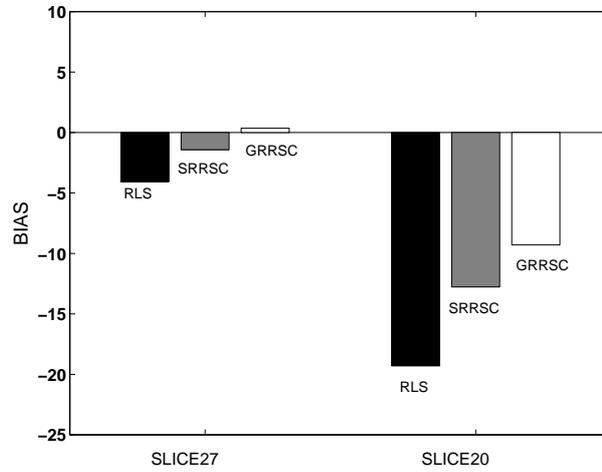


Figure 2.1:  $k_3$  BIAS averaged over all pixels, obtained with RLS, SRRSC and GRRSC is displayed for slice  $n^\circ 27$ (left) and  $n^\circ 20$ (right).

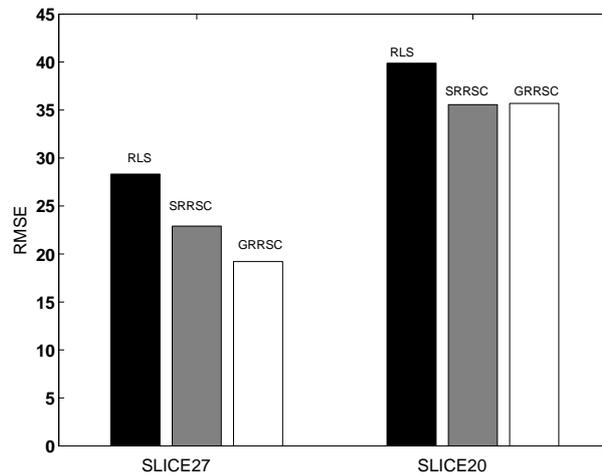


Figure 2.2:  $k_3$  RMSE averaged over all pixels, obtained with RLS, SRRSC and GRRSC is displayed for slice  $n^\circ 27$ (left) and  $n^\circ 20$ (right).

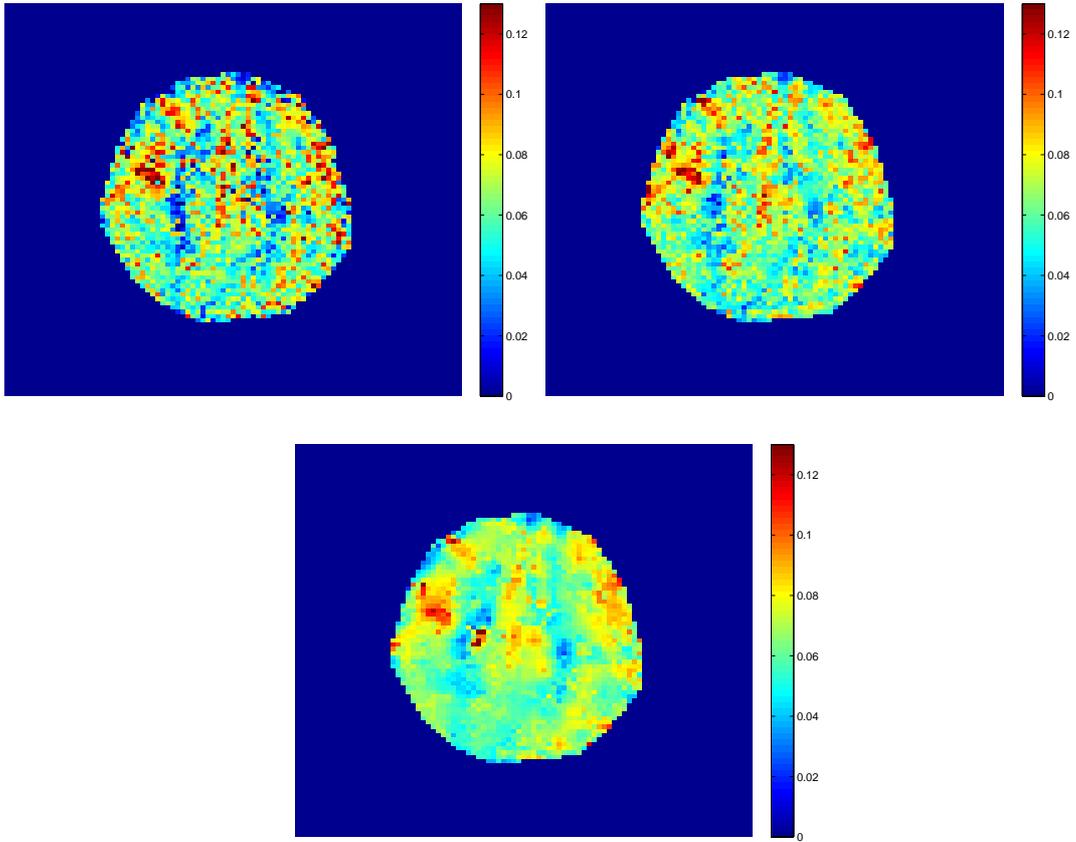


Figure 2.3: *The parametric map of  $k_3$  of one repetition of the simulated slice n°27 obtained with RLS(left), SRRSC(left) and GRRSC(down) is displayed.*

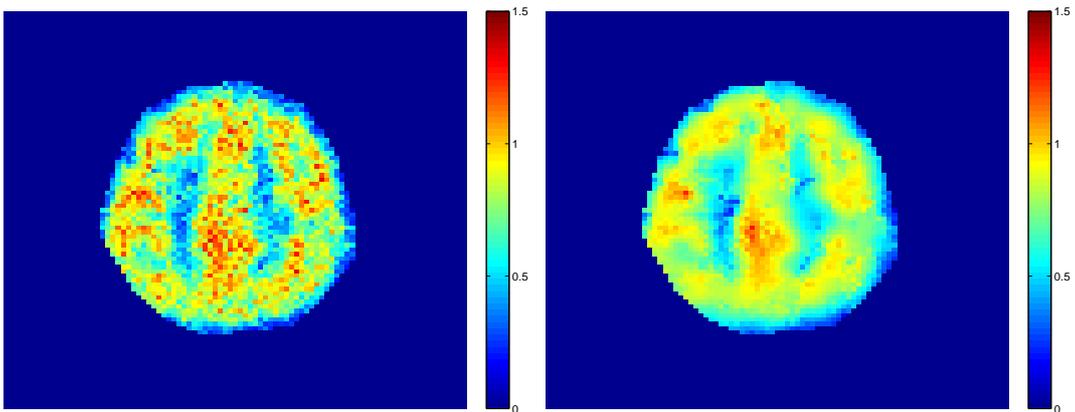


Figure 2.4: *The parametric map of the relative delivery  $R$  of one repetition of the simulated slice n°27 obtained with RLS(left) and GRRSC(right) is displayed.*

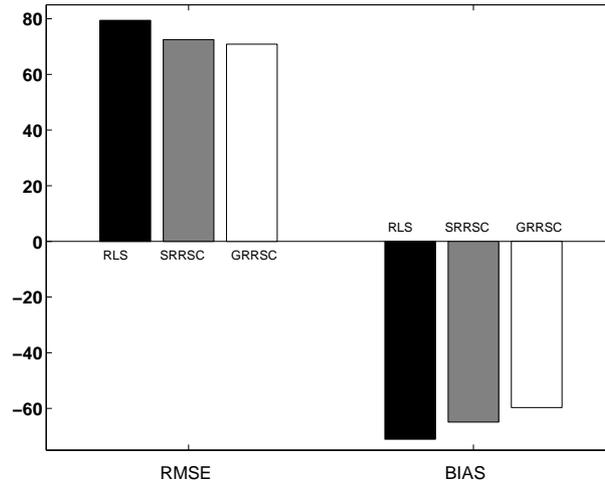


Figure 2.5: Average RMSE(left) and BIAS(right) of  $k_3$  obtained with RLS, SRRSC and GRRSC for the portion of slice n°20 containing the thalamus are displayed.

It is well known (see for instance [5]) that RLS performs poorly when the parameter  $k_3$  of the target region is high and that in this case  $k_3$  estimates are characterized by a negative bias. The thalamus is a region characterized by a high AChE activity and therefore by high values of  $k_3$ : RLS estimates of this parameter in that region were consequently affected by a systematic negative BIAS. This explains the global negative BIAS of slice n°20, which is due to the presence of the thalamic region in it. Figure (2.5) reports, as previously, BIAS and RMSE but only for the portion of slice n°20 containing the thalamus (mask drawn manually). It can be seen that, besides the obvious worsening of performances, in this case the amount of improvement is very limited due to poor initial estimates.

The conclusion is that linear ridge regression techniques are powerful instruments to improve the reliability of parametric maps. Generalized ridge regression is preferable to simplified ridge regression due to its flexibility in weighting differently the contribution of the Bayesian term for each parameter. Large BIAS and RMSE reductions were observed and resulting parametric maps appeared smoother at visual impact. The amount of improvement ranged from moderate to significant according to the parameter under examination. When initial estimated were very poor however -as for the case of thalamic  $k_3$ - the amount of improvement brought by these methods was low because of the intrinsic limits of the linear method from which ridge regression techniques started. Although these conclusions were derived only on the basis of only one simulated data set it is likely that they are tenable for most tracers, also because they are coherent with what found in ([7] and in [8]). The amount of improvement, on the other side, will be clearly dependent on the complexity of the considered model and on the noise level of the data analyzed.

### 2.3 Non-linear ridge regression

In this paragraph non linear problems, in which measured data are a generic non linear function  $h(p)$  of the parameters of interest, will be considered. It is known that any

non linear problem can be suitably rewritten in a linear form  $h(p)=Xp$ , but that the estimation of the linearized version gives rise to bias in the parameter estimates, due to the correlation between the measurement noise and the noise of the columns in the matrix X containing terms dependent on the pixel TAC. In [17] Feng and colleagues proposed a method, called GLLS (Generalized Linear Least Squares), which can be used to correct for the bias introduced by the linearization process. One could think to apply the linearization and GLLS to any non linear problem, employing then linear ridge regression. GLLS, however, hasn't been validated on complex models, such as the two-compartment 4-constant model, and hasn't found a wide diffusion in the PET world. The need to consider non linear models, therefore, appears necessary.

As pointed out previously, linear ridge regression can be employed for non-linear models as well. Instead of minimizing a function like (2.2), the function to consider assumes now the form:

$$[y - h(p)]'W[y - h(p)] + (p - p_0)'H(p - p_0) \quad (2.18)$$

For each pixel, therefore, a non-linear estimation has to be performed twice, the first time using measured data only to compute  $p_0$  and H, the second to find the point of minimum of (2.18). This procedure is not feasible at pixel level, especially when generating 3D parametric maps, because it is too time consuming.

### 2.3.1 Prior based on WNLLS: NLRR1

If in (2.18), however,  $h(p)$  is substituted by its first order Taylor expansion around the point  $p^*$ , with  $p^*$  equal to the preliminary estimate at the pixel of interest, preliminary computed through WNLLS, the second estimation problem becomes linear and can be efficiently solved. This version of Non Linear Ridge Regression was called NLRR1.

The most reasonable way to address the problem is to approximate the non-linear function  $h(p, t_i)$ ,  $i = 1, 2, \dots, N$  through

$$h(p^*) + S\Delta p \quad (2.19)$$

with  $p^*$  equal to the WNLLS estimate of the parameter vector made using the pixel data only,  $\Delta p$  equal to the  $m \times 1$  column vector  $(p - p^*)$ , and with S, the  $N \times m$  sensitivity matrix, containing at each row the gradient of  $h(p, t_i)$  evaluated at  $p^*$ . H can then be computed exactly as for ridge regression -either in its Simplified, or, preferably, Generalized form- and in the same way the prior can be estimated, smoothing the initial WNLLS parametric maps through a filter of size  $3 \times 3$  centered in the pixel of interest, with equal weights for the pixels in it. This same filter was used for all the versions of NLRR (see the following paragraphs), as well as it was used for linear ridge regression.

Substituting to  $h(p)$  the approximation expressed by (2.19), and  $(\Delta p + p^*)$  to p into (2.18), one gets

$$Q(\Delta p) = (y - h(p^*) - S\Delta p)'W(y - h(p^*) - S\Delta p) + \quad (2.20)$$

$$(\Delta p - (p_0 - p^*))'H(\Delta p - (p_0 - p^*))$$

This equation is linear in the unknown parameter "increment"  $\Delta p$ . Its solution is given by

$$\theta = (S^TWS + H)^{-1}[S^TW(y - h(p^*)) + H(p_0 - p^*)] \quad (2.21)$$

The previous equation is the usual expression for the point of minimum of linear function with a linear Bayesian term as in (2.6), the only difference being the fact that, here, the "data" are represented by the difference ( $y - h(p^*)$ ) between the measured and fitted data and that the unknown variable is the increment  $\Delta p$  rather than the parameter vector itself.

Besides testing the simple Gauss-Newton approach proposed, in which the increment  $\Delta p$  is given by (2.21) and the final estimate is therefore given by  $p^* + \Delta p$ , a modified version of the algorithm -the same proposed in [9] and originally developed in [16]- was also employed. An additional scalar parameter  $\rho$  was computed as:

$$\frac{0.75Q(0) - Q(0.5\theta) + 0.25Q(\theta)}{Q(0) - 2Q(0.5\theta) + Q(\theta)} \quad (2.22)$$

and the final estimate  $p_{final}$  was then computed, for each pixel, as  $p^* + \Delta p$ .  $\Delta p$ , this time, is computed as  $\rho \theta$ : the "direction" of the increment does not change, but the amount varies as it is modulated by  $\rho$ .

Whatever method is used for the computation of the increment  $\Delta p$ ,  $p_{final}$  will be a sum of the solution of the standard WNLLS problem  $p^*$  and of a term  $\Delta p$  depending on the prior, determined averaging the preliminary estimates obtained with WNLLS in the neighbourhood of the pixel of interest.

### 2.3.2 Priors based on clusters: NLRR2

In the previous paragraph the assumption that one and only one non-linear problem can be solved at each pixel, due to computational reasons, was implicitly made. In NLRR1 the non linear function  $h(p)$  was minimized to determine  $p^*$  from which the prior was computed, and then the problem was linearized to avoid the second non linear minimization. Why not compute the prior differently so that the only non linear minimization assumed to be available can be used in the end, in the formulation (2.18) in which the Bayesian term is comprised? This version of ridge regression was called NLRR2.

To compute appropriate priors, cluster analysis can be an elegant and efficient solution. Clustering means subdividing the voxels of a given slice or region into homogeneous groups; typically, but not necessarily, voxels are divided on the basis of their TACs. Clustering algorithms can be partitioned between "hard" and "soft": "hard" clustering algorithms ([18]) assign each pixel to one and only one cluster, while "soft" algorithms ([19]) compute for each pixel a set of N positive weights which are proportional to the probability that the pixel belong to each of the N clusters. Both strategies are feasible for the computation of the prior.

Each cluster must include an adequate number of pixels so that each mean kinetics  $C_i(t)$ ,  $i=1,2,\dots,N$ , has a high signal-to-noise ratio for WNLLS model fitting. In fact, the step after clusterization is to fit the model appropriate for the tracer under examination to the kinetics of every cluster means: a parameter vector  $p_i$  is thus obtained for each cluster.

If a "hard" algorithm was used we associate then to each pixel the parameter vector of the cluster to which the pixel was assigned to: this criterion to assign priors will be called "direct attribution" technique. If a "soft" method was used a linear combination of the parameter vectors of the clusters, with weights derived from the clusterization

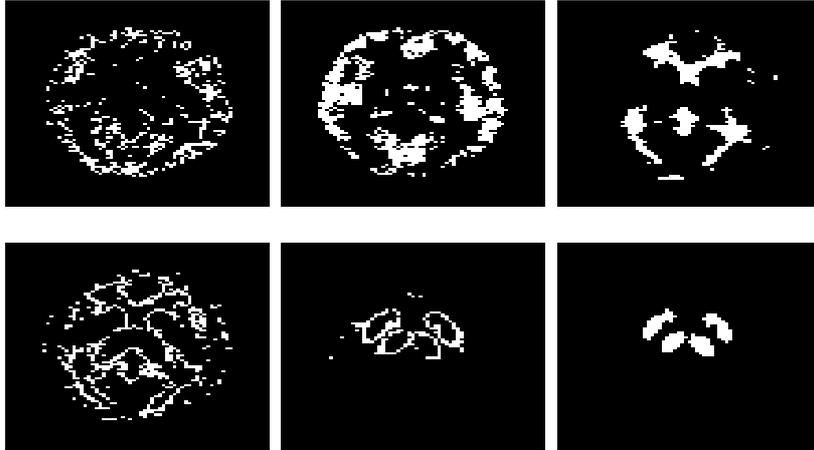


Figure 2.6: *Example of how a dynamic MP4A image was segmented employing hierarchical cluster analysis with 6 clusters*

algorithm, can be instead reasonably employed. Being more reliable, the "hard" hierarchical cluster analysis ([18]) was employed to segment the image both for the direct attribution technique and for CRM (see below in this paragraph); the same technique was also used for the algorithm described in the following paragraph.

The last step is to smooth the resulting parametric image to compute for each pixel an estimate  $p_0^j$  of the parameter vector that is to be used both as prior and as starting point for the minimization of (2.18).

Figure (2.6) shows an example of how a dynamic MP4A image was segmented by hierarchical cluster analysis with 6 clusters. Notice how symmetrical the clusters are and how well the striatum was segmented.

$\sigma^2$ , which is necessary for the construction of H in ridge regression, can be estimated considering the WRSS between the TAC of the j-th pixel and the fitted kinetics  $C_j$  computed using the prior  $p_0^j$  at the pixel. With this technique pixels with noisy TACs, which will therefore give rise to high values of  $\sigma^2$ , will consistently have priors with high strength and viceversa.

Another possible way to estimate priors from clusters was proposed in [9]. From  $p_i$  one can compute the fitted kinetics  $C_i^*(t)$  of each cluster. A set of N positive weights  $\pi_i$  is then to be calculated, at each pixel, from

$$C_j(t) = \sum_{i=1}^N \pi_i C_i^*(t) \quad (2.23)$$

$j=1,2,\dots,M$ , where  $C_j(t)$  indicates the noisy kinetics of the j-th pixel. The problem can be efficiently solved using the Non-Negative Linear Least Squares algorithm (NNLLS). A preliminary prior is then computed simply as

$$\frac{\sum_{i=1}^N \pi_i p_i}{\sum_{i=1}^N \pi_i}$$

and the resulting parametric image is spatially smoothed to yield for each pixel an estimate  $p_0^j$  of the parameter vector that is used as prior and also as starting point

for the algorithm.  $\sigma^2$  can be computed as above. This formulation was called in [9] Component Representation Model (CRM).

Basically, whereas with the direct attribution technique the parameter vector of the cluster is assigned directly to the pixels belonging to that cluster, in the CRM the parameter vector of a generic pixel is computed as weighted sum of the parameters of the clusters, with weight depending on the similarity between the pixel TAC and the clusters TACs.

Both approaches for the attribution of the prior to each pixel were tested. To test how heavily results depended on the number  $N$  of clusters employed, the algorithm was tested with  $N=3,6,9$  respectively.

### 2.3.3 Elimination of non linear minimizations: NLRR3

In [9] a different version of NonLinear Ridge Regression was developed to totally avoid non linear minimizations (NLRR3). The prior is computed from cluster analysis as in the previous paragraph and the usual minimization problem expressed by (2.18) has then to be solved. The second non-linear estimation, which yields the final parameter estimates, is avoided as the original non linear problem is linearized around the current estimate. The significant difference with respect to the linear case, NLRR1, NLRR2 is that the ridge regression matrix  $H$ , in NLRR3, is not computed and kept fixed but it is calculated iteratively. At each iteration  $H$  is computed from the residuals of the linearized problem; then the parameter vector is estimated using this time the Bayesian component as well. The process -linearization + computation of  $H$  + Bayesian solution- is then iterated until a convergence criterion is satisfied.

At iteration  $(j + 1)$ ,  $j = 0, 1, 2, \dots$ , the non-linear function  $h(p, t_i)$ ,  $i = 1, 2, \dots, N$  is approximated through

$$h(p^j) + S^j \Delta p \quad (2.24)$$

as before, where  $p^j$  is the estimate of the parameter vector at the previous iteration,  $\Delta p$  is equal to  $(p - p^j)$ , and  $S^j$  is the sensitivity matrix evaluated at  $p^j$ . At each fixed  $t_i$ , (2.24) represents the first order Taylor expansion of the scalar function  $h(p, t_i)$  around the point  $p^j$ .

Substituting to  $h(p)$  this approximation and to  $p$  ( $\Delta p + p^j$ ) into (2.18) one gets

$$\begin{aligned} Q(\Delta p) &= [y - h(p^j) - S^j \Delta p]' W [y - h(p^j) - S^j \Delta p] \\ &+ (\Delta p - (p_0 - p^j))' H (\Delta p - (p_0 - p^j)) \end{aligned} \quad (2.25)$$

which is totally equivalent to (2.20). This equation is linear in the unknown parameter "increment"  $\Delta p$ , with the remaining issue that  $H$  is unknown. An estimate of the residuals at the current iteration, which is essential for the estimation of  $H$ , can be computed considering only the first addend  $\Phi$  of (2.25), (the first line of (2.25)). Being  $\Phi$  a linear function of  $\Delta p$ , the expression for point of minimum of  $\Phi$  exists in closed-form and it is given by

$$(S^{j'} W S^j)^{-1} [S^{j'} W (y - h(p^j))] \quad (2.26)$$

The computation of  $\Delta p$  allows to estimate the WRSS as

$$[S^j \Delta p - (y - h(p^j))] W [S^j \Delta p - (y - h(p^j))] \quad (2.27)$$

which can be used to compute  $\sigma^2$  from  $WRRS/(N-m)$ , as in (1.7).

Being the prior  $p_0$  already known from cluster analysis and being an estimation of the variance  $\sigma^2$  now available, (2.3) and (2.4) -depending on the choice to use SRRSC or GRRSC- allow to compute the ridge regression matrix  $H$  for each pixel. The minimization of (2.25) becomes now a linear problem in the parameter increment  $\Delta p$  with everything else known. Analogously to (2.21), its solution is given by

$$\theta = (S^{j'} W S^j + H)^{-1} [S^{j'} W (y - h(p^j)) + H(p_0 - p^j)] \quad (2.28)$$

so that the new estimate  $p^{j+1}$  is set for each pixel to  $p^j + \theta$ .

As in NLRR1, the modified version Gauss-Newton approach proposed in [16] was also tested. The additional parameter  $\rho$  was computed as before as:

$$\frac{0.75Q(0) - Q(0.5\theta) + 0.25Q(\theta)}{Q(0) - 2Q(0.5\theta) + Q(\theta)} \quad (2.29)$$

and the final estimate  $p^{j+1}$  set, for each pixel, to  $p^j + \Delta p$ , with  $\Delta p$  given by  $\rho \theta$ .

Following [9] a global relative convergence criterion was used. Iterations were considered to have converged if the total pixelwise cost function, summing over all pixels, changed less than a pre-specified threshold  $\varepsilon$  (e.g. 1%). In other words, the algorithms ended when

$$\left| \frac{\Sigma Q(\Delta p) - \Sigma(Q(0))}{\Sigma Q(0)} \right| < \varepsilon \quad (2.30)$$

where  $Q(\Delta p)$  is the same of (2.25) and the sum is performed over all pixels.

Even though the mathematical details may appear heavy, the rationale underlying the proposed approach is simple. The prior is computed fitting the non linear model of interest to the kinetics of a handful of clusters, exactly as in NLRR2. At the generic iteration  $j$ , then, for each pixel, the non linear function is linearized with a first order Taylor expansion around the current value of the parameter vector and the corresponding linear problem is solved computing the optimum increment  $\Delta p$ . The associated residuals and the values of the parameters vectors at the current iterations allow a computation of the ridge matrix at each pixel as in linear ridge regression. A second and definitive estimation of  $\Delta p$  is then performed including the Bayesian term. The process is then iterated and convergence is declared when the overall variation of the cost functions is sufficiently low.

NLRR2 and NLRR3 compute the priors from clusters in the same way -either with direct attribution of the prior of the appropriate cluster to the pixel or via CRM- and keep it fixed during the rest of the algorithm. However, while NLRR2 simply uses the prior to create a non-linear function in a standard Bayesian framework, NLRR3, through an iterative linearization of the non-linear problem at each pixel around the current estimate, updates the ridge matrix  $H$  at every iteration from the values of residuals and parameters vectors at the current iteration. In NLRR2, priors derived from clusters are also used to compute  $\sigma^2$  and consequently  $H$ , which is then kept fixed. Aside from performance assessment, which will be discussed in the following paragraph, an obvious advantage of NLRR3 over NLLRR2 (and over NLLRR1) is computational as no solution of any non linear problem is required. This brings about a significant reduction of computational time.

## 2.4 Evaluation of non-linear ridge regression

### 2.4.1 The tracer [<sup>11</sup>C]WAY100,635

To test Non Linear Ridge Regression algorithms, synthetic data of the tracer [<sup>11</sup>C]WAY100,635 were generated. This tracer was chosen both because it is widely used in PET and particularly because its 2T-4K model, with two reversible compartments and four rate constants, is by far the most diffused for neuroreceptorial studies. WAY100,635 (N-(2-(4-(2-methoxyphenyl)-1-piperazinyl) ethyl)-N-(2-pyridinyl) cyclohexane carboxamide) is a serotonin 1A antagonist with high affinity and selectivity for serotonin 1A receptors as described in [20].

The differential equations describing the model are:

$$\frac{C_{f+ns}(t)}{dt} = K_1 C_{pl}(t) - (k_2 + k_3) C_{f+ns}(t) + k_4 C_s(t) \quad (2.31)$$

$$\frac{C_s(t)}{dt} = k_3 C_{f+ns}(t) - k_4 C_s(t) \quad (2.32)$$

with initial conditions  $C_s(0) = C_{f+ns}(0) = 0$ . In (2.31) and (2.32),  $C_{pl}$  is the plasma tracer concentration corrected for metabolites,  $C_f$  the concentration of free ligand,  $C_{ns}$  the concentration of nonspecifically bound ligand and  $C_s$  the concentration of specifically bound ligand. The free and non specific compartments are merged in one compartment, called non-displaceable compartment, to make the model a-priori uniquely identifiable.  $K_1$  [ml/ml/min] and  $k_2$  [ $\text{min}^{-1}$ ] are, respectively, the rates of transfer to the free plus non specifically bound ligand compartment from plasma and to plasma from the free plus non specifically bound ligand compartment, while  $k_3$  and  $k_4$  [ $\text{min}^{-1}$ ] are the constants of transfer to and from the specifically bound compartment. The measurement equation, assuming negligible blood volume, is:

$$C(t) = C_s(t) + C_{f+ns}(t)$$

In addition to  $K_1, k_2, k_3, k_4$ , it is also possible to estimate the binding potential BP

$$BP = f \frac{Bmax}{Kd} = \frac{k_3}{k_4}$$

where  $f$ ,  $Bmax$  and  $Kd$  represent, respectively, the free fraction of the tracer in the non-displaceable compartment, the local maximum concentration of receptors and the local dissociation rate constant of the reaction between the tracer and its specific receptors.  $C(t)$  can be expressed as a non-linear function of the micro-parameters  $K_1, k_2, k_3, k_4$  ([13]), which, assuming  $C_{pl}$  known, as in our case, can then be estimated through WNLLS.

The generation of simulated TACs was performed using the same principles described in (2.2.3). For completeness the procedure is described again here. A random healthy subject was chosen in the data set presented in [22]; slice n°23 out of the 45 of the image was selected and used for the generation of simulated maps. WNLLS was used to estimate model parameters for each pixel of that slice in the real PET image using the measured plasma TAC. Both in the quantification of the original image and in the analysis of simulated data, weights were set equal to  $C_i/\Delta_i$ , with  $C_i$  and  $\Delta_i$  indicating the concentration of the noise free TAC at frame  $i$  and the duration of the  $i$ -th frame. The estimates at each pixel were then kept fixed and used together with the plasma

TAC to generate error-free continuous time activity curves, which were then sampled at the midtimes of the frames of the real scanning protocol used for the [ $^{11}\text{C}$ ]WAY100,635 dataset. Noisy TACs were then generated adding to the error-free TACs Gaussian noise with zero mean and variance equal to

$$\alpha \frac{C_i}{\Delta_i}$$

with  $C_i$  and  $\Delta_i$  as above. To choose reliable values for  $\alpha$  - constant for each pixel but varying from pixel to pixel-  $\alpha$ -maps were generated during the analysis of the real image by setting for each pixel  $\alpha = \text{WRSS}/(N-m)$ , where WRSS is the value of the weighted cost function evaluated at its minimum, N the number of frames and m the number of model parameters (34 and 4 respectively).

## 2.4.2 Results

Results, presented as usual through BIAS and RMSE, are visualized in Figure (2.7) and (2.8) for 5 different methods: standard WNLLS (1), NLRR1 (2), NLRR2 with CRM to compute priors (3) and NLRR3 with the two different techniques previously described to estimate priors, direct attribution and CRM (4 and 5). For all the algorithms making use of clustering (NLRR2 and NLRR3), results refer to the case of 6 clusters. For both NLRR1 and NLRR3 the modified version of Hurltley ([16]) yielded performances significantly better than standard Gauss-Newton method, and only results computed with Hurltley's approach are therefore reported.

The results deserve a few comments:

- Although changes in BIAS did not follow a clear pattern, reductions of RMSE with respect to standard WNLLS were observed, with all ridge regression methods, for 3 out of 4 parameters. Reductions ranged from medium to relevant.
- On the basis of RMSE, the best results were obtained with NLRR2 and NLRR3, with either method for computing the prior when NLRR3 was used.
- Maps generated with NLRR2 appeared, at a visual assessment, too smooth and resembled original images less than NLRR3 images. WNLLS maps, as expected from RMSE indices, were the noisiest and least regular. These facts are illustrated in Figure (2.9).

Table (2.1) and (2.2) contain RMSE and BIAS computed through NLRR3 with direct attribution of the priors, for different values of the number of clusters employed for the computation of the priors. Notice how results got worse with increasing values of the number of clusters employed, which is probably due to the fact that a higher number of clusters give rise to a poorer WNLLS fit of the cluster mean kinetics due to the higher noise of the TACs. Results corresponding to 3 and 6 clusters, however, were quite similar, whereas there was a relevant worsening with  $N=9$ . The use of a low number of clusters seems therefore advisable, and, as long as this number is kept low, no strong dependence of results on the number of clusters is expected. Employing one cluster would make no sense -all pixels would be assigned the same priors and the algorithms NLRR2 and NLRR3 couldn't work; the use of two clusters only, similarly, does not seem to allow a sufficient differentiation of priors among pixels. Therefore the minimum advisable number is three.

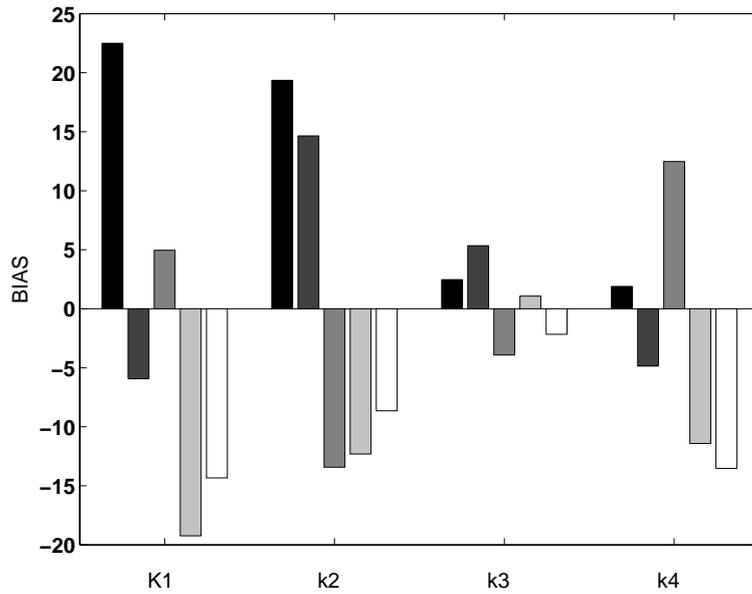


Figure 2.7: *BIAS* averaged over all pixels for the four kinetic parameters of the simulated  $[^{11}\text{C}]\text{WAY100,635}$  data set. Results are displayed respectively for WNLLS, NLRR1, NLRR2 with CRM, and NLRR3 with two different criteria for computing priors (direct attribution and CRM.)

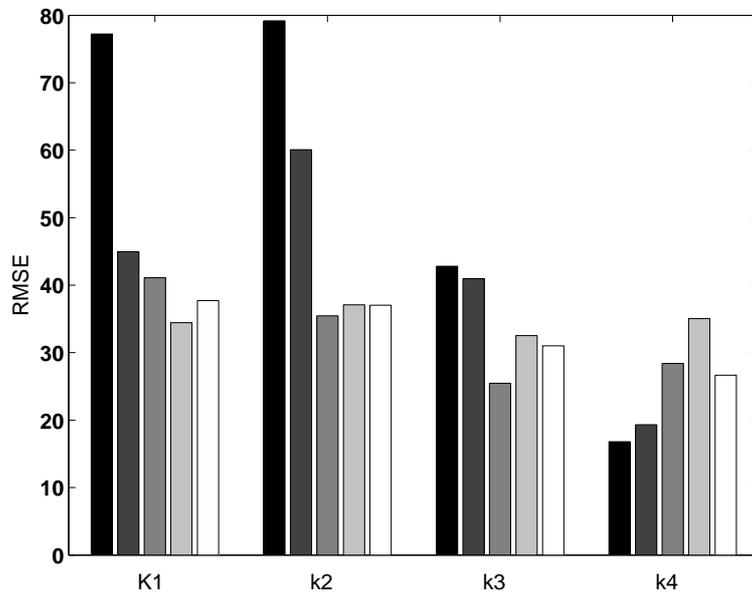


Figure 2.8: *RMSE* averaged over all pixels for the four kinetic parameters of the simulated  $[^{11}\text{C}]\text{WAY100,635}$  data set. Results are displayed respectively for WNLLS, NLRR1, NLRR2 with CRM, and NLRR3 with two different criteria for computing priors (direct attribution and CRM.)

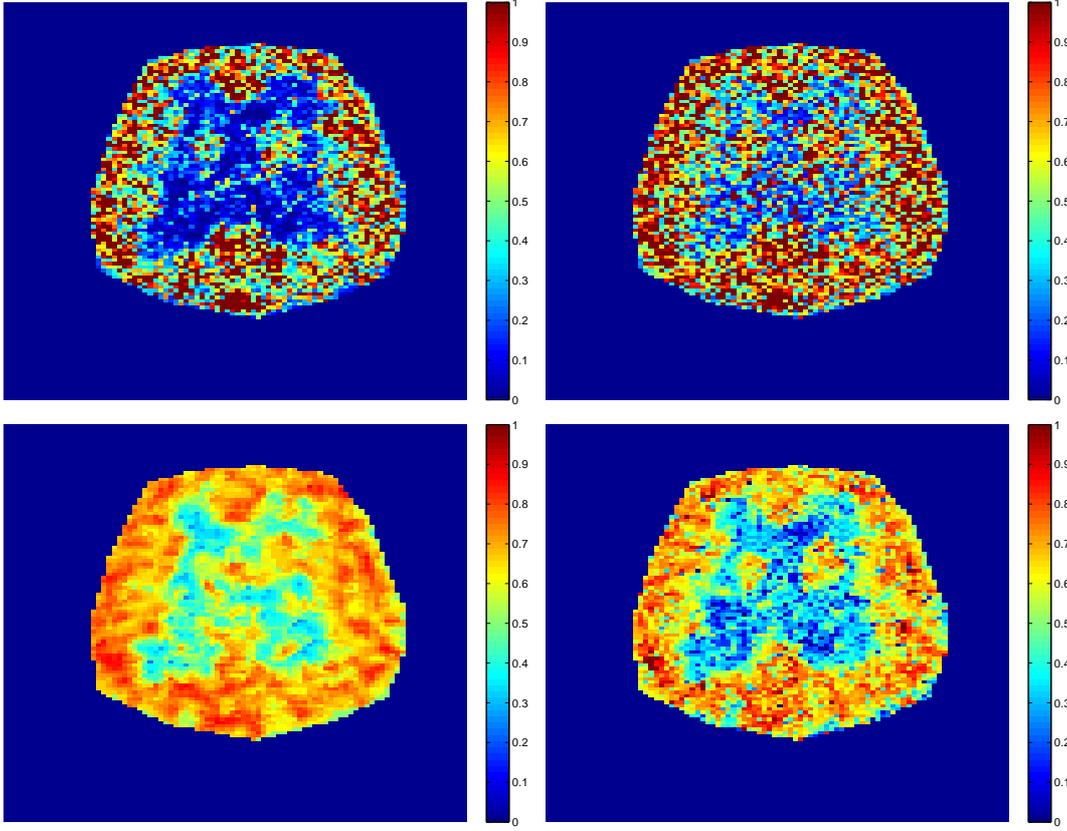


Figure 2.9: "True"  $K_1$  parametric map (upper left), NLLS map (upper right), NLRR2-CRM map (lower left), and NLRR3-CRM map (lower right) are displayed for one repetition of the simulated  $[^{11}\text{C}]$  WAY100,635 data set.

RMSE	3 cluster	6 cluster	9 cluster
$K_1$	34.4	35.2	46.0
$k_2$	39.7	42.6	47.5
$k_3$	32.5	33.7	38.7
$k_4$	35.0	38.5	58.9

Table 2.1: RMSE of NLRR3 with direct attribution of the priors for different values of the number of clusters employed for the computation of the priors.

BIAS	3 cluster	6 cluster	9 cluster
$K_1$	-19.2	-20.1	-14.5
$k_2$	-5.4	-6.6	-11.3
$k_3$	1.1	-1.3	-4
$k_4$	-11.4	-13.3	-39.7

Table 2.2: BIAS of NLRR3 with direct attribution of the priors for different values of the number of clusters employed for the computation of the priors.

### 2.4.3 Comments on computational cost

The issue of computational cost was raised several times before, but it was never analyzed quantitatively. In the linear ridge regression section it was stated that it took roughly thirty seconds to generate maps for a single simulated slice. Even though it was doubled with respect to RLS, in the linear case computational time remains low and does not constitute a problem, even in application to 3D parametric imaging.

In the non-linear scenario the situation is different. With the performances of the computer employed for the analysis it took roughly one hour to generate maps for a simulated [ $^{11}\text{C}$ ]WAY100,635 slice. This value, clearly, depends also on other settings such as the tolerance threshold employed to declare convergence at each non linear problem and on the number of pixels included in the analysis. The total computational time required by NLRR1 and NLRR2 was, in the same way, one hour, as the additional cost for solving the linearized problems at each pixel (phase 2 of NLRR1) or for computing priors from clusters (phase 1 of NLRR2) were negligible if compared to that of the non linear minimizations. With NLRR3, on the other side, it took roughly fifteen minutes to obtain results for one simulated slice, which was mainly required for the precise computation of the sensitivity matrix at each pixel. This value refers to the case of one iteration; the total cost is clearly proportional to the number of iterations performed by the algorithm.

Even if computational time depends remarkably on the tracer, model and settings, it is clear that NLRR3, avoiding the solution of non linear problems, is very appealing from a computational point of view, especially thinking of possible application to 3D generation of parametric maps.

### 2.4.4 Discussion on NLRR

Keeping into account all the comments presented in the results and computational cost paragraphs, the conclusion is that NLRR3, with the use of a low number of clusters, is the best solution. To compute the priors no noticeable difference was noticed between CRM or direct attribution, but it is probably advisable the use of CRM in that it is slightly more flexible in the attribution of priors. As to the number of iterations required, it was noticed that, even though it may take not few iterations to reach convergence with the threshold fixed to 1%, after one or two iterations results were very stable in terms of BIAS and RMSE, and therefore performing a fixed low number of iterations (1 or 2) is a good solution to achieve good results and save computational time.

## 2.5 Parametric imaging of [ $^{18}\text{F}$ ]FDG in skeletal muscle

It may happen that the estimation problem, when the number of parameters is too high, can not be solved at pixel level but only at ROI level due to the lower noise of ROI TACs. In other words when pixel TACs are used, the numerical procedure employed to estimate the parameters of interest may fail to converge to a solution due to the noise level which is too high in relation to the complexity of the model. One possible solution is to simplify the model reducing the number of parameters; this typically causes, however, a loss of physiological information.

In paragraph (2.3) several versions of non linear ridge regression were presented. NLRR1 requires a preliminary fit of pixel TACs to compute the prior and is therefore

unsuitable in this situation, as these fits are assumed unfeasible. The third version of non-linear ridge regression (NLRR3) can instead be used to overcome this problem. NLRR3 was chosen over NLRR2 -which could have been used as well- because, as pointed out in paragraph (2.3), it provides more reliable parametric maps and requires a much lower computational time.

To exemplify this issue real data from a 90 minutes  $[^{18}\text{F}]\text{FDG}$  scan of human skeletal muscle were considered. A detailed discussion on the tracer  $[^{18}\text{F}]\text{FDG}$ , on its model and on simplified procedures for parameter estimation will be given in the following chapter. For the purpose of this paragraph it is sufficient to know that it has been shown in [28] that a five-constant three-compartment irreversible model (5K-model) is the most appropriate to describe the kinetics of  $[^{18}\text{F}]\text{FDG}$  in human skeletal muscle. In this model  $K_1$  and  $k_2$  represent, respectively, the rate of delivery of the tracer to the extracellular pool from the blood pool and viceversa,  $k_3$  and  $k_4$  the delivery to the intracellular pool from the extracellular pool and viceversa, and  $k_5$  is the rate of phosphorylation of  $[^{18}\text{F}]\text{FDG}$ . Assuming the blood volume fraction  $V_b$  equal to 0, five coefficients remain to be identified using standard WNLLS; at pixel level, at least with the available data set (described in [27]), the minimization problem does not converge to a solution. Figure (2.10) displays a random pixel TAC; notice the high level of noise which prevents the identification. The natural simplification of the 5K model leads to the traditional 3K model, in which the extracellular and intracellular compartments of the 5K model are merged together. The simpler three-constant two-compartment irreversible model (3K) is the most used for  $[^{18}\text{F}]\text{FDG}$  studies in brain, but, when applied to human skeletal muscle, the 3K model brings about a loss of physiological information as the extracellular and intracellular compartments are merged together. For skeletal muscle therefore, the use of the 5K model is preferable when its identification is possible.

A  $[^{18}\text{F}]\text{FDG}$  scan of an healthy subject in basal state was considered (see [27] for details on protocol) and a random slice out of the 62 transaxial slices of the PET image was selected in the middle (slice n°30). After manually creating a mask comprising only pixels belonging to the image (i.e. to the legs of the subject) a hierarchical cluster algorithm was applied to segment those pixels into 3 homogeneous clusters and the corresponding mean TACs were computed averaging the appropriate pixel TACs.

Figure (2.11) shows how pixels were divided: note how well the algorithm worked, with the 3 clusters roughly corresponding to the outer, central and inner part of the leg. After that the 5K parameter model was fitted to the 3 clusters kinetics with the measured plasmatic tracer concentration of the subject used as input for the model and weights expressed by

$$\frac{\Delta t_i}{C_i^{ROI}} \quad (2.33)$$

After that, NLRR3 was implemented with the direct attribution criterion to determine priors.

Figure (2.12) and figure (2.13) display the results obtained for the kinetic parameters  $K_1$  and  $k_4$ , respectively. On the left is the initial parametric image that was used as prior and on the right the output of the NLRR3 (Generalized version). Different scales were employed for the representation of the same map, because if the same scale had been used one image would have been characterized by a high loss of details. As expected the images representing the priors are smoother and characterized by a significant loss of details with respect to NLRR images. The final parametric image of  $K_1$  obtained with NLRR3 was of good quality and anatomical aspects such as the presence of the

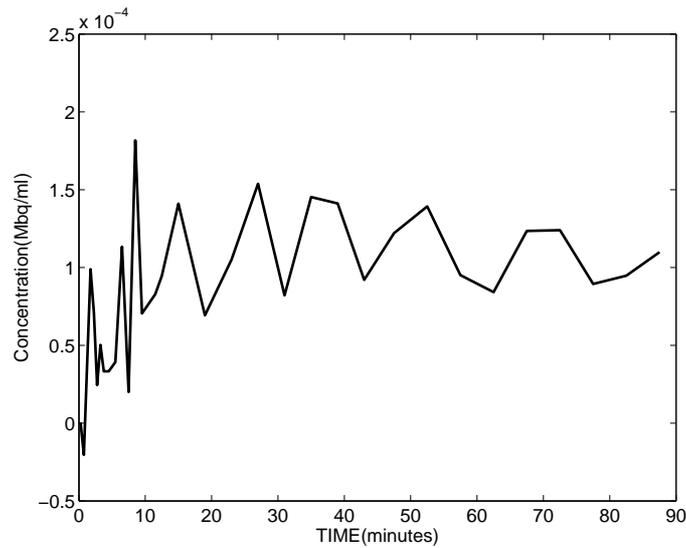


Figure 2.10: A randomly chosen pixel TAC is displayed; notice the high level of noise, which makes the application of the 5K model prohibitive.

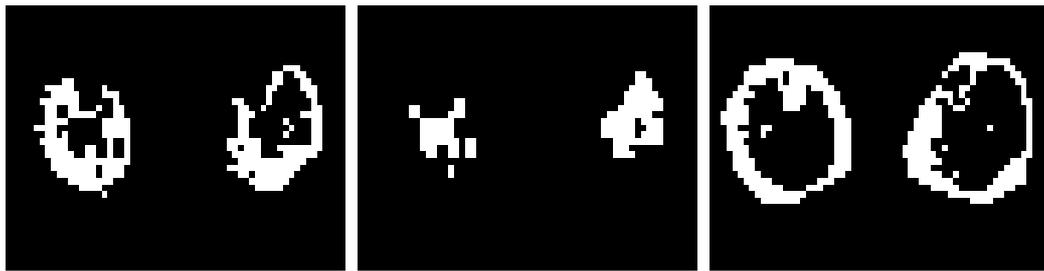


Figure 2.11: The segmentation of pixels into 3 clusters by means of hierarchical clusterization is displayed.

bones (upper part, right leg especially) and of the vascular components (veins and arteries, center of both legs) became visible only in the map computed with NLRR3.  $k_4$  parametric maps computed with NLRR3, on the other side, despite showing a certain symmetry and being more detailed than the corresponding  $k_4$  priors, preserved a high degree of noisiness and irregularity. The same considerations could have been drawn for  $k_3$  and  $k_5$ ;  $k_4$  was chosen just for the sake of space.

The conclusion is that, in this context, non-linear ridge regression is a powerful tool in that it allows to compute parametric maps that couldn't be created with conventional WNLLS. The quality of these parametric maps, judged by visual assessment, clearly depends on the amount of noise in the original image and, for a given data set, on the specific kinetic parameter considered.

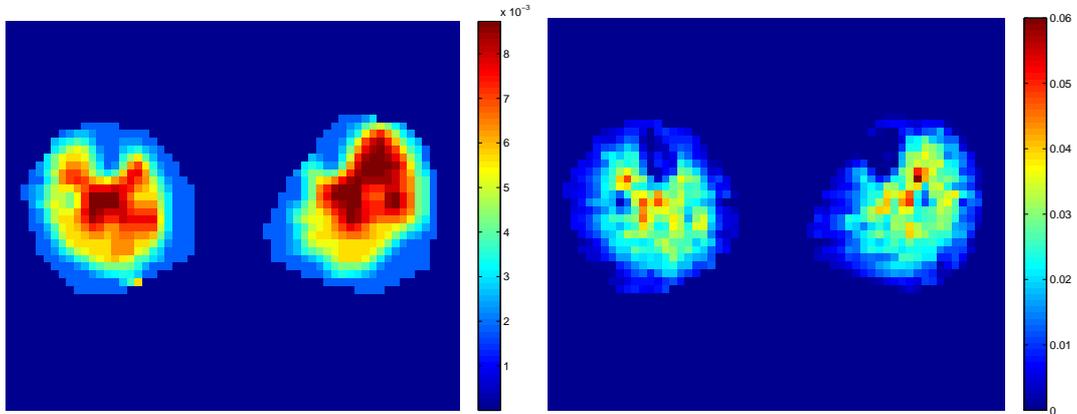


Figure 2.12: The parameter  $K_1$  is displayed. On the left is the initial parametric image that was used as prior and on the right the output of the NLRR3. Note that different scales were employed.

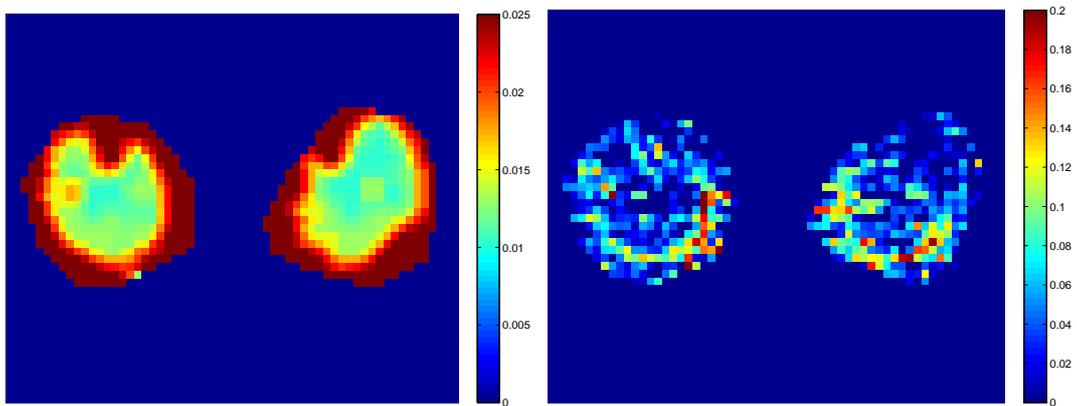


Figure 2.13: The parameter  $k_4$  is displayed. On the left is the initial parametric image that was used as prior and on the right the output of the NLRR3. Note that different scales were employed.

## 2.6 Discussion

In this chapter the mathematical theory of ridge regression in both its linear and non linear versions was introduced and its performances assessed through simulated data sets. Linear ridge regression was employed in the form presented in [7]; for the non-linear counterpart, besides employing the original version ([9]), two alternatives, NLLRR1 and NLLRR2, making use respectively of linearization of the non linear problem and of clusterization to compute priors were tested. In the end, however, the original form of NLLRR, NLLRR3, with a low number of clusters and a fixed low number of iterations (1 or 2), emerged as the best approach, especially keeping into account its low computational cost.

Being Generalized Ridge Regression as simple as Simplified Ridge Regression from the point of view of the equations and of computational burden, the use of the former is preferable as it has a higher degree of flexibility allowing weights of the prior to be different for each kinetic parameter.

Parametric maps obtained using ridge regression were characterized by much lower RMSE with BIAS almost unchanged; at visual assessment they appeared more regular than those obtained with the "gold standard" employed (Weighted Least Squares in its linear or non linear form). However, as shown in the [ $^{11}\text{C}$ ]MP4A analysis, if original estimates are characterized by a too high BIAS and/or a too high RMSE, the improvements brought by ridge regression may be low, because it starts from initial estimates which are too poor. In addition, the example of application of non-linear ridge regression on the 5K model employed to describe [ $^{18}\text{F}$ ]FDG in human skeletal muscle illustrated the power of this analytical tool, which allowed to compute parametric maps that couldn't be created with conventional WLLS. The quality of these maps, judged by visual assessment, clearly depends on the amount of noise in the original image and on the specific kinetic parameter considered.

A word on the computational cost is in order. Linear ridge regression simply requires another linear problem to be solved at each voxel, and in this case, even if doubled, computational time remains in any case low. The version of non linear ridge regression which emerged as optimal (NLLRR3), suitably and brilliantly combining clusterization at the beginning and iterative linearizations avoids the solution of non linear problems and appears very appealing from the point of view of computational time.

## Chapter 3

# Population approaches: theory and validation on simulated data

### 3.1 A few general ideas

As already pointed out previously, pixel-by-pixel quantification may yield unreliable parameter estimates due to the noise of PET data at pixel level. Another possible way to address this problem is provided by the so-called "population" approaches, which are techniques originally developed in the pharmacokinetic/pharmacodynamic sector (see for instance [23]). Population methodologies, as suggested in [24], can be divided into two broad classes. If sufficient measurements are available for each individual to allow the estimation of individual parameter estimates, such estimates may be used as building blocks for further inference: this kind of approach is called a Two-Stage (TS) method. If such data are not available, methods based on linearization (First Order FO, or First Order Conditional Expectation FOCE) must be employed.

In PET, normally, dynamic data allow the estimation of kinetic parameters at pixel level. Therefore, emphasis was put on the two-stages methods with emphasis on the Global-Two-Stage (GTS) and, to a minor extent, on the Iterative-Two-Stage (ITS).

In the population scenario all the homogenous subjects which underwent the same experimental protocol are studied together, with the aim of identifying, in each subject, the individual kinetic model parameters. The intersubject homogeneity is then exploited to improve the parameter estimates for those subjects for which standard WNLLS yielded unreliable estimates or failed to converge. The population parameter estimation paradigm, i.e. poor individual data sets borrow strength from the others if there is some homogeneity, can be clearly applied in a number of situations different from the classical one just described. In the case considered here, for instance, the "subjects" whose estimates are to improve are pixels instead of individuals and the "population" can be considered as made of those pixels belonging to a same ROI or slice or, more generally, showing some homogeneity according to a certain criterion. From now on the expression "individuals" or pixels will be used interchangeably, to refer to the members of the population.

The population paradigm is not completely new in the field of PET as [25] describes a possible application to dynamic PET data. In that work, however, interest was addressed to analysis at ROI level and no study was performed to assess the applicability at pixel level. Besides, only ITS was studied; as already pointed out this method has

a very high computational burden and its application for improving parametric maps is totally unfeasible. The analysis centered on GTS and the application at pixel level, therefore, constitute a new and original approach to PET studies.

## 3.2 The two-stages methods

### 3.2.1 TS

After obtaining an initial estimate of the parameter vector for each of the  $M$  individuals using WNLLS according to the model under investigation and individual data separately, the TS method requires to compute the mean population vector simply averaging the  $M$  estimates as

$$p_{pop} = \frac{1}{M} \sum_{i=1}^M p_i \quad (3.1)$$

and the sample population covariance matrix as

$$\Sigma_{pop} = \frac{1}{M} \sum_{i=1}^M (p_i - p_{pop})(p_i - p_{pop})' \quad (3.2)$$

No update of the estimates of the individual parameter vector is obtained and the method is therefore useless if one aims at improving the individual parametric estimates. No use of the individual covariance matrices is made, leading to a biased estimate of  $\Sigma_{pop}$ . TS has been mentioned both for the sake of completeness and for it is the basis of ITS and GTS.

The last equation can sometimes be found in the form

$$\Sigma_{pop} = \frac{1}{M-1} \sum_{i=1}^M (p_i - p_{pop})(p_i - p_{pop})' \quad (3.3)$$

because

$$\frac{1}{M-1} \sum_{i=1}^M \sigma^2$$

is an *unbiased* estimator for the variance of  $M$  independent identically distributed random variables each of which with variance  $\sigma^2$ , whereas

$$\frac{1}{M} \sum_{i=1}^M \sigma^2$$

is slightly downward biased. In the case considered here, assuming to work with a high number of pixels, the difference is totally negligible.

A schematic representation of the steps performed by ITS and GTS algorithms will be given in the following two paragraphs.

### 3.2.2 ITS

*Iteration 0* Calculate an initial estimate of the parameter vector  $p_i^0$  and of its covariance matrix  $\Sigma_i^0$  for each of the M individuals -typically through WNLLS estimation- using individual data separately. The prime refers to the number of iteration. In the application of this method to pixels of a PET image, the computation of  $\Sigma_i^0$  through WNLLS requires the a-posteriori estimation of  $\sigma^2$  according to (1.7).

*Iteration k (k=1,2,...)*: Compute the mean population vector  $p_{pop}^k$  averaging the M individual estimates  $p_i^{k-1}$  obtained at iteration (k-1) as

$$p_{pop}^k = \frac{1}{M} \sum_{i=1}^M p_i^{k-1} \quad (3.4)$$

and compute the population covariance matrix

$$\Sigma_{pop}^k = \frac{1}{M} \sum_{i=1}^M \Sigma_i^{k-1} + \frac{1}{M} \sum_{i=1}^M (p_i^{k-1} - p_{pop}^k)(p_i^{k-1} - p_{pop}^k)' \quad (3.5)$$

The first term of  $\Sigma_{pop}^k$  reflects the uncertainty of the individual estimates while the second keeps into account the inter-subject variability. Obtain then, for each individual, new estimates  $p_i^k$  for the parameter vector, minimizing (1.11) which we report here with an appropriate change of notation

$$[y - h(p)]' \Sigma^{-1} [y - h(p)] + (p - p_{pop}^k)' (\Sigma_{pop}^k)^{-1} (p - p_{pop}^k) \quad (3.6)$$

in which the mean population vector  $p_{pop}^k$  and the population covariance matrix  $\Sigma_{pop}^k$  were used as the natural substitute to the prior  $\mu$  and covariance of the prior  $\Delta$ . Compute the new covariance matrix  $\Sigma_i^k$  for each individual using (1.13), which is reported here with a simple change of notation,

$$\Sigma_i^k = [(S_i^k)' \Sigma_i^{-1} S_i^k + (\Sigma_{pop}^k)^{-1}]^{-1} \quad (3.7)$$

where  $S_i^k$  and  $\Sigma_i$  are the sensitivity matrix at pixel i and iteration k and the error covariance matrix at pixel i (independent on the iteration number).

Repeat until convergence or until a fixed number of iterations has been performed.

### 3.2.3 GTS

*Iteration 0* As for ITS, calculate an initial estimate of the parameter vector  $p_i^0$  and of its covariance matrix  $\Sigma_i^0$  for each of the M individuals.

*Iteration k (k=1,2,...)*: Compute the mean population vector  $p_{pop}^k$  and the population covariance matrix  $\Sigma_{pop}^k$  according to (3.4) and (3.5) as for ITS. Obtain then new individual estimates  $p_i^k$  of the parameter vector as:

$$p_i^k = [(\Sigma_i^0)^{-1} + (\Sigma_{pop}^k)^{-1}]^{-1} [(\Sigma_i^0)^{-1} p_i^0 + (\Sigma_{pop}^k)^{-1} p_{pop}^k] \quad (3.8)$$

which is equivalent to (1.12), the closed-form solution for the MAP estimator in case the model is linear and  $X = I_n$ , with  $I_n$  denoting the identity matrix of dimension n. Compute the new individual covariance matrix  $\Sigma_i^k$  as

$$[(\Sigma_i^0)^{-1} + (\Sigma_{pop}^k)^{-1}]^{-1} \quad (3.9)$$

Repeat until convergence or until a fixed number of iterations have been performed.

### 3.2.4 Comparison between ITS and GTS

The differences between ITS and GTS are evident. For GTS the measured data of each individual are used only once at the beginning, in the process of estimation of  $p_i^0$  and  $\Sigma_i^0$ , which are then kept *fixed* throughout the algorithm. In ITS, on the other side, individual data are used at each step, since there is the need to minimize (3.6) at each iteration to obtain  $p_i^k$ . An important consequence is that, when the original model is non linear, for ITS a non-linear minimization has to be performed to solve (3.6) at each iteration for each individual making the algorithm computationally cumbersome. If the model is non-linear therefore, ITS is unfeasible due to its computational cost when the number of individuals becomes high, as it would be for the generation of parametric maps. This problem does not hold for GTS, as (3.8) provides a closed-form formula to estimate  $p_i^k$ .

Whereas the interpretation of ITS in a Bayesian framework is straightforward -to the usual cost function one adds a term penalizing the distance from the prior, the population parameter mean- the closed form expression (3.8) of GTS deserves further explanation. By simple algebraic manipulation (3.8) can be written as:

$$(\Sigma_i^0)^{-1}(p_i^k - p_i^0) = (\Sigma_{pop}^k)^{-1}[-(p_i^k - p_{pop}^k)] \quad (3.10)$$

This new expression lends itself to a very elegant interpretation. At iteration k the estimate  $p_i^k$  of individual i is the one such that the weighted difference between  $p_i^k$  itself and the initial parameter estimate of that individual  $p_i^0$  equals the opposite of the weighted difference between  $p_i^k$  and the population average at that iteration  $p_{pop}^k$ . As  $p_i^k$  is somewhere in the middle between  $p_i^0$  and  $p_{pop}^k$  the two quantities  $(p_i^k - p_i^0)$  and  $(p_i^k - p_{pop}^k)$  have different signs and this explains the presence of the sign minus in (3.10) to make the equality hold. The weights clearly equals, from (3.10), the inverse of the appropriate covariance matrices: in the scalar case for instance, the smaller the variance is, the higher its inverse is and therefore the higher the weight is.

It is interesting to notice that GTS and ITS do coincide when the model is linear. In fact, in this case  $h(p)$  can be written as  $Xp$  and the closed-form solution of

$$(y - X p)' \Sigma^{-1} (y - X p)$$

becomes

$$(X' \Sigma^{-1} X)^{-1} (X' \Sigma^{-1} y) \quad (3.11)$$

with covariance matrix equal to

$$(X' \Sigma^{-1} X)^{-1} \quad (3.12)$$

Substituting (3.11) and (3.12) to  $p_i^0$  and  $\Sigma_i^0$  respectively into (3.8) -which is the expression necessary to obtain  $p_i^k$  according to GTS- one obtains:

$$(X' \Sigma_i^{-1} X + (\Sigma_{pop}^k)^{-1})^{-1} (X' \Sigma_i^{-1} y_i + (\Sigma_{pop}^k)^{-1} p_{pop}^k) \quad (3.13)$$

But (3.13) is also the closed-form solution of the MAP estimator when the model is linear, as expressed by (1.12), and it therefore gives the minimum of (3.6), which is the function to be minimized by ITS to compute the estimates of single individuals at each iteration.

### 3.3 Important remarks on ITS and GTS

It is interesting to discuss what happens if the WNLLS estimation made using individual data only, which is the preliminary step for both algorithms, fails to converge for one or more individuals in the population due to noisy data and/or to the complexity of the model. Can these estimates be computed through a population approach? It is quite clear from (3.8) that GTS is useless for this purpose in that it requires, at each iteration, the preliminary estimates that are assumed not to be available for some individuals, who are therefore to be discarded from the analysis. With ITS, on the other side, the use of (3.6) may lead to the estimation of the parameter vector  $p_i^k$  even if  $p_i^0$  couldn't be computed for individual  $i$  as the original function couldn't be minimized. It may also happen, even if it is not common, that  $p_i$  can't be estimated at the  $k$ -th iteration due to the failure of minimization of (3.6), while at  $(k+1)$ -th iteration the estimation is obtained. The introduction of the Bayesian term may, in other words, have a positive impact for the estimation of the parameter vector for some of the individuals for which individual data only were not sufficient for this purpose.

#### 3.3.1 Termination criteria

There are two issues that should be addressed carefully in the application of these iterative approaches in that they heavily affect the final estimates. One is the criterion employed to stop the algorithms. One can finish either when a fixed number of iterations has been performed or when the relative difference of every element of the parameter vector of each individual between the current and the previous iteration is lower than a fixed threshold (e.g. 1%). These criteria can obviously be combined or modified ; for instance one can stop when the relative difference of any parameter between the current and the previous iteration is lower than the threshold, but only for a predefined fixed percentage of individuals, e.g. 90%. Different criteria were tested in order to select the best one.

The issue of convergence -whether the algorithm converges to a solution or diverges- is slightly different from that of termination. If "convergence" is defined in the most natural way as the iteration at which the relative variation of the parameter of interest with respect to the previous one is lower than a specified "small" threshold for all pixels, one can see how convergence and termination are issues strictly related. One could also be tempted to use the just defined convergence criterion as termination criterion. If a too strict criterion to declare convergence is used (e.g. a low threshold such as 0.01%) it may happen -and it often did happen with the simulated data analyzed- that final maps are characterized by very high bias. In other words, convergence in its narrower meaning does not necessarily mean better results in terms of RMSE and BIAS. Therefore less stringent criteria to terminate iterations were heuristically defined and tested.

#### 3.3.2 Segmentation into sub-populations

The other important aspect to address is the decision concerning how to construct the "population(s)". In an application in which all the individuals involved can be considered homogeneous according to a certain criterion (e.g. age, experimental protocol used, kind of disease) it is natural to consider them as part of a unique population. If, on the other side, homogeneity does not hold, it is more convenient to group the subjects into two or more homogeneous sub-populations; for instance data from healthy and ill subjects

are available it is natural to treat the healthy subjects and the patients separately as two distinct groups. Unfortunately, the decision is often not so straightforward.

In applications to parametric maps, a natural and elegant way to address the issue is to group pixels on an *anatomical basis*, that is to consider pixels which are part of the same ROI (cerebellum, thalamus, occipital cortex ..) as belonging to the same sub-population. This approach requires therefore to associate to each ROI the set of pixels which are part of that ROI, which is a time-consuming and not automatic task -results will vary depending on the user performing the segmentation- besides requiring noticeable technical expertise. The possibility to make to this procedure automatic through the normalization of the given image into a stereotaxic space and the use of a predefined atlas will be described later. Even if it may be difficult to employ if applied to the whole 3D image -if one is not making use of the normalization and of the stereotaxic atlas- the use of anatomical criteria may become feasible, for instance, if interest is addressed in a few specific cerebral areas (e.g. striatum in Huntington disease), as it sometimes happens.

A simpler approach is to consider as part of the same population pixels belonging to the *same slice* (transaxial slice) of the reconstructed PET image: in this case the homogeneity criterion is spatial vicinity. In a given slice of a reconstructed image, however, there is commonly a great tissular heterogeneity as gray matter, white matter, cerebro-spinal fluid, skull and blood pool are all present. A third possible approach, less coarse than the previous one, is therefore to further subdivide pixels of the same slice by means of *clustering algorithms* which can group them on the basis of their entire TAC or on the basis of the value of one kinetic parameter. All these methods have been considered and will be presented in detail later.

### 3.4 Comparison between ridge regression and Two-Stages algorithms

After presenting the mathematical formulation of linear and non-linear ridge regression (RR) in Chapter 2, and of Two-Stages algorithms in the previous paragraphs, it is interesting now to compare them in order to highlight their similarities and differences. To better appreciate the relationship between RR and TS algorithms it is convenient to divide the basic equation (2.18), which expresses the function to minimize in RR, by  $\sigma^2$ . It is obvious that the point of minimum is not influenced by the division by a constant and therefore (2.18) is totally equivalent to

$$[y - h(p)]'\Sigma^{-1}[y - h(p)] + (p - \mu)'\Gamma^{-1}(p - \mu) \quad (3.14)$$

in which  $\Sigma^{-1}$  was employed instead of  $W/\sigma^2$  from (1.4) and in which the symbol  $\Gamma^{-1}$  was used for  $H/\sigma^2$ .

#### 3.4.1 Analogies

- The prior is simply computed as mean of the parameters of the population (TS) or of the pixels in the neighborhood of the pixel of interest (RR).
- If (3.14), which is the function to be minimized in RR as just shown, is compared to (1.11), which is the function minimized by ITS at every iteration, it appears

clear that the structure of the two expressions is the same. The first term in both equations, besides, is identical; the only difference is that the prior covariances (indicated on purpose with  $\Gamma^{-1}$  for RR and with  $\Delta^{-1}$  for ITS) are different.

### 3.4.2 Differences

- In RR, prior and prior covariance are *different* for each pixel whereas they are *common* for all the pixels of the population in TS approaches.
- Ridge regression methodologies (except NLRR3) compute the parameters only once, whereas TS approaches are *iterative*, and the estimation of individual parameters, prior and covariances is performed several times until a convergence criterion is satisfied.
- In both RR -always referring to the formulation (3.14)- the covariance of the prior  $\Gamma$  has on its diagonal the squared differences between the individual parameter  $p_i$  and the prior  $p_{i0}$  (averaged over all parameters for SRRSC, see (2.4) and (2.3)). In TS methods the covariance matrix  $\Delta$  (previously referred to as  $\Sigma_{pop}^k$  to indicate its dependence on the iteration number k), contains as in RR the squared differences between the individual parameter  $p_i$  and the prior as expressed by the second term of equation (3.5). In this case, however, these difference are averaged over all the pixels of the population. As the first term of (3.5) shows, besides, in the covariance matrix  $\Sigma_{pop}^k$  the reliability of the individual parameter estimates are taken into account, which is not true for RR.

In conclusion, if one think of RR as a particular case of population approach in which each pixel has its own population - the one comprised by the pixels in its neighborhood- one can notice that the similarities between RR and TS algorithms are high. The iterative nature of TS methods and the difference in the construction of the prior covariance, which in the case of TS keeps into account also the reliability of individual estimates besides their variability, are, however, elements of undeniable difference.

## 3.5 The tracer $[^{18}\text{F}]\text{FDG}$

Most of the simulations of this chapter will deal with the tracer  $[^{18}\text{F}]\text{FDG}$  (fluorodeoxyglucose). It is therefore worth to give a brief introduction on this tracer, on its model and on a simplified technique for estimating its kinetic parameters.

$[^{18}\text{F}]\text{FDG}$  is an analog of glucose, which means that, although it is related to glucose, the parameters which characterize its kinetics are different from those of glucose. Nonetheless it is by far the most used tracer to study glucose metabolism in vivo and also one of the most used tracers in PET.

It was previously stated that a five-constant three-compartment irreversible model (5K-model) is the most appropriate to describe the kinetics of  $[^{18}\text{F}]\text{FDG}$  in skeletal muscle. In the 5K-model the two reversible compartments represent  $[^{18}\text{F}]\text{FDG}$  in the extracellular and intracellular fluids whereas the irreversible compartment quantifies  $[^{18}\text{F}]\text{FDG}$  phosphorylated ( $[^{18}\text{F}]\text{FDG-6P}$ ). For  $[^{18}\text{F}]\text{FDG}$  brain studies the simpler three-constant two-compartment irreversible model (3K) is on the other hand used. The different structure is due to the particular physiological conditions in the human brain created by the presence of the blood-brain barrier. In the 3K-model the reversible

compartment represents [ $^{18}\text{F}$ ]FDG in the tissue pool, with no distinction between intracellular and extracellular pools.  $K_1$  [ml/ml/min] is the delivery from the blood to the tissue pool, but  $k_2$  [ $\text{min}^{-1}$ ] and  $k_3$  [ $\text{min}^{-1}$ ], however, do not have in the 3K model a clear physiological meaning as they are combination of  $k_2, k_3, k_4, k_5$  of the 5K model.  $k_2$  [ $\text{min}^{-1}$ ], therefore, does not represent the rate of delivery from the tissue to the blood pool, and, analogously,  $k_3$  [ $\text{min}^{-1}$ ] is not simply the rate of tracer phosphorylation. As for [ $^{11}\text{C}$ ]MP4A, assuming for the sake of simplicity the blood volume component  $V_b$  equal to 0, the concentration of the tracer in a given pixel or ROI at time  $t$  (time 0 being the beginning of the scan), is given by

$$C(t) = \frac{K_1 k_2}{k_2 + k_3} \int_0^t C_{pl}(x) e^{-(k_2+k_3)(t-x)} dx + \frac{K_1 k_3}{k_2 + k_3} \int_0^t C_{pl}(x) dx \quad (3.15)$$

where  $C_{pl}$  denotes the measured plasmatic concentration of the tracer. A similar but more complicated equation relates the 5 parameters of the 5K model to the measured concentration.

An important parameter for [ $^{18}\text{F}$ ]FDG is the so-called irreversible uptake rate constant  $K_i$  [ml/ml/min], which equals the amount of tracer which enters the irreversible compartment for unity of time and unity of volume: basically the parameter quantifies the rate at which, in stationary conditions, the tracer is irreversibly trapped. In the case of [ $^{18}\text{F}$ ]FDG this parameter is related to the intensity of the phosphorylation process, the rate at which glucose is metabolized into cells. In fact,  $K_i$  is proportional to Metabolic local Rate of Glucose consumption MRGlu [mg/dl/min] according to

$$MRGlu = \frac{K_i C_{gl}}{LC} \quad (3.16)$$

where LC [unitless] is the Lumped Constant which summarizes the differences between [ $^{18}\text{F}$ ]FDG and glucose kinetic parameters ([28]) and where  $C_{gl}$  [mg/dl] represents glucose concentration in plasma.  $K_i$  has therefore an important biological meaning, being more relevant than any single kinetic parameter itself. The relation between  $K_i$  and the kinetic parameters is given by

$$K_i = \frac{K_1 k_3}{k_2 + k_3}$$

for the 3K model and by

$$K_i = \frac{K_1 k_3 k_5}{k_2 k_4 + k_2 k_5 + k_3 k_5}$$

for the 5K model.

### 3.5.1 Patlak plot

Patlak plot is one of the most popular approaches for the quantification of irreversible tracers. The method can be applied to irreversible models only (i.e. models which have at least one compartment in which the tracer is irreversibly trapped) and requires to consider the discrete quantities

$$x_i = \int_0^{t_i} C_{pl}(\tau) d\tau / C_{pl}(t_i)$$

and

$$y_i = C_{Target}(t_i) / C_{pl}(t_i)$$

for every scan time  $t_i$  greater than an appropriate  $t^*$  such that, for all  $t$  greater than  $t^*$ , all the reversible compartments are in equilibrium.  $C_{Target}(t_i)$  indicates the tracer concentration in a ROI or pixel at frame  $i$ , while,  $C_{pl}(t)$ , as usual, denotes the time course of the tracer concentration in plasma. The method allows to estimate the irreversible uptake rate constant  $K_i$  previously described.

The assumption made by Patlak method is that there exists an instant  $t^*$  such that, for  $t$  greater than  $t^*$ , all the reversible compartments are in equilibrium. In other words the ratio of the tracer concentrations of any reversible compartment to the plasmatic compartment is assumed constant for all  $t$  greater than  $t^*$ , or, equivalently, the time-derivative of the tracer concentration of all reversible compartment is assumed to be equal to zero for  $t$  greater than  $t^*$ . This is typically verified within the duration of the scan, although the determination of the smallest  $t^*$  is not easy.

The central idea in Patlak plot is that the relation between  $y$  and  $x$  defined before, for  $t$  bigger than  $t^*$ , can be shown to be linear with slope equal to the irreversible uptake rate constant which can therefore be estimated using linear least squares.

If  $t_1$  is such that the equilibrium hypothesis is verified,  $t_2$  greater than  $t_1$  is assumed to be a valid value as well. It is therefore safer to use a high value for  $t^*$  especially when the scan length is high allowing to plot a sufficient number of points for  $x$  and  $y$  even with a large  $t^*$ . In fact, the use of a value lower than the minimum  $t^*$  may lead to BIAS in the estimation of  $K_i$ . The choice of  $t^*$ , on the other side, must be such that the number of mid-scan values greater than  $t^*$  is not too low, as this would give rise to unreliable estimates for  $K_i$  in the linear least squares estimation.

### 3.6 Performances of Two-Stages algorithms

In this section performances of ITS and GTS on a simulated data set will be compared analyzing the influence of factors such as

- noise level of data
- intrinsic variability of the population
- number of subjects forming the population
- amount of data available for each subject

on the overall performances of the algorithms.

The simulations described here are addressed to the study of general aspects of population techniques; an analysis more focused on realistic PET data will be the presented later on in this chapter.

#### 3.6.1 Generation of the basic simulated data set

A real data set comprised of 8 healthy subjects which underwent a 90-minutes [ $^{18}\text{F}$ ]FDG scan of their legs was used to construct the simulated data set ([27]). The 3K model was used because, as explained later in this paragraph, the estimation of kinetic parameters was performed both at ROI and at pixel level, and the identification of the 5K model is not feasible at pixel level, as previously illustrated. The blood volume component  $V_b$  was set equal to 0. ROIs were placed over the anterior tibialis and soleus muscles of the subjects using the corresponding MRI scans and in this way the corresponding ROI

TACs were extracted from the image; the soleus muscle only, however, was used in the analysis. Further details on the experimental protocol and procedure used to acquire the images can be found in [27].

The 3 kinetic parameters  $K_1$ ,  $k_2$  and  $k_3$  were estimated for each subject using the gold standard WNLLS with the measured plasmatic concentration time course. A multivariate Gaussian distribution was then assumed for the parameter vector  $K=[K_1, k_2, k_3]$ , with mean  $\mu$  set equal to the mean of the 8 estimated parameter vectors  $K_i$ ,  $i=1,2,..8$ , and covariance  $\Sigma$  simply computed as  $\frac{1}{8} \Sigma(K_i - \mu)(K_i - \mu)'$ .

A simulated data set made of 1000 subjects was then created using the measured plasmatic concentration time course of one of the eight subject and random samples drawn from the previous multivariate Gaussian distribution for the parameter vector  $K$ . 5 repetitions were performed obtaining in this way 5000 simulated TACs; in other words 5 populations of 1000 subjects each were created. These continuous TACs were then sampled to yield discrete TACs according to the time grid employed for the real [ $^{18}\text{F}$ ]FDG data set: 8 frames of 0.5 minutes each, 9 of 1 minute, 8 of 4 minutes and 9 of 5 minutes for a total of 90 minutes and 34 frames.

To simulate noise levels resembling the ones which are seen at ROI and pixel level respectively, (1.7) was used to estimate the proportionality coefficient  $\sigma^2$ . More precisely, for the ROI level,  $\sigma_{ROI}^2$  was obtained averaging the 8 values  $\sigma_i^2$  computed for each subject after the WNLLS estimation of the soleus TACs. For the pixel level noise of a single slice of the PET scan of a subject randomly chosen was used: WNLLS and the the measured plasmatic concentration function were employed to estimate parameters, thus obtaining an estimate of  $\sigma^2$  at each pixel, whose average was set equal to  $\sigma_{pixel}^2$ . This simulated data set, with pixel level noise, will be referred to as *basic* or *high noise* data set.

Performance were assessed employing the usual RMSE and BIAS criteria, considering the single kinetic parameter  $K_1$ ,  $k_2$ ,  $k_3$  or, alternatively, the macro-parameter  $K_i$ , which is of particular importance for [ $^{18}\text{F}$ ]FDG. Results were averaged over the 5 populations. Results of the different simulations were not presented each time all the kinetic parameters, but one among  $K_1$ ,  $k_2$ ,  $k_3$  and  $K_i$  was chosen according to which one best illustrated the aspect under examination.

A reasonable observation is that little can be inferred from these simulations regarding applications to real PET data at pixel level. Which is, for instance, the correlation between the parameter variability of the pixels of a given slice or region on one hand and the inter-subject parameter variability used for this simulated data set, on the other? The obvious answer is that there is no correlation and that these simulations were designed in this way only in order to illustrate some general aspects of the TS methods. Application to PET data simulated at pixel level will be presented later in this chapter.

### **Variation of the basic simulated data set**

As stated above, several settings were changed with respect to the basic case in order to study the behavior of the algorithms in different conditions.

### **Intrinsic variability of the population**

To change the level of homogeneity of the population two new sets of kinetic parameters were generated, both with Gaussian distribution with the same mean value as above, but with a covariance matrix respectively equal to  $0.25\Sigma$  (more homogeneous population) and to  $4\Sigma$  (more inhomogeneous population), where  $\Sigma$  is the covariance matrix

of the basic data set. These values were chosen in order to reduce and to increase the standard deviations by a factor 2 each time. After that, TACs were generated in the way described above. In this analysis, as well as in the following ones, GTS only was considered as ITS provided similar results in the first set of simulations, but its computational burden was much higher.

### **Number of subjects forming the population**

As to the dependance on the number of subjects forming the population, the basic data set was used again, but this time the 1000 simulated subjects, instead of being used together, were divided into 10 groups of 100, 20 groups of 50, and 100 groups of 10 subjects respectively. GTS was then applied each time using these new sub-populations separately, and final results were computed by averaging the results of each sub-population.

### **Amount of data available for each subject**

In the end, to test what happens with a reduced scan time, two new sets of WNLLS estimates were computed, using only the first 28 and 25 frames out of the 34, which correspond to a scan time of 60 and 45 minutes respectively. GTS was then applied to these new estimates and to their corresponding variances to study what happens when poorer initial WNLLS estimates are available. The aim was to assess the potential applicability of population approaches to reduce the scan time in real applications.

### **3.6.2 Results: different noise levels**

The first outcome of the analysis was, as expected, that for noise level typical of ROIs the estimates provided by GTS and ITS coincide with those computed with WNLLS. In other words, when signal-to-noise ratio is high, the estimates provided by WNLLS alone are already of good quality and the application of TS methods is simply useless in that it does not determine any appreciable difference.

Noticeable improvements, on the other side, were observed when noise level typical of pixels was considered. Figure (3.1) and (3.2) show, respectively, RMSE and BIAS for WNLLS, ITS and GTS. The relative reduction of RMSE, for instance, was roughly of 25%; ITS and GTS, besides, yielded almost identical performances.

Due to the results of this preliminary analysis, which is totally coherent with what was expected, "pixel" TACs only will be considered in the rest of the paragraph.

### **ITS and GTS computational cost**

Therefore, having provided performances similar to GTS, ITS was not be considered any more in the rest of the thesis as its computational cost makes it totally unsuitable when interest is addressed in generating parametric maps.

### **3.6.3 Results: population homogeneity**

Figure (3.3) displays RMSE for the parameter  $K_1$  for the low and high level of population variability. Horizontal lines refer to WNLLS results, whereas curved lines display the performances of GTS as a function of the iteration number. It can be seen that, for the low level of population variability, WNLLS gave a value of 53% whereas the error of GTS, at its minimum, reached 37% with a relative decrease of 32%. For the high

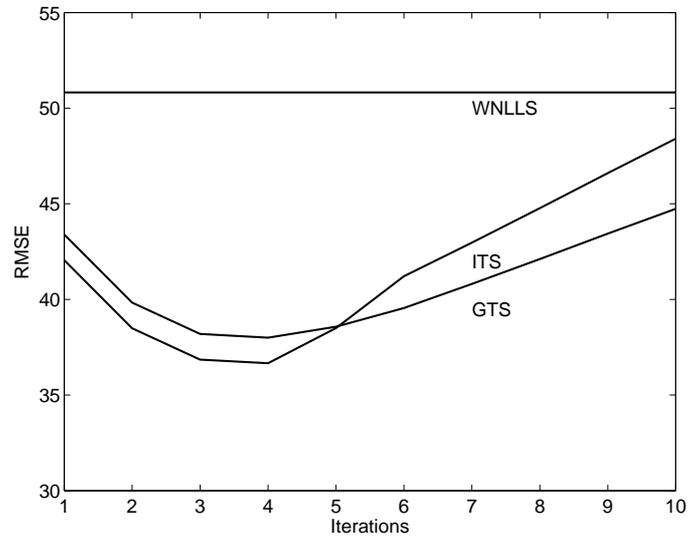


Figure 3.1: *RMSE for  $K_i$  [ml/ml/min], computed first with WNLIS and subsequently with GTS and ITS, is displayed for the pixel level noise simulation.*

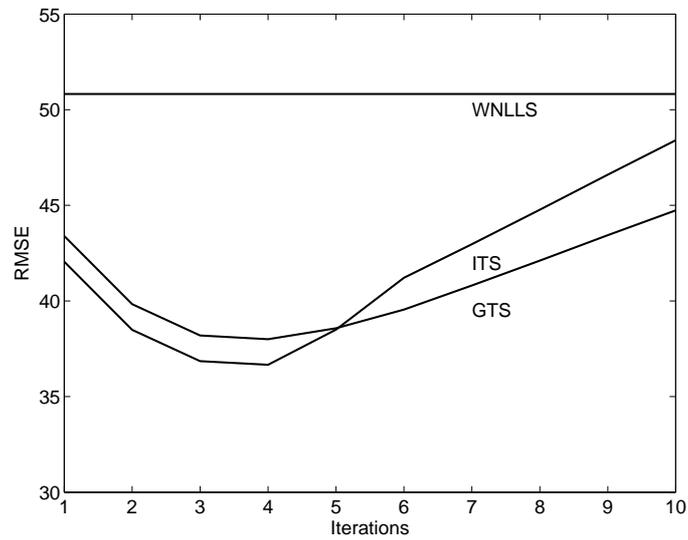


Figure 3.2: *BIAS for  $K_i$  [ml/ml/min], computed first with WNLIS and subsequently with GTS and ITS, is displayed for the pixel level noise simulation.*

population variability the relative decrease was only of 19%, from 59% of WNLLS to 48% of GTS. The results, as expected, illustrate the obvious idea that a lower population homogeneity is associated to lower improvements of TS with respect to WNLLS because of the higher intrinsic population variability. This simple results suggests that attention should be given to population homogeneity when applying GTS, as lower parameter variability is typically associated to higher potential improvements.

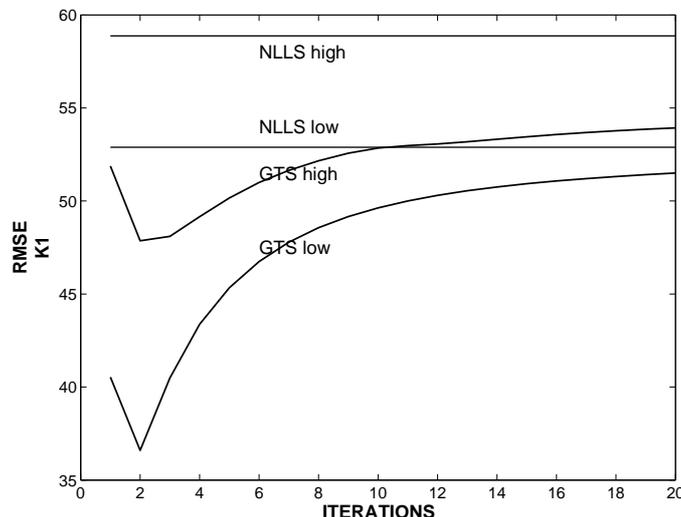


Figure 3.3: *RMSE for  $K_1$  [ml/ml/min], computed first with WNLLS and subsequently with GTS, is displayed for the high and low levels of population variability.*

### 3.6.4 Results: number of subjects forming the population

Figure (3.4) and (3.5) display respectively RMSE and BIAS for  $K_i$ , for different population sizes, computed with GTS. The main outcome was that performances of the algorithm were monotonously dependent on the number of subjects forming the population: the higher the number, the better the results.

### 3.6.5 Results: amount of data available for each subject

The effect of reducing the scan length on GTS is illustrated in figure (3.6). The percentage reduction of RMSE between WNLLS and GTS, expressed by

$$100\left(\frac{RMSE_{WNLLS} - RMSE_{GTS}}{RMSE_{WNLLS}}\right) \quad (3.17)$$

is displayed for  $K_1$ ,  $k_2$  and  $k_3$  for the 3 different scan lengths (90,60 and 45 minutes) employed to compute WNLLS estimates. GTS RMSE clearly depends on the iteration number; here its minimum value was taken. The selection of the best stop criterion will be the object of the following paragraphs. The bottom line is that, except for  $k_3$  RMSE reduction which remained substantially unchanged, the shorter the scan time was, the higher the improvement was.

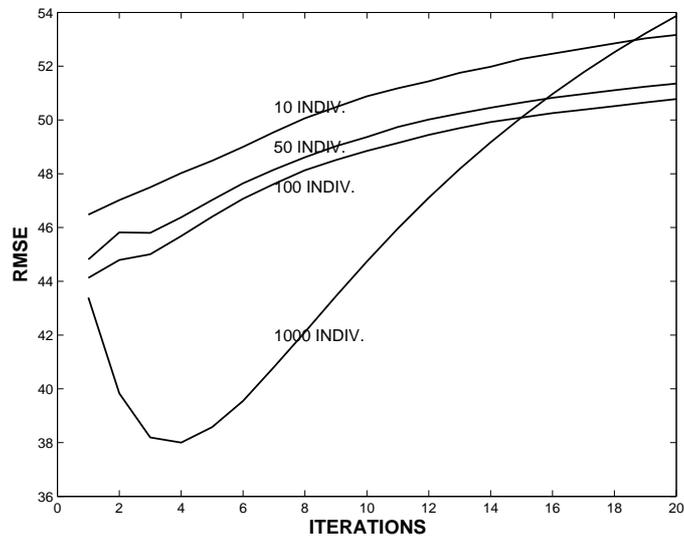


Figure 3.4: *RMSE of  $K_i$  [ml/ml/min], computed with GTS, is displayed for different population sizes.*

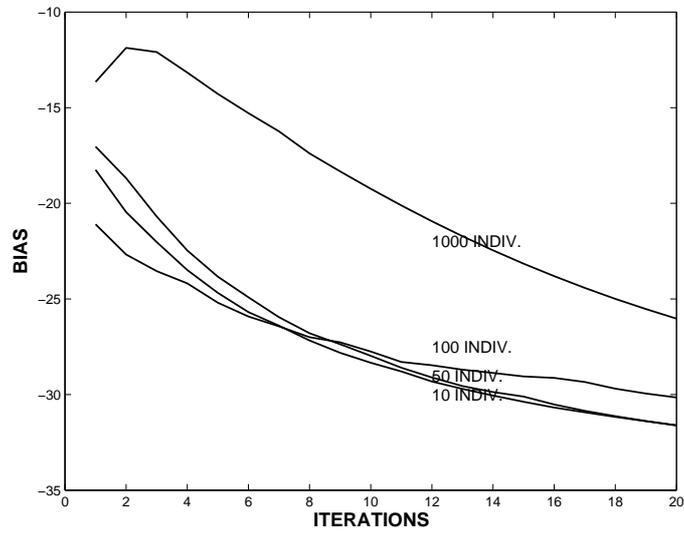


Figure 3.5: *BIAS of  $K_i$  [ml/ml/min], computed with GTS, is displayed for different population sizes.*

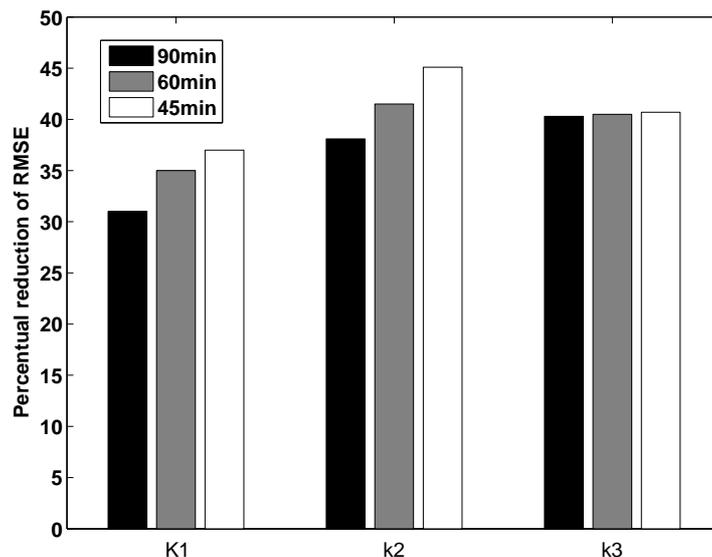


Figure 3.6: The percentage reduction of RMSE between GTS and WNLLS, as expressed by (3.17), is displayed for  $K_1$ ,  $k_2$  and  $k_3$  for the 3 different scan lengths (90, 60 and 45 minutes) employed to compute WNLLS estimates.

### 3.6.6 Brief summary of the main outcomes

The main results of the previous simulations can be summarized as follows. No difference between traditional WNLLS and the population approaches was noticed when considering data with a noise level typical of ROIs. In that case the estimates provided by WNLLS alone are already of good quality. Appreciable improvements took place instead when noisier data were considered, with a large decrease of RMSE and BIAS.

The use of more homogeneous and inhomogeneous populations, as expected, provided improvements respectively of higher and lower extent with respect to the case of "medium" homogeneity.

Interesting was the fact that performances were monotonously dependent on the number of individuals in the populations: the higher the number, the better the results. It is clear that, in a real application not only to PET data but also in a more general pharmacokinetic scenario, increasing the number of subjects lead to a more inhomogeneous population, and therefore a trade-off between these two different requirements should be searched.

Last, the application of GTS to WNLLS estimates obtained from a reduced scan time yielded even larger improvements, according to the general principle that, within certain limits, the worst the initial estimates are, the higher the improvement of GTS is. The reason is that parameter estimates which are unreliable, completely or partly, are to a certain extent "corrected" by the application of the algorithm.

## 3.7 Pixel level simulations: linear case

The simulations of the previous paragraph were performed in order to illustrate some general aspects of the two-stages-methods. Little, however, can be inferred from them

regarding applications to real PET data, except for the fact that population approaches can be helpful only at pixel level.

3 important questions, however, remain unanswered. Is the parameter variability inherent in the pixels of a given slice or region low enough to allow to employ population approaches successfully? Which is the optimal tradeoff between the need for homogeneity on one side and the need to employ "big" populations on the other? Which is the best criterion to stop the iterative algorithms? In the last paragraph, little attention was paid to these aspects, and especially to the last one, as results were presented as a function of the iteration number. In real application, clearly, one and only one final estimate is required and the need for a stop criterion which is optimal from the BIAS/RMSE point of view and relatively independent on the dataset/tracer used is crucial.

### 3.7.1 Simulated data set

One subject of the [ $^{18}\text{F}$ ]FDG data set presented in the previous paragraph and described in [27] was employed to generate the simulated pixel TACs. As the subject considered was studied under fasting conditions, this first data set will be referred to as "basal state" data set. The procedure employed to generate realistic TACs was already described but will be nonetheless recalled here. A random slice in the middle of the field of view (slice n°30 out of the 62) was selected and the measured plasmatic function of the subject was employed to compute  $K_1$ ,  $k_2$  and  $k_3$  at each pixel using the standard weights  $\Delta_i C_i$  and WNLLS for the estimation. The 3 rate constants just computed and the plasmatic function were then employed at each pixel to generate noise-free TACs using the time protocol of the real scan. Gaussian noise with zero mean and variance equal to  $\alpha C_i / \Delta_i$  was then added. To generate TACs with a noise level similar to that of real data  $\alpha$ -maps were generated during the quantification of the original [ $^{18}\text{F}$ ]FDG image, by computing for each pixel  $\alpha = \text{WRSS} / (\text{N} - \text{m})$  as in (1.7), where WRSS, as usual, is the value of the weighted cost function evaluated at its minimum, N the number of frames and m the number of model parameters (34 and 3 respectively in this case). These maps were then used in the process of noise generation.

In an identical way the same slice of another subject who underwent a [ $^{18}\text{F}$ ]FDG scan during a euglycemic insulin infusion at 30 mU/min per  $\text{m}^2$  of Body Surface Area was employed to create a simulated "insulin" dataset. 100 repetitions of both slices (basal and insulin) obtained adding different random samples with the statistic distribution just described were generated

Patlak plot was used both as the method to which compare results to and for the generation of the initial parameter and variance estimates from which GTS starts. Being Patlak model linear in the parameters of interest, in this context GTS coincides with ITS, as previously explained. The irreversible uptake rate constant was used in the simulation as the parameter for performance assessment. The last 12 frames were used for the application of Patlak plot: this corresponds to a  $t^*$  of 33 min for the stationary state.

### 3.7.2 Segmentation

As pointed out previously, there are two settings in the application of population approaches that are user-dependent and which may strongly affect final results: the criteria used respectively to group pixels into sub-populations and to declare convergence.

As to the strategy employed for the segmentation, several possibilities were tested. First GTS was applied to all the pixels of the image: this approach will be referred to as GTSraw. Then segmentation based on hierarchical cluster analysis using pixel TACs was performed, with the employment of 2 and 3 clusters (the use of one clusters would correspond to the use of GTSraw).

In the end, pixels were segmented into 2 and 3 clusters using, this time, the preliminary estimates of  $K_i$  obtained through the application of Patlak approach -that would be in any case the starting point for the application of GTS. In other words after applying Patlak method to get an estimate of  $K_i$  at each pixel, these values were used to segment the images into 2 or 3 sub-populations. This method was called GTSpop, where "pop" stands for the populations into which pixels were segmented. The idea is that if one is interested in a specific kinetic parameter  $\alpha$  and the need arises to segment pixels according to their homogeneity with respect to  $\alpha$ , clusterization based on dynamic data may not be totally appropriate because, if two pixels have "similar" TACs, they do not necessarily have similar values of  $\alpha$ . In this "static" segmentation scenario besides, hierarchical cluster analysis was tried but then discarded. First the method is very time-consuming and secondly, due to the unavoidable presence of outliers in the parametric image used for the segmentation, it often happened that one or more clusters were comprised of just few pixels with unreasonably high or low initial estimates whereas all the remaining were clustered together. A simple division of the physiological range of the kinetic parameter of interest, computed according to Patlak preliminary estimates, into a number of intervals equal to the desired number of clusters, was instead used. This simple procedure allowed to segment pixels into clusters homogeneous with respect to the parameter of interest in a very efficient fashion.

5 different segmentations (none, 2 based on dynamic data, 2 based on the subdivision of preliminary estimates) were therefore performed for each repetition and GTS was then employed with each sub-population treated separately in the realization of the algorithm. A low number of clusters was employed because the number of the considered pixels in the images was low (400-500). A higher number is probably more appropriate when performing analysis on brain images.

### 3.7.3 Termination criteria

For all the techniques employed results were first computed as a function of the iteration cycle.

Three different stop criteria were then tested (variation less than 1% for  $K_i$  for at least 90%, 95% and 99% of individuals of each sub-population). These versions of the algorithms with the use of a stop criterion will be called GTSraw\* and GTSpop\*, respectively. As there is no single "optimal" number of iterations, because it clearly depends on factors such as the amount of noise and on the compartmental model analyzed, it is not reasonable to employ a fixed number of iterations for all tracers. The search for a stable criterion to terminate the algorithm, which is fairly tracer and noise independent, is therefore important. The simulations performed with other tracers (see next paragraph) are interesting in this sense in that they allowed to test whether a criterion that seems appropriate -maybe optimal- in one situation remains robust with different compartmental model, amount of noise, parameters considered and estimation technique employed.

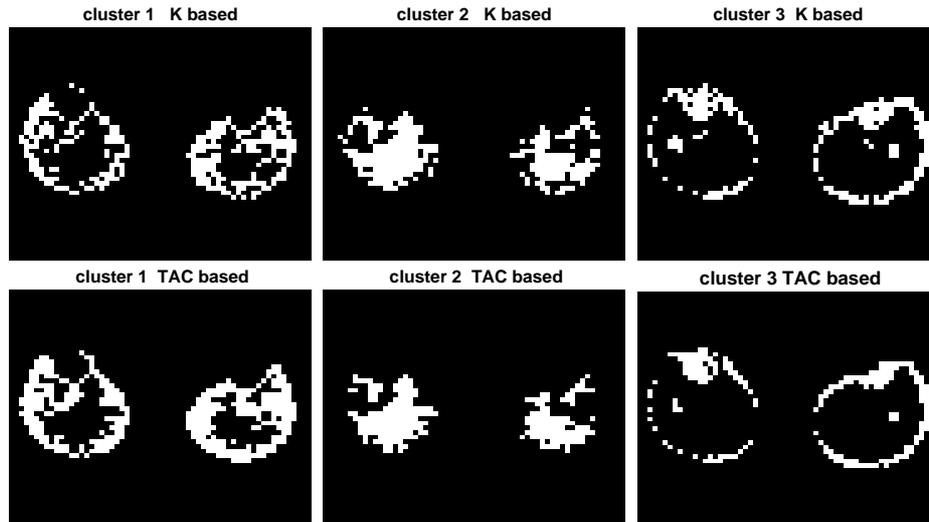


Figure 3.7: *The segmentation of one simulated slice of the insulin data set into 3 clusters is displayed. Above is the segmentation based on the irreversible uptake rate constant, below hierarchical clusterization based on pixel TACs.*

### 3.7.4 Results

A first comparison between results obtained using clusters based on pixel TACs and those based on the preliminary Patlak estimates showed that, even if differences were sometimes negligible, the "static" segmentation based on the irreversible uptake rate rate yielded slightly better performances. In the rest of the paragraph, therefore, results (RMSE and BIAS) obtained with the TAC-based segmentation will be omitted for the sake of space. In addition, no noticeable difference was noticed between results obtained with 2 and 3 populations, and only those corresponding to  $n=3$  will be presented. Some differences, on the other side, were found between GTSpop and GTSraw, and therefore results corresponding to both methods will be discussed.

Figure (3.7) shows the output of the segmentation of one simulated slice for the insulin data set. In the upper part the segmentation based on the irreversible uptake rate constant is presented, in the lower part of the figure the division based on the application of hierarchical clusterization to pixel TACs is presented. Note the quality of the clusterization, with the image segmented in three zones corresponding roughly to the outer, central and inner part of the legs; note also the similarity between the two different techniques of segmentation, which, on the other side, provided different results for the basal data set.

The stop criterion which was selected as the best one was the 90% one -variation less than 1% for at least 90% of the pixels of the population under examination; results presented here refer to this criterion. But, as illustrated below, where evaluation indices are plotted as a function of the iteration number, results stabilized quickly after a few iterations (2-3) and the differences between RMSE and BIAS obtained after 3 and after 20 iterations were never larger than 3%-4%. This is a positive fact in that it shows how GTS, at least with these simulated data sets, is robust with respect to the number of iterations performed, and, consequently, to the criterion used to stop the algorithms.

Figures (3.8) and (3.9) show BIAS and RMSE for the basal state simulation. The dotted line refers to plain Patlak method, dashed lines correspond to GTS-based methods when a convergence criterion was used, and the solid lines represent performances of GTSpop and GTSraw as a function of the iterations. Figures (3.10) and (3.11) represent the same parameters for the insulin data set, with the same symbology.

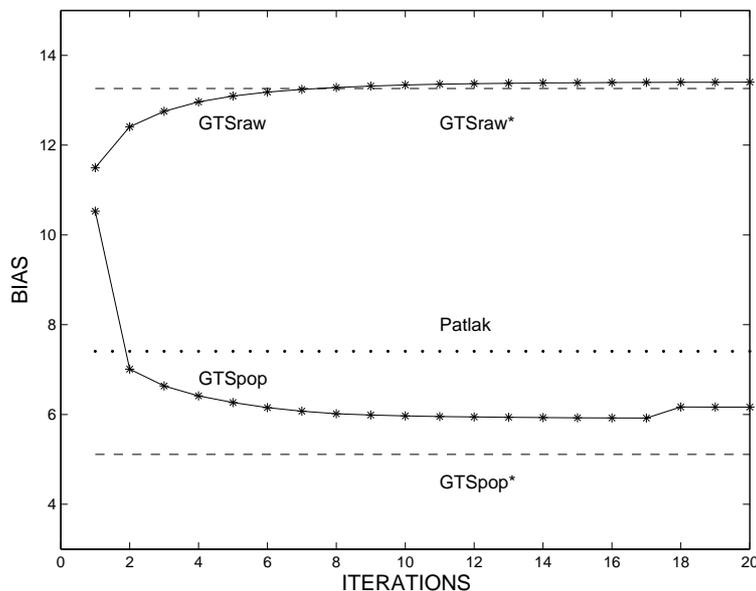


Figure 3.8: *BIAS for the simulation in the basal state is displayed. The dotted line refers to Patlak method, dashed lines correspond to GTS-based methods when a convergence criterion was used, and the solid lines represent performances of GTSpop and GTSraw as a function of the iterations.*

GTS-based approaches differed from Patlak method in terms of BIAS only slightly; in the basal state data set GTSpop yielded performances better than GTSraw, which gave rise to a slight increase of BIAS. Noticeable improvements were noticed instead in terms of RMSE. Considering for instance GTSpop\*, decreases from 49.1% to 39.7% (relative reduction of roughly 20%) in the basal data set and from 28.3% to 19.2% (relative reduction of roughly 30%) in the insulin data set were noticed. GTSpop always outperformed GTSraw, even though differences were typically slight.

Differences between GTSpop\* and GTSraw\* on one side and their iterative counterparts on the other were low, never higher than 3%.

A few iterations (3 to 6) were necessary to achieve convergence, with higher values for GTSraw because a more heterogeneous population clearly requires more iterations to stabilize.

### 3.7.5 Comparison with linear ridge regression

Figure (3.12) displays the irreversible rate constant for one simulated slice of the basal data set. The map obtained with Patlak method is on the upper left part, GTSpop map on the upper right, and linear ridge regression map -which was computed for completeness to allow a visual comparison- is located in the lower section. Notice the

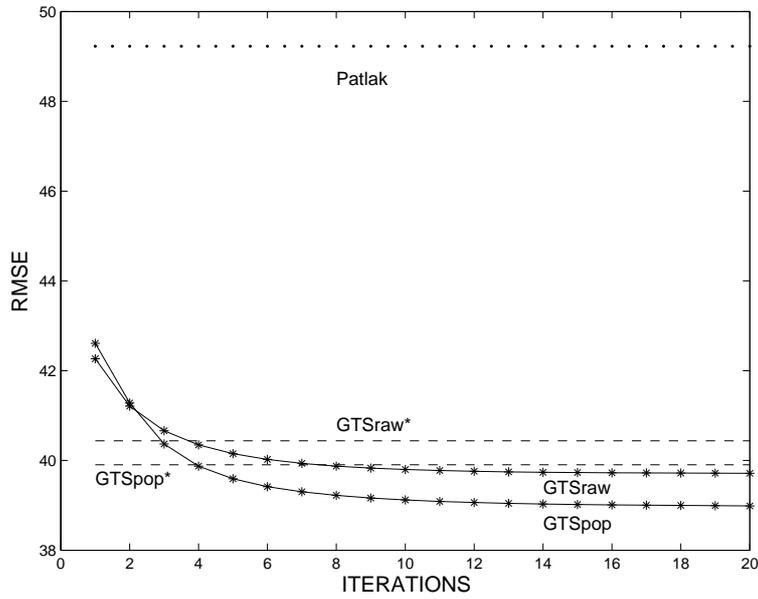


Figure 3.9: *RMSE* for the simulation in the basal state is displayed. The dotted line refers to *Paltak* method, dashed lines correspond to *GTS*-based methods when a convergence criterion was used, and the solid lines represent performances of *GTSpop* and *GTSraw* as a function of the iterations.

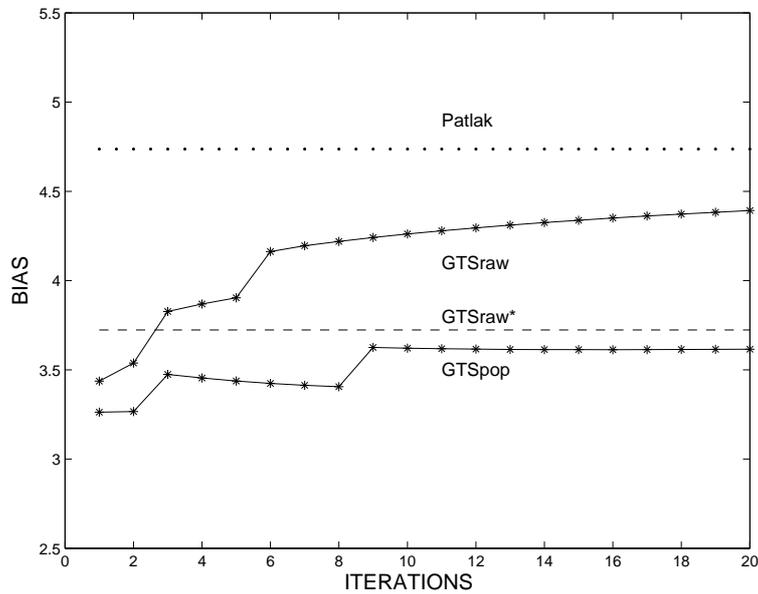


Figure 3.10: *BIAS* for the insulin data set is displayed. The dotted line refers to *Paltak* method, the dashed line corresponds to *GTS*-based methods when a convergence criterion was used (*GTSpop*\* gave a result equal to *GTSraw*\*), and the solid lines represent performances of *GTSpop* and *GTSraw* as a function of the iterations.

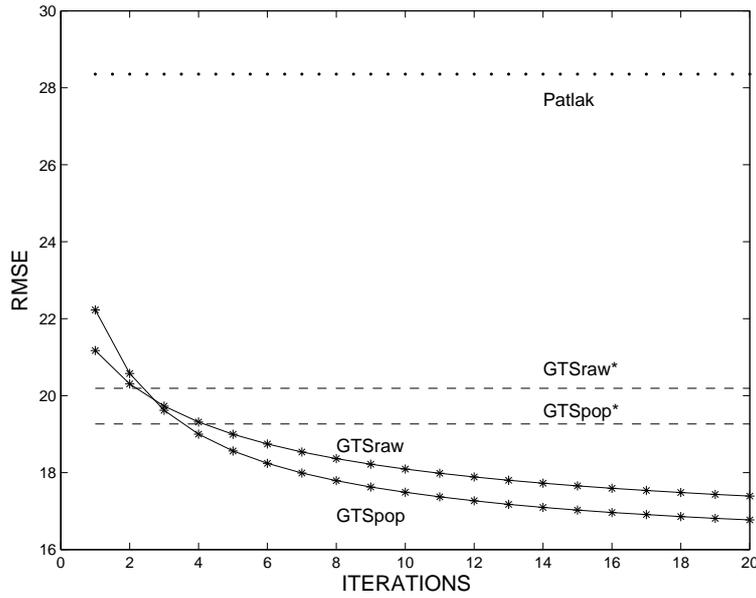


Figure 3.11: *RMSE for the insulin data set is displayed. The dotted line refers to Patlak method, dashed lines correspond to GTS-based methods when a convergence criterion was used, and the solid lines represent performances of GTSpop and GTSraw as a function of the iterations.*

improvement in terms of regularity brought by both GTS and ridge regression. Graphic differences for the simulated insulin data set were of lower amount, as original Patlak images were already of high quality.

Table (3.1) compares results obtained using GTS and linear ridge regression, in its generalized version. From it emerges a slight superiority of GTS in terms of both BIAS and RMSE. Both methods, however, left mean BIAS basically unchanged while reducing average RMSE.

	GTSpop*(basal)	GRRSC(basal)	GTSpop*(insulin)	GRRSC(insulin)
RMSE	39.7	43.9	19.2	24.5
BIAS	3.7	9.3	5.2	4.7

Table 3.1: *Comparison of performances of GTS and linear ridge regression on the simulated  $[^{18}F]$ FDG data sets*

### 3.8 Pixel level simulations: nonlinear case

In order to test performances of GTS when applied to different tracers and models, another simulated data set was analyzed. In particular, as in the previous paragraph a linear method was considered, the standard 2T-4K reversible model which is commonly used in receptorial studies was considered to study GTS in a non linear scenario and WNLLS was used as the method of comparison. The same simulated  $[^{11}C]$ WAY100,635

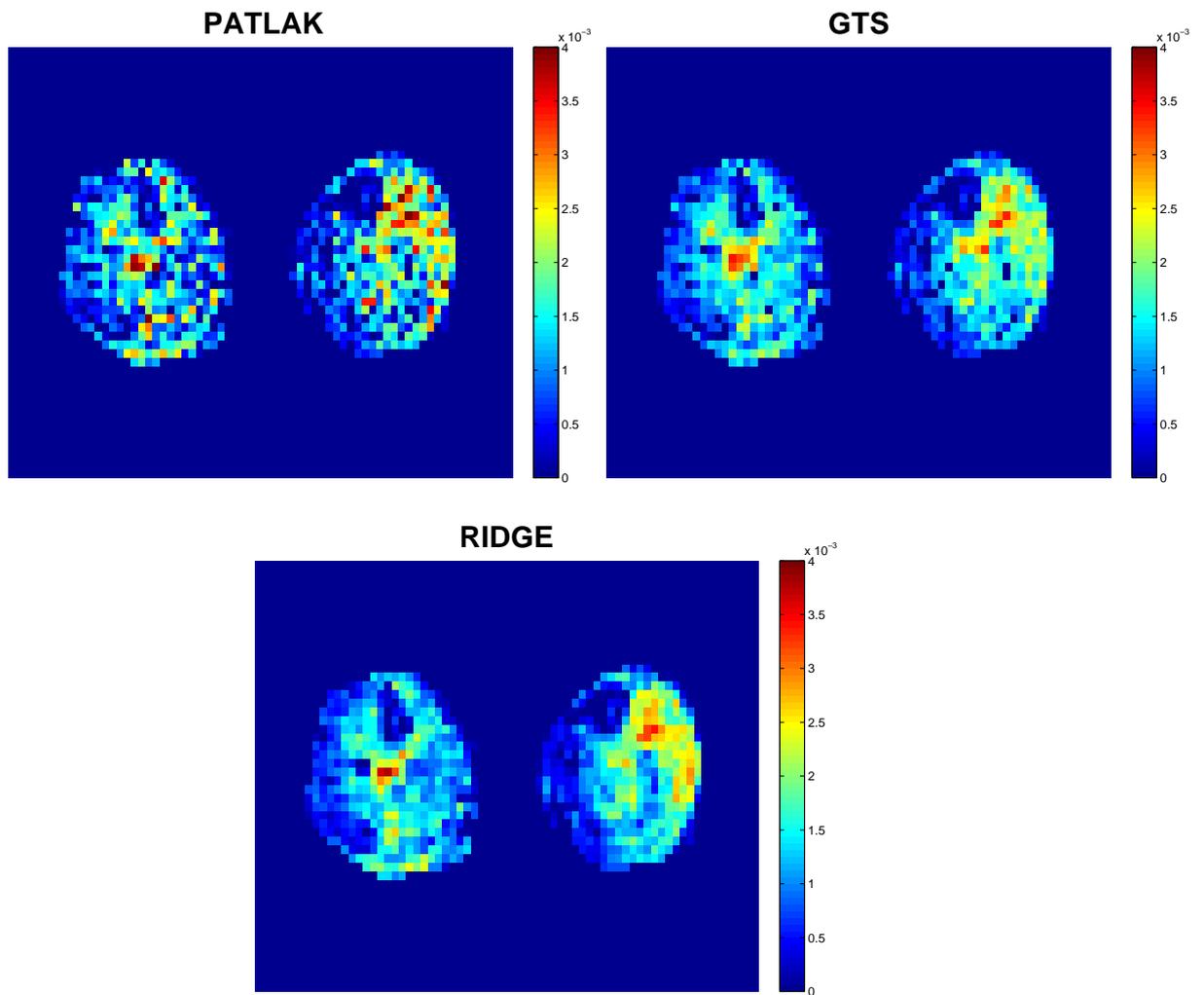


Figure 3.12: *The irreversible rate constant for one simulated slice of the basal data set, obtained respectively with Patlak method (left), GTSpop (right) and linear ridge regression (down), is displayed.*

data set employed for the evaluation of non linear ridge regression (see Chapter 2) was used also here; this allowed also a comparison between the performances of GTS and of non linear ridge regression algorithms. Besides considering GTSraw (all pixels together in the algorithm) GTSpop was used with the number of sub-populations  $n$  set first to 3 and then to 6 and 9 in order to test the dependence of results on the number of sub-populations used. For  $[^{11}\text{C}]\text{WAY100,635}$ , segmentation was performed separately for each kinetic parameter and was based on the preliminary estimates provided by WNLLS; no TAC-based clusterization was made in this case.

### 3.8.1 Results

A good independence of results on the iteration number was noticed. The robustness of GTS with respect to the number of iterations performed, and, consequently, with respect to the stop criterion, which was also found for the synthetic  $[^{18}\text{F}]\text{FDG}$  data, was therefore confirmed. Results presented here refer only to the case when a stop criterion was used. As for the  $[^{18}\text{F}]\text{FDG}$  data sets iterations were stopped when the relative variation of the parameter of interest was less than 1% for at least 90% of the pixels of the population under examination.

In addition, as displayed in tables (3.2) and (3.3), no remarkable difference was noticed between results obtained with the use of 3,6 and 9 populations. Results corresponding to  $n=3$ , however, were the best both from in terms of the RMSE and BIAS and when referring to GTSpop, from now on, the case  $n=3$  will be implicitly considered. Noticeable differences, on the other side, were found between GTSpop and GTSraw, and therefore results corresponding to both methods will be discussed.

	$K_1$	$k_2$	$k_3$	$k_4$
n=3	61.0	65.4	31.3	16.7
n=6	66.3	71.8	34.5	16.9
n=9	68.1	72.3	36.2	16.8

Table 3.2: Comparison of performances of GTSpop for different number of sub-populations ( $n=3,6,9$ ) on the simulated  $[^{11}\text{C}]\text{WAY100,635}$  data set: average RMSE is displayed.

	$K_1$	$k_2$	$k_3$	$k_4$
n=3	15.0	9.8	-6.6	1.2
n=6	18.7	16.9	-2.7	1.7
n=9	19.8	17.6	-1.3	1.9

Table 3.3: Comparison of performances of GTSpop for different number of sub-populations ( $n=3,6,9$ ) on the simulated  $[^{11}\text{C}]\text{WAY100,635}$  data set: average BIAS is displayed.

Figures (3.13) and (3.14) show BIAS and RMSE for the four kinetic parameters  $K_1, k_2, k_3$  and  $k_4$ . Black bars refer to standard WNLLS, gray to GTSraw and white to GTSpop, with a subdivision of the pixel population into 3 clusters. It emerged that:

- Whereas GTSpop gave always satisfactory results in terms of BIAS, GTSraw yielded poor performances for  $k_2$  and  $k_3$ . The reason is that, when grouping non homogeneous pixels as the ones of a whole slice, the final estimates of those pixels whose initial estimates are characterized by a large variance may deviate towards the estimates of those pixels with an initial low variance, potentially giving rise to a bias that may be more or less significant according to the homogeneity of the initial image.
- Differently from what found for the  $[^{18}\text{F}]\text{FDG}$  data sets, GTSraw outperformed GTSpop in terms of RMSE: for  $K_1$  and  $k_2$  RMSE reduction when GTSraw was employed was relevant.
- No difference between GTS-based algorithms and WNLLS was noticed for  $k_4$ , which was the best estimated parameter. This confirms what found for the analysis at ROI level: population approaches change little or nothing if original estimates are of good quality, i.e. have an associated small variance.

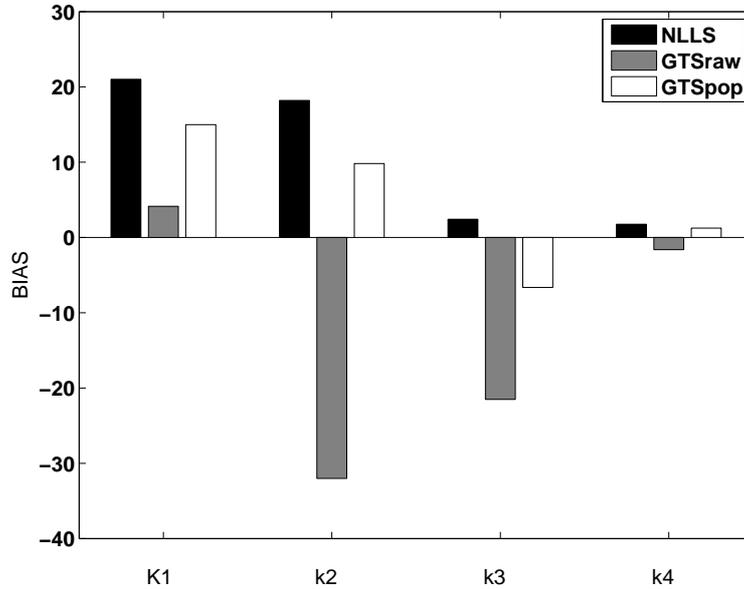


Figure 3.13: *BIAS for  $[^{11}\text{C}]\text{WAY100,635}$  data is displayed. Black bars refer to standard WNLLS, gray to GTSraw and white to GTSpop, with a subdivision of the population into 3 clusters.*

### 3.8.2 Comparison with non linear ridge regression

Although GTS always yielded improvements in terms of RMSE, the results presented in the previous paragraphs can not be considered totally satisfactory. GTSraw provided the largest reduction of RMSE, but, with approach, 2 of the 4 parameters were, on average, heavily underestimated. GTSpop, on the other side, didn't increase BIAS but reduction of RMSE were low for  $K_1$  and  $k_2$ .

If for the  $[^{18}\text{F}]\text{FDG}$  data set, as previously illustrated, GTS emerged as slightly superior to ridge regression, for  $[^{11}\text{C}]\text{WAY100,635}$  data the situation is inverted. This

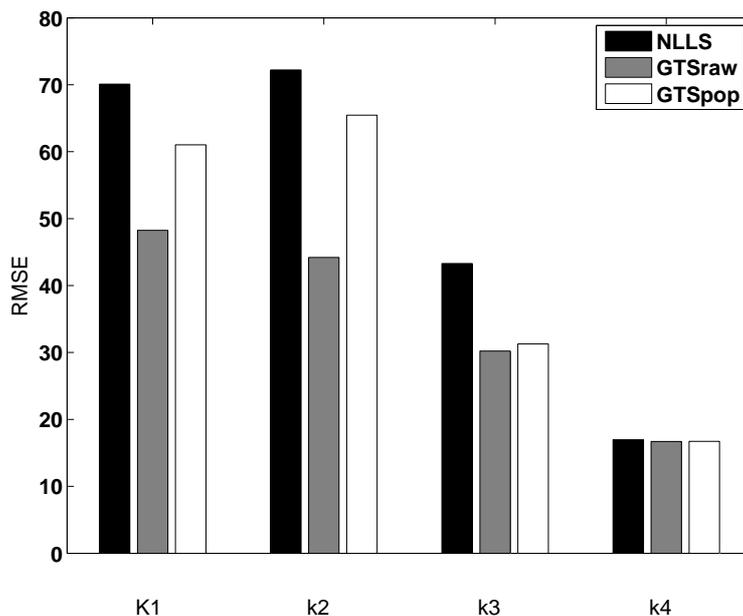


Figure 3.14: *RMSE for  $[^{11}C]WAY100,635$  data is displayed. Black bars refer to standard WNLLS, gray to GTSraw and white to GTSpop with a subdivision of the population into 3 clusters.*

can be seen from tables (3.4) and (3.5) which summarize results already presented, displaying the usual RMSE and BIAS for both GTS-based approaches and for NLRR3 with CRM to compute priors, which was selected as the best approach in the non linear ridge regression section. Differences are particularly clear for  $K_1$  and  $k_2$  RMSE.

Figure (3.15) displays, using the same color scale for the different figures, the 'true'  $k_3$  parametric map (upper left), and those obtained for one synthetic slice using GTSraw (upper right), GTSpop (lower left) and NLLRR3 (lower right);  $k_3$  was chosen because it illustrates particularly well the differences of final parametric maps. It emerges that non linear ridge regression provided the most appealing map. GTSraw map is too smooth and high  $k_3$  estimates are shrunk towards the global mean, thus causing a loss of physiological information and the BIAS previously described. GTSpop map resemble the "true" one more than GTSraw, but it has the drawback that pixels seem to be categorized into 3 groups, corresponding to the 3 sub-populations into which they were segmented. In other words each pixel estimate seems to be determined exclusively by the mean of the corresponding group, with a consequent loss of possibly important details.

	$K_1$	$k_2$	$k_3$	$k_4$
GTSraw	48.3	44.2	30.2	16.7
GTSpop	61.0	65.4	31.3	16.7
NLRR3	34.4	37.1	32.5	35.1

Table 3.4: *Comparison of performances of GTS (raw and pop) and non linear ridge regression (version 3 with CRM to compute priors) on the simulated  $[^{11}C]WAY100,635$  data set: average RMSE is displayed.*

	$K_1$	$k_2$	$k_3$	$k_4$
GTSraw	4.1	-32.0	-21.5	-1.6
GTSpop	15.0	9.8	-6.6	1.2
NLRR3	-19.2	-12.3	1.1	-11.4

Table 3.5: Comparison of performances of GTS (raw and pop) and non linear ridge regression (version 3 with CRM to compute priors) on the simulated [ $^{11}\text{C}$ ]WAY100,635 data set: average BIAS is displayed.

### 3.9 Recovery of information in unreliable scans

It has been shown that population approaches and, more generally, Bayesian methods are not useful when working at ROI level because in this scenario the quality and accuracy of parametric estimates is typically of good level due to the low noise of ROI TACs. In this paragraph, however, an interesting application of ridge regression and GTS at region-of-interest level is presented.

It may happen for a variety of reasons that a PET scan must be interrupted before the end, for instance because the patient suddenly starts to suffer from claustrophobia or needs to be helped due to an unexpected event (cough, tremor...). This is clearly more likely to happen for longer scans and therefore for tracers labelled with  $^{18}\text{F}$ , and to a minor extent, for tracers labelled with  $^{11}\text{C}$ . This problem does not arise for instance for [ $^{15}\text{O}$ ]H $_2$ O, whose scans last typically 5-8 minutes because of the very short half-life of  $^{15}\text{O}$ . It may also happen, more generally, that the last part of the scan is totally unreliable; imagine for instance that the patient has moved too much so that no re-alignment algorithm is able to correct for this sudden and unexpected movement. These conditions are not infrequent especially considering that many scans are performed on elderly subjects and that many of them are selected in that they show some neurologic pathology, which makes their motion and behavior less controllable.

Are these scans to be totally discarded from the quantitative analysis? Or is it possible to recover, at least partially, some of the information making use of the fact that these "problematic" patients show a certain degree of homogeneity with other patients for whom there is a complete and reliable scan available?

To investigate this problem a simulated data set was generated. [ $^{18}\text{F}$ ]FDG was selected and the real [ $^{18}\text{F}$ ]FDG data set previously described was considered, with 8 healthy subjects who underwent a 90-minutes scan of their legs used to construct the simulated data set. It was stated several times before that a five-constant three-compartment irreversible model (5K) provides the most appropriate description of the kinetics of [ $^{18}\text{F}$ ]FDG in the skeletal muscle and this model was therefore employed here. A ROI was placed over the soleus muscles of each subject using his/her MRI scan and in this way the corresponding time activity curve was extracted from the image.

It was clearly necessary to simulate a distribution for the parameter vectors which resembled as much as possible the real situation, otherwise results would not have been realistic, causing underestimation or overestimation of the potential benefits of Bayesian approaches. To achieve this goal the 5 kinetic parameters  $-V_b$  was set to 0- were estimated in the usual way using the measured plasmatic tracer concentration and standard WNLLS. Mean  $\mu$  and sample covariance  $\Omega$  of the parameter vector  $\mathbf{K}$  were then com-

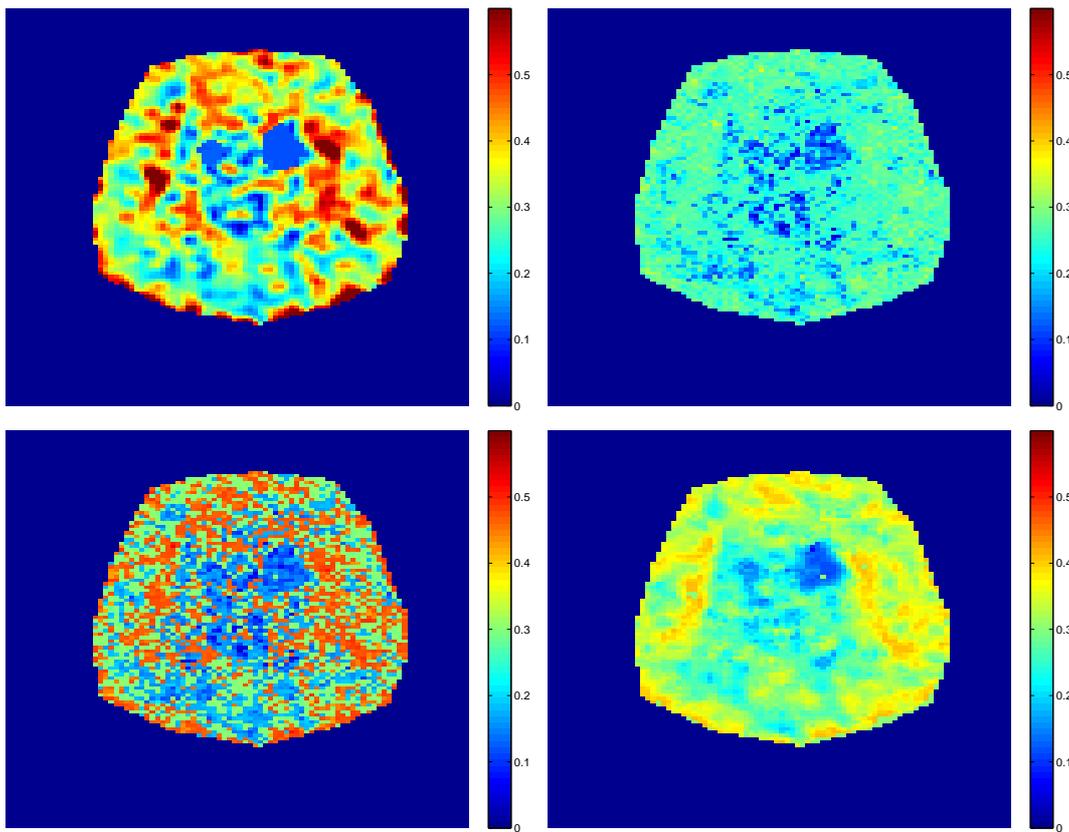


Figure 3.15: The "true"  $k_3$  parametric map (upper left), and those obtained for one simulated slice using *GTSraw* (upper right), *GTSpop* (lower left) and *NLLRR3* (lower right) are displayed, using the same color scale for all the figures.

puted, with  $\Omega$  set to  $\frac{1}{8}\Sigma(K_i - \mu)(K_i - \mu)^T$  and  $K_i$  equal to the parameter vector of the  $i$ -th subject. A multivariate Gaussian distribution with mean  $\mu$  and covariance  $\Omega$  was then employed to generate 1000 simulated parameter vectors (or, alternatively, 1000 simulated subjects). A randomly selected real plasmatic curve was used to create the corresponding 1000 noise free TACs using the parameters just generated and the same time protocol of the real scan. Noise was added as previously explained by summing for each subject and frame Gaussian noise with zero mean and variance equal to

$$\sigma^2 \frac{C_i}{\Delta_i}$$

with  $C_i$  and  $\Delta_i$  indicating respectively the concentration of the noise free TAC at frame  $i$  and the duration of the  $i$ -th frame. The proportionality constant  $\sigma^2$  was set equal to the mean of  $\sigma_i^2$  among the subjects, with  $\sigma_i^2$  computed from the WNLLS fit of the soleus TAC as in (1.7).

To simulate the scenario of reduced scan times, 4 sets of kinetic parameters were estimated for each subject assuming as available the full scan (90 minutes, set 1), 60 minutes (set 2), 41 minutes (set 3) and 33 minutes (set 4). As in real situations the number of subjects undergoing the same study (same scanner, scan protocol, tracer) is not high, the 1000 simulated subjects were divided into groups made of  $N$  individuals.  $N$  was initially set to 10 and later on, for GTS only, to 4,6,8,20 to test the dependence of results on this parameter.

For each of the 3 reduced scan length, each of the populations of  $N$  individuals was considered as made of 1 subject with a reduced scan length and the remaining ( $N-1$ ) with the full scan length; this was repeated for all subjects. These populations were used in the construction of the prior for both ridge regression and GTS as explained later. Results were evaluated as usual in terms of BIAS and RMSE: those obtained applying to every population ridge regression and GTS were compared to the ones derived using for each subject only data from the reduced scan, with no use of the information available from the other subjects.

### 3.9.1 Adaptation of GTS to the problem

GTS was modified to keep into account the specificity of the problem. It wouldn't make sense to employ the algorithm in its original form as it was previously presented, because in that case all parameter estimates would be modified, even the ones belonging to subjects for whom the complete scan is available. The goal is, on the other hand, to modify and potentially improve the estimates of the subject(s) for whom only the first part of the scan is reliable. The 2 basic GTS equations [3.8] and [3.9] for the computation of the parameter vector  $p_i^k$  and corresponding covariance matrix  $\Sigma_i^k$  of individual  $i$  at iteration  $k$ , which are presented here

$$\begin{aligned} & [(\Sigma_i^0)^{-1} + (\Sigma_{pop}^k)^{-1}]^{-1} [(\Sigma_i^0)^{-1} p_i^0 + (\Sigma_{pop}^k)^{-1} p_{pop}^k] \\ & [(\Sigma_i^0)^{-1} + (\Sigma_{pop}^k)^{-1}]^{-1} \end{aligned}$$

were still used. This time, however, only the parameter vector and covariance matrix corresponding to the subject with the reduced scan were modified, whilst the other ( $N-1$ ) were kept unchanged. In this way the only variations in the mean parameter vector  $p_{pop}^k$  and covariance matrix  $\Sigma_{pop}^k$  were clearly due only to modifications of the parameters

associated to the subject with the reduced scan. As to the stop criterion, iterations were ended when the relative difference, between the current iteration and the previous one of all kinetic parameters was lower than 1% for the subject under examination, as validated previously.

### 3.9.2 Adaptation of ridge regression to the problem

As stated in the previous subsection, the goal in this scenario is to modify only the estimates of the subject(s) for whom the first part of the scan is available and it is therefore obvious to apply ridge regression only to them, consistently with what said for GTS. The standard equation for GRRSC (2.4), recalled here,

$$h_i = \frac{\sigma^2}{(p_i - p_{i0})^2} \quad i = 1, 2, \dots, M$$

was employed here as well to compute H so that final estimates were estimated minimizing (2.18)

$$[y - h(p)]'W[y - h(p)] + (p - p_0)'H(p - p_0)$$

The only difference was that, instead of considering the voxels belonging to the 3x3 window centered in the voxel of interest to compute both the prior and the weights in the ridge matrix, in this context the *real subjects'* estimates were employed for both purposes. In other words the 3x3 group of pixels employed in classic ridge regression was substituted, here, by the N-individual population. The prior  $p_0$  and  $\sigma^2$ , therefore, were simply computed averaging the estimates of the subjects belonging to the population made of N individuals to which the subject with the reduced scan belongs to.

### 3.9.3 Results

Results were presented using the macroparameter

$$K_i = \frac{K_1 k_3 k_5}{k_2 k_4 + k_2 k_5 + k_3 k_5}$$

to compare "real" and estimated values.  $K_i$ .

Figure (3.16) displays average BIAS and RMSE for standard WNLLS (black), GRRSC (gray) and GTS (white) with the number N of individuals set to 10. Improvements were evident in terms of both indices and they clearly increased with decreasing scan time, when the initial estimates were poorer and the Bayesian paradigm was more helpful in improving them. The negative BIAS present for short scan times, in particular, was remarkably reduced by both approaches. GTS provided better performances with respect to GRRSC, but the difference was slight.

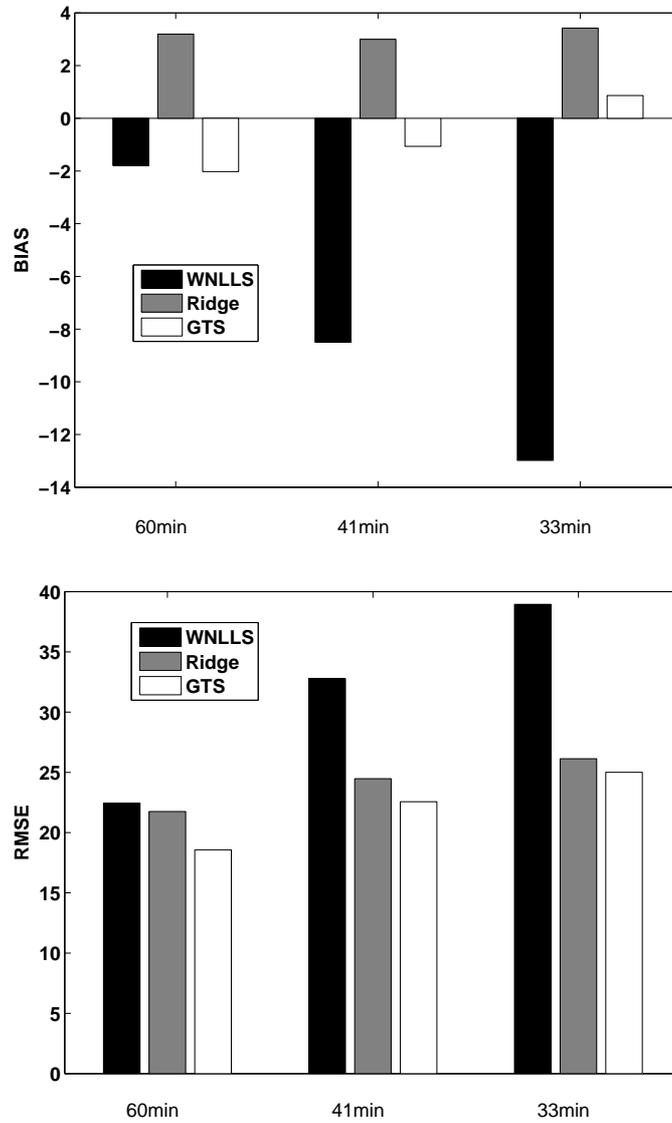


Figure 3.16: *BIAS* and *RMSE* averaged over the 1000 simulated subjects are displayed for standard *WNLIS* (black), *GRRSC* (gray) and *GTS* (white) with  $N=10$  and 3 different reduced scan times.

No improvements took place when N was equal to 4; in this case the number of individuals was too low. With N set to 6 improvements were already noticed according to the principle, previously illustrated, that as long as homogeneity is preserved, the higher the number of individuals is, the higher improvements are. Performances therefore got better for increasing values of N, but the amount was so slight that differences between the case N=6 and N=20 were negligible.

A reduced scan length of 75 minutes was also considered, but in this case WNLLS provided the best results; it is clear that, when only few frames (2-4) are missing, the application of standard WNLLS yields the best results because the scan time is still sufficient to get estimates of good quality.

Some words caution on the idea of "homogeneity" are in order: even though this concept hasn't been quantitatively characterized it is clear, for instance, that if the scanned subjects comprise ill and healthy individuals, or young and elderly subjects, the application of these algorithms may be problematic. The use of the suggested approaches is therefore advisable only when the subjects analyzed form an homogeneous group at least from the point of view of age and health condition.

In general Bayesian algorithms are useless at ROI level because estimates are typically more than satisfactory. This paragraph, however, has shown an interesting application of ridge regression and population techniques in this scenario. The outcome of the analysis was that, as long as there is a certain homogeneity between the scanned subjects and as long as their number is not too low, ridge regression or GTS can be successfully employed to provide estimates that are better than the ones that would be obtained using only the data from the scan of reduced length. Improvements were clearly more relevant for shorter scans and didn't show much dependence on the number of subjects employed for the application of Bayesian algorithms.

### 3.10 Discussion

The simulations presented in the previous paragraphs didn't allow to conclude unequivocally that ridge regression is better than GTS, or viceversa. Some ideas can however be inferred;

- While BIAS didn't not show a clear trend, there was an unequivocal *reduction* of RMSE when ridge regression or GTS were used with respect to standard WNLLS. This reduction ranged from slight to very large; in no case there was an increase of RMSE.
- While ridge regression gave, in general, improvements that were relatively constant with respect to the parameter and tracer analyzed, GTS yielded performances that were less stable, its improvements ranging from high to slight.
- In general GTSpop is advised over GTSraw because the use of the latter may cause an unacceptable loss of resolution and anatomical details of the final parametric maps due to the shrinking of "high" and "low" estimates towards the global mean value. This can not happen with GTSpop due to preliminary segmentation; what may happen with GTSpop, however, is that there is shrinking of estimates towards the mean of each subpopulation. These effects were not evident at all for the [<sup>18</sup>F]FDG simulation but noticeable for [<sup>11</sup>C]WAY100,635 data.

GTS, differently from RR, has two user-dependent settings which can strongly influence final results. Performances of GTS didn't show to be significantly dependent on the method employed to stop iterations; in any case, the criterion to end when the relative difference of the parameter of interest was lower than 1% for at least 90% of the individuals in the considered population proved to be quite robust in every situation tested.

Results, on the other side, appeared to be more dependent on the segmentation technique employed. A segmentation of each slice of the image into a small number of clusters performed on the basis of each parameter separately, is a good solution which keeps computational time low and allows to perform GTS on more homogeneous populations. Fortunately, it was found both in [ $^{18}\text{F}$ ]FDG and in [ $^{11}\text{C}$ ]WAY100,635 simulations that BIAS and RMSE were not significantly dependent on the number of sub-populations employed to segment the image. Alternatively, if the image is normalized to a stereotaxic space, a predefined atlas (see chapter 4) can be employed to create sub-populations made of pixels belonging to the same anatomical ROI. This last solution eliminates the need for the choice of the number of sub-populations to be employed, and integrates anatomical information into parametric maps, which are both very positive facts.

The comparison between RR and GTS on the basis of RMSE of the two simulated data sets ([ $^{18}\text{F}$ ]FDG and [ $^{11}\text{C}$ ]WAY100,635) showed that sometimes GTS performed slightly better, while sometimes the contrary happened; in any case differences were never relevant.

## Chapter 4

# Applications to a clinical dataset: $[^{11}\text{C}]$ -(R)-PK11195

In the previous chapters ridge regression and population approaches have been applied to simulated data sets only in order to assess their performances when "real" kinetic parameters were known. The conclusion was that these methodologies allow to obtain better results with a little increase of computational time. The amount of the improvement may be low or high depending on the level of noise and on the complexity of the kinetic model employed. The simple general rule is that, within certain limits, the poorer the initial estimates are, the higher the improvement is.

In this chapter application of ridge regression and GTS to real clinical data sets is presented. It is not straightforward to compare different methodologies when applied to real data sets. The criteria chosen to compare parametric maps generated with different methods were

- simple visual assessment
- correlation between the kinetic parameters averaged over all pixels of a given ROI and the same parameters estimated from the TAC of that ROI
- ability to differentiate between healthy and ill subjects or pixels measured by sensitivity and specificity

### 4.1 Introduction to $[^{11}\text{C}]$ -(R)-PK11195

Most of the chapter will deal with results concerning the tracer  $[^{11}\text{C}]$ -(R)-PK11195 which therefore deserves a brief introduction.

$[^{11}\text{C}]$ -(R)-PK11195 is a selective ligand for the peripheral benzodiazepine receptor, or binding site (PBBS). PBBS is a protein abundant in peripheral organs, particularly in adrenal glands and kidney but also in heart and lungs ([29]). In the normal brain PBBS is expressed in the endothelium and smooth muscles of intra-cerebral arteries and in the endothelial walls of the veins.

What is important, in relation to  $[^{11}\text{C}]$ -(R)-PK11195, it is that PBBS is also expressed in activated microglia ([30]). Microglia are the intrinsic *cerebral immune system* and they are involved in the response to traumatic, inflammatory, degenerative and

neoplastic disease. Simplifying the complicated and still not totally understood underlying biological mechanisms, microglia express PBBS not in resting state but only when in activated state. [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195, therefore, binding to PBBS, most of which are expressed by activated microglia, serves as a very generic indicator of neuronal damage in the central nervous system. ([31]) .

Quantification of [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 PET studies has been approached in the great majority of cases by the application of the simplified reference tissue model (SRTM) ([32], [33]). Only a recent work ([34]) has studied [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 modelling when the plasmatic input is measured, but all the other works in the last 10 years employed SRTM, which was therefore considered here. In the past either the cerebellum or clustering algorithm were employed to compute the reference TAC, but, in both cases, problems were found because of the inclusion of voxels associated to regions containing PBBS, which are expressed in the whole brain. In a recent work it has been demonstrated that a supervised selection algorithm which compares the TAC of each pixel with a database of tissue kinetics (normal gray and white matter, vascular, muscle, skull and pathological tissue with high active microglia density) is able to define a reference gray matter tissue devoid of specific PBBS binding and is therefore the best methodology to extract the reference region ([35]). To extract the reference TACs employed in this thesis, the supervised algorithm proposed in [35] was always employed.

All the [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 studies analyzed here were performed on an ECAT EX-ACT 3D (CTI/Siemens) PET camera with 23.4 cm axial field of view, 95 transaxial planes, spatial resolution of 4.8 mm full width at half maximum (FWHM) (transaxial) and 5.6 mm FWHM (axial). Three-dimensional sinograms of emission data were then acquired over 60 minutes as 18 time frames (30s background frame, 1x15s-frame, 1x5sframe, 1x10s-frame, 1x30s-frame, 4x60s-frames, 7x300s-frames and 2x600s-frames). Volumetric T1 weighted MRIs images were obtained on a 1.0 TeslaPicker HPQ scanner (Picker, Cleveland, OH) at the Robert Steiner MR Unit, Hammersmith Hospital, London.

The data set used consisted of 9 patients suffering from Huntington's disease (HD) and 12 age-matched normal controls (NC), described more deeply in [36]. The MRI of each patient was co-registered with his-her PET summed image using Statistical Parametric Mapping SPM5 (Functional Imaging Laboratory, Wellcome Department of Imaging Neuroscience, UCL, London). The co-registered MRIs were then normalized to the MNI/ICBM512 space (Montreal Neurological Institute stereotaxic space) and the parameters obtained from the normalization process for each subject were applied to each frame of the PET image to obtain a *normalized* dynamic PET image. Region of interest TAC extraction was performed on the normalized PET images using Hammersmith maximum probability atlas ([37]). This atlas assigns to the main cerebral regions (e.g. cingulate gyrus, cerebellum, thalamus) the set of voxels which belong to each region: every ROI is divided between its left and right part for a total of 66 ROIs. This assignation is made on the MNI stereotaxic space; in other words the atlas is a matrix with dimensions [91,109,91], with only a subgroup of the 91x109x91 voxels assigned to a specific ROI. The normalization process, therefore, was necessary in order to be able to apply the atlas to the images; in this way TAC extraction could be performed automatically and without the need for specific anatomical competence.

## 4.2 Inclusion of the vascular components into the SRTM

### 4.2.1 Standard SRTM equation

The classical SRTM equation

$$C_{Tar}(t) = RC_{Ref}(t) + (k_2 - \frac{k_2 R}{1 + BP}) \int_0^t C_{Ref}(x) e^{-\frac{k_2}{1+BP}(t-x)} dx \quad (4.1)$$

with  $C_{Tar}$  and  $C_{Ref}$  indicating respectively the time course of the target and reference regions, is used to estimate the ratio  $R$  of delivery constants  $K_1^{Tar}/K_1^{Ref}$ , the rate of outflux of the tracer  $k_2$  and the binding potential  $BP$ . Equation (4.1) is non-linear in the parameter vector  $[R, k_2, BP]$ ; the well-known "basis function method" (BFM) ([38]) is typically applied to it in order to reduce its computational cost. The basis function method requires to fix a value for the parameter  $\vartheta = \frac{k_2}{1+BP}$ , to compute the convolution integral

$$\int_0^t C_{Ref}(x) e^{-\frac{k_2}{1+BP}(t-x)} dx$$

which appears on the right side of (4.1) and then to solve the estimation problem expressed by (4.1), problem which becomes linear in the parameter vector [

$$p_1, p_2] = [R, k_2 - \frac{k_2 R}{1 + BP}$$

] once  $\vartheta$  is fixed. The procedure is repeated for each value of  $\vartheta$  belonging to a predefined grid and for each of these values the problem, linear in the unknown vector  $[p_1, p_2]$ , is solved. through the usual expression

$$p = (X'WX)^{-1}(X'Wy)$$

with  $X$   $N \times 2$  matrix with  $C_{Ref}(t)$  and  $\int_0^t C_{Ref}(x) e^{-\frac{k_2}{1+BP}(t-x)} dx$  in its columns,  $y$  containing the  $N$  measured data, and  $W$   $N \times N$  diagonal matrix containing suitable weights. The weights for all the analysis performed for  $[^{11}C]-(\text{R})\text{-PK11195}$  were set equal to the scan interval lengths  $\Delta_i$ . Plausible values for  $\vartheta_{min}$  and  $\vartheta_{max}$  are typically determined from the knowledge of physiological values for  $k_2$  and  $BP$  for the tracer under examination and then a grid of 50 or more values is used. At the end the value of  $\vartheta^*$  which gave rise to the smallest weighted sum of residuals is retained, together with  $p_1^*$  and  $p_2^*$ , the solution of the linear problem corresponding to  $\vartheta^*$ . The estimate of  $R$ ,  $k_2$  and  $BP$  is then easily achieved by suitably combining the values of  $\vartheta^*$ ,  $p_1^*$  and  $p_2^*$  according to

$$\begin{aligned} R &= p_1^* \\ k_2 &= p_2^* + \vartheta^* R \\ BP &= \frac{k_2}{\vartheta^*} - 1 \end{aligned}$$

To obtain a model which is completely linear one has simply to integrate the standard differential equations regulating the SRTM model in order to make the first derivatives disappear thus obtaining:

$$C_{Tar}(t) = RC_{Ref}(t) + k_2 \int_0^t C_{Ref}(x) dx - \vartheta \int_0^t C_{Tar}(x) dx \quad (4.2)$$

with  $\vartheta = \frac{k_2}{1+BP}$  as above. This approach is faster than the previous one because at each pixel one and only one linear problem has to be solved whereas a larger number is needed (at least greater than 20-30 depending on the settings) when applying SRTM with a basis function approach. The linearized equation (4.2), unfortunately, has a well-known drawback in that, when applied to noisy TACs such those at pixel levels, it provides biased estimates of parameters. Using the usual notation  $y=Xp+e$  to express the linear problem, the NxM known matrix X is not, in this case, noise free in that one of its column contains the integral of the noisy pixel TAC. The usual expression

$$p = (X'WX)^{-1}(X'Wy)$$

which is used to compute the parameter estimates yields biased estimates as a consequence of the correlation between the measurement error  $e$  and the error of the column in X. The noisier the TAC is, the higher the bias is: the linearized version of SRTM was not employed to avoid such phenomenon.

## 4.2.2 Modifications of the standard SRTM equation

An implicit assumption of the SRTM is that  $V_B^T = V_B^R = 0$ , with  $V_B$  indicating the vascular fraction within the region or pixel of interest (the primes T and R stand for target and reference respectively). This simplification may be totally legitimate but, depending on the tracer, it may in certain situations be inappropriate. For instance, when analyzing healthy and ill subjects together, the implicit assumption of a constant vascular fraction in healthy conditions and in disease may not be correct. In the case of [<sup>11</sup>C]-(*R*)-PK11195 this would be equivalent to postulate an identical PBBS expression in the vascular components in normal and ill conditions, which may be questionable. To modify the standard SRTM to keep this fact into account (this new model will be referred to as SRTMV), (4.1) can be easily transformed as follows. Consider the equations

$$C_{Tar}^{Meas}(t) = (1 - V_B^T)C_{Tar}^{True}(t) + V_B^T C_B(t) \quad (4.3)$$

and

$$C_{Ref}^{Meas}(t) = (1 - V_B^R)C_{Ref}^{True}(t) + V_B^R C_B(t) \quad (4.4)$$

which express the measured concentrations in target and reference regions as a function of the "true" (unknown) concentrations and of the vascular term, with  $C_B(t)$  indicating the whole blood tracer time course. Solving (4.3) and (4.4) for  $C_{Tar}^{True}$  and  $C_{Ref}^{True}$  and inserting the results into (4.1) one obtains

$$C_{Tar}(t) = V_B^T C_B(t) + \frac{1 - V_B^T}{1 - V_B^R} \{ R(C_{Ref}(t) - V_B^R C_B(t)) + (k_2 - \frac{k_2 R}{1 + BP}) \int_0^t (C_{Ref}(x) - V_B^R C_B(x)) e^{-\frac{k_2}{1+BP}(t-x)} dx \} \quad (4.5)$$

The parameters to be estimated in this new framework are  $R$ ,  $k_2$  and  $BP$  as before, plus  $V_B^T$  and  $V_B^R$ , the blood volume fractions in the target and reference region respectively. (4.5) can be used to estimated the parameters through WNLLS: with this approach, however,  $V_B^R$  is attributed for every target region or pixel of interest a different value whereas it does have a fixed, albeit unknown, value. Besides, the application to WNLLS with 5 unknown parameters often yielded unreliable values for  $V_B^T$  and  $V_B^R$ .

As suggested in [39], where (4.5) was first proposed, fixing  $V_B^R$  to a reasonable value makes the estimation of the remaining four parameters more stable. According to [39] a standard value of 0.05 corresponding to 5% was used here. In [40] it is shown that, although the value of  $V_B^R$  is monotonically dependent on the value fixed for  $V_B^T$ , the other 3 parameters are almost independent on  $V_B^R$  and this is particularly important since the analysis presented here was centered on  $BP$ s.

(4.5) can be used to estimate the other parameters as with SRTM. Once  $V_B^R$  is fixed, the estimation problem expressed by (4.5) can be solved by means of a basis function method, with

$$\vartheta = \frac{k_2}{1 + BP}$$

as above and

$$[p_1, p_2, p_3] = [V_B^T, R(1 - V_B^T), (1 - V_B^T)(k_2 - \frac{k_2 R}{1 + BP})]$$

It easily follows that

$$\begin{aligned} V_B^T &= p_1 \\ R &= \frac{p_2}{1 - V_B^T} \\ k_2 &= \frac{p_3 + R\vartheta}{1 - V_B^T} \\ BP &= \frac{k_2}{\vartheta} - 1 \end{aligned}$$

which must be employed for the computation of  $R$ ,  $k_2$ ,  $V_B^T$  and  $BP$ .

A grid of 50 equally spaced values between 0 and 0.5 was employed  $\vartheta$ . The use of a higher number of values (100) was tested but didn't yield noticeable differences in terms of the resulting parametric maps and was therefore discarded for computational reasons.

### 4.2.3 Estimation of the whole blood TAC

The whole blood time activity curve  $C_B(t)$  was not available from invasive measurements and had to be estimated from the image itself. For this purpose the following algorithm was employed. In the whole brain the 10 pixels with the maximum value in the first 5 frames -which are the ones where the blood tracer concentration has its peak- were selected. The mean of the selected pixel TACs was used as  $C_B(t)$ . This simple procedure is remarkably faster than any approach based on cluster or on manual selection of the vascular components, besides being user's independent. The choice to select 10 pixels (and not more than 10 or less) was based on a compromise between the need to obtain a sharp peak for  $C_B(t)$  -which would require to use the smallest possible number of pixels- and the need to obtain reliable and regular curves. Kinetic parameters for each subject were also computed estimating  $C_B(t)$  using the 5 and 20 pixels with the highest values in the first frames to test the robustness of the algorithm with respect to the number of pixels employed for the estimation of  $C_B(t)$ .

The procedure used for the extraction of the blood component from images is far less sophisticated than others that are available for a more accurate estimation of the blood tracer concentration time course. They make either use of cluster analysis ([41])

or of accurate definition of the vascular region with appropriate correction for spillover and partial volume ([42]) or of resolution of non-linear problems to determine the arterial concentration for each subject ([43]). However, these procedures often requires one or more blood sample making them slightly invasive. Besides, as the estimated arterial input function was used in those works as the forcing function of the system, its estimate had to be absolutely reliable. In SRTMV, the forcing function is  $C_{Ref}(t)$  and parameter estimates are much less dependent on  $C_B(t)$ . Hence, in this context a reliable approximation of the whole blood tracer concentration only was necessary and the simple yet effective algorithm described was selected due to its computational speed and simplicity. This approach, besides, displayed a remarkable robustness with respect to the number of pixels employed for the blood TAC computation when  $V_B^T$  and BPs obtained with different  $C_B(t)$  were compared ([40]).

### 4.3 Application of ridge regression

#### 4.3.1 Non linear version

SRTM and SRTMV are non-linear models and the simplest idea, therefore, is to apply non-linear ridge regression. In Chapter 2 it was shown on simulated data how NLRR3 provided the optimal results at the lowest computational cost among the non linear ridge regression methods. This version of NLRR, with the use of the component representation model CRM to compute priors, was therefore implemented.

The linearization of (4.5) -i.e. the computation of the sensitivity matrix S- which is central in the application of NLRR. Two of the four parameters,  $k_2$  and  $BP$ , appear both outside and inside the integral in (4.5) and the computation of the derivatives of (4.5) with respect to them, even if not difficult, may be tricky. The possibility to switch the derivative and integral operators, necessary for deriving with respect to  $k_2$  and  $BP$ , is legitimate in the case of a regular function as (4.5). For the sake of space the known functions

$$(C_{Ref}(t) - V_B^R C_B(t))$$

and

$$\int_0^t (C_{Ref}(x) - V_B^R C_B(x)) e^{-\frac{k_2}{1+BP}(t-x)} dx$$

. will be denoted with  $g(t)$  and  $h(t)$ , respectively, and the derivatives of (4.5) with respect to the four parameters  $R, k_2, BP, V_b$  are reported here for completeness.

$$\frac{\partial f}{\partial R} = \frac{1 - V_B^T}{1 - V_B^R} g(t) - \frac{k_2}{1 + BP} h(t) \quad (4.6)$$

$$\begin{aligned} \frac{\partial f}{\partial k_2} = \frac{1 - V_B^T}{1 - V_B^R} \left\{ \left(1 - \frac{R}{1 + BP}\right) h(t) + \left(k_2 - \frac{k_2 R}{1 + BP}\right) \right. \\ \left. \left[ -\frac{t}{1 + BP} h(t) + \int_0^t g(x) e^{-\frac{k_2}{1+BP}(t-x)} \frac{x}{1 + BP} dx \right] \right\} \end{aligned} \quad (4.7)$$

$$\begin{aligned} \frac{\partial f}{\partial BP} = \frac{1 - V_B^T}{1 - V_B^R} \left\{ \frac{k_2 R}{(1 + BP)^2} h(t) + \left(k_2 - \frac{k_2 R}{1 + BP}\right) \right. \\ \left. \left[ \frac{k_2}{(1 + BP)^2} t h(t) - \int_0^t g(x) e^{-\frac{k_2}{1+BP}(t-x)} x \frac{k_2}{1 + BP} dx \right] \right\} \end{aligned} \quad (4.8)$$

$$\frac{\partial f}{\partial V_B^T} = C_B(t) - \frac{1}{1 - V_B^R} g(t) \quad (4.9)$$

Each derivative is time dependent, as it is to be computed in the correspondence of the N mid-scan times yielding a column vector. The four column vectors together form the Nx4 sensitivity matrix S.

### Computation of priors

The computation of priors is an issue which deserves a deeper analysis as, clearly, final estimates are heavily dependent on priors and, besides, the computation of priors on three-dimensional images deserves further comments. In NLRR3, the version which emerged as the optimal one and which was used here, priors are computed fitting the appropriate kinetic model to clusters' TACs.

Which is, however, the best way to cluster a 3D image? Segmenting all voxels together is computationally not feasible, at least with the clusterization technique which was employed. In fact, hierarchical cluster analysis is very time consuming and its computational time, besides, is non linearly dependent on the number of elements to be clustered: a two-fold increase of voxels to be segmented may give rise to a five to eight fold increase of computational time. The clusterization of hundreds of thousand of voxels is far beyond the possibility of current computers.

The most reasonable idea would probably be to cluster each slice separately. Even if feasible, this approach would cause, in any case, a relevant increase of computational cost, as, in this application, one should cluster 91 slices. To cluster each slice separately would have required, on the computer employed, roughly one hour and a half for each subject. Is this really necessary? Isn't enough to consider a single slice which contains voxels with remarkably different kinetics (gray matter voxels, white matter voxels, blood pool voxels)? This is the approach which was employed and which proved to work reasonably well, at least from a visual inspection of the parametric maps of the priors. The only thing to check was that the resulting final cluster TACs -the standard number of 6 clusters was employed- were not all similar to each other but were, in a certain sense, representative of the different brain tissues in order to obtain, before the application of CRM, sets of kinetics parameters reasonably different from each other. One slice was sufficient to achieve this goal with all the subjects considered. A drawback of the suggested approach, however, was that results were dependent on the slice chosen. Slight differences were noticed, as expected, when two contiguous slices were employed for the generation of priors but differences became noticeable to the naked eye when two slices far apart from each other (e.g. slice n°1 and n°40) were employed. Slice n°15 was heuristically selected and employed for all subjects. In the previous chapters the different versions of NLRR were referred to as NLRR1, NLRR2, NLRR3. As stated above, only NLRR3 with CRM was used here; the method just described for the computation of priors -clusterization of one slice before CRM- will be referred to as NLRR3a.

Alternatively, one could consider the additional information provided by the anatomical atlas, which allows to automatically extract a set of ROI TACs. In this new version of NLRR3 (NLRR3b) pixels which, according to the atlas, belonged to a specific ROI, were simply assigned, as prior, the kinetic parameters derived fitting the TAC of that ROI. Priors for pixels which were not assigned to a specific ROI, were, on the other side, derived exactly as in NLRR3a. In other words the two methods computed priors

differently only for those pixels which the atlas assigned to a specific ROI, for which the parameter vector of that ROI was used as prior.

In NLRR3b, therefore, the same priors were assigned to all pixels of the same ROI. One could fear that the basic equation to compute the ridge matrix in GRRSC, reported here,

$$h_i = \frac{\sigma^2}{(p_i - p_{i0})^2} \quad i = 1, 2, \dots, m \quad (4.10)$$

can not be employed because the denominator of (4.10) is 0 for all parameters, as the prior  $p_{i0}$  is computed averaging the estimates in the neighbourhood of the pixel of interest and, if these estimates are all equal to each other as in NLRR3b, one would get  $p_i = p_{i0}$ . This does not happen because, as explained in paragraph (2.3.3), (4.10) is not employed directly on parametric maps of priors, but after the computation of

$$\Delta p = (S'WS)^{-1}[S'W(y - h(p))] \quad (4.11)$$

with S denoting the appropriate sensitivity matrix at pixel i, y the TAC and p the prior at the same pixel. The parameter vector p is then updated as  $p + \Delta p$ , and (4.10) can then be successfully applied. All these details were already presented, but were reported here again in order to clear potential doubts that may have risen.

After the initial attribution of priors, the two version of NLRR3 (NLRR3a, NLRR3b) worked in the same way according to the algorithm described previously.

### 4.3.2 Linear version

Although SRTM and SRTMV are non-linear models the basis function approach can be successfully employed to linearize the problem allowing to employ linear ridge regression even in presence of a Bayesian term. To analyze the more general case, SRTMV was considered, as already said ; equations concerning SRTM are simpler and can be obtained by setting  $V_B^T = V_B^R = 0$ . With the same identical notation of above ( $\vartheta = \frac{k_2}{1+BP}$ ,  $[p_1, p_2, p_3] = [V_B^T, R(1 - V_B^T), (k_2 - \frac{k_2 R}{1+BP})(1 - V_B^T)]$ ), and with the additional position  $C^*(t) = \frac{1}{1-V_B^R} (C_{Ref}(t) - V_B^R C_B(t)) =$  known quantity, (4.5) can be trivially rewritten as

$$C_{Tar}(t) = p_1 C_B(t) + p_2 C^*(t) + p_3 \int_0^t (C^*(x) e^{-\vartheta(t-x)}) dx \quad (4.12)$$

This is just a change of notation to reduce the length of the equation. Once the elements of  $p = [p_1, p_2, p_3]$  are estimated through a basis function approach, the prior  $p_0$  and the 3x3 ridge matrix H can be computed at each pixel as explained in the linear ridge regression section. For each fixed value of  $\vartheta$  belonging to a predefined reasonable set, as above, if we let X be the matrix whose columns X(i), i=1,2,3 are respectively given by  $C_B(t), C^*(t), \int_0^t (C^*(x) e^{-\vartheta(t-x)}) dx$ , the estimate of p in the linear ridge regression framework is the one which minimizes the cost function

$$(y - Xp)W(y - Xp)' + (p - p_0)H(p - p_0)' \quad (4.13)$$

which is given by

$$p_{optimal}(\vartheta) = (X'WX + H)^{-1}(X'Wy + Hp_0) \quad (4.14)$$

with W as usual denoting a diagonal matrix containing the weights. At the end the value of  $\vartheta$  whose correspondent  $p_{optimal}(\vartheta)$  gave rise to the minimum of the cost function (4.13)

is retained and the elements of  $\beta = [R, k_2, BP, V_B^T]$  are computed exactly as explained before. The simple yet effective idea to "linearize" the original model by setting  $\vartheta$ , the parameter which makes the model non-linear, to a fixed value is therefore used here as well with the only difference that the selected value  $\vartheta_{optimum}$  is now the one which realizes the optimum trade-off between the weighted distance from the measured data  $y$  and from the prior  $p_0$ . The application of linear ridge regression to SRTMV, "linearized" through the application of BFM, will be referred to as LRR (linear ridge regression) to distinguish it from the non linear version version described in the previous paragraph.

## 4.4 Application of GTS

The application of GTS deserves some special comments as well because several strategies are, in principle, feasible and only the one which was selected as the most appropriate was employed.

One possible method is to perform the population analysis on the vector  $p = [p_1, p_2, p_3]$  and then, from the new values of  $p$  computed using GTS, evaluate the elements of  $\beta = [R, k_2, BP, V_B^T]$  which are the ones of interest. This approach, however, leaves unchanged the value of  $\vartheta_{optimum}$  determined from the application of the basis function method to the measured data only. One could think to apply GTS also to the value of  $\vartheta_{optimum}$  computed at each pixel from the preliminary application of the BFM, but this is not possible since this method does not allow to compute the variance of this parameter, which is essential for the application of population algorithms. More reliable estimates for  $p$  obtained through the application of GTS, therefore, do not necessarily result in better estimates for  $\beta$ , more precisely for  $k_2$  and  $BP$  which need  $\vartheta_{optimum}$  to be computed; the reason is that the "new" values of  $p$  and the "old" value of  $\vartheta_{optimum}$  are combined together. This approach was tried but then discarded.

The already mentioned equations

$$\begin{aligned} V_B^T &= p_1 \\ R &= \frac{p_2}{1 - V_B^T} \\ k_2 &= \frac{p_3 + R\vartheta}{1 - V_B^T} \\ BP &= \frac{k_2}{\vartheta} - 1 \end{aligned}$$

allow to compute at each pixel not only the elements of  $\beta$  but also their associated variance. To achieve this goal the linear transfer model (LTM) was applied. With the LTM one computes an approximation of the variance of a parameter which is function of one or more parameters for which the covariance matrix is available. Let  $y$  be a known function  $y = f(x_1, x_2)$  of two parameters  $x_1$  and  $x_2$  and assume that a point estimate  $x_1^*$  and  $x_2^*$  of  $x_1$  and  $x_2$  and of their covariance matrix is also available.

The LTM approximates the variance of  $y$   $\sigma_y^2$ , according to

$$\left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + \frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} cov_{x_1 x_2} \quad (4.15)$$

where the partial derivatives of  $f$  are computed in correspondence of  $x_1^*$  and  $x_2^*$ . It is straightforward how (4.15) can be generalized to the case of 3 or more variables: with

3 variables there will be 6 terms, 3 with the squared of the first derivatives and 3 with the mixed derivatives.

In the case of SRTMV,  $x_1, x_2$  and  $x_3$  are represented by  $p_1, p_2, p_3$  and, assuming a basis function method is employed, the covariance matrix of the parameter vector  $p$  is computed from the usual expression

$$\Sigma_p = \sigma^2(X'WX)^{-1}$$

used for the linear models in PET. Applying (4.15) to compute the variances of the parameters of interest it follows easily, for instance, that

$$\sigma^2_{V_B} = \sigma^2_{p_1}$$

and

$$\sigma^2_R = \frac{\sigma^2_{p_2}}{(1 - V_B)^2} + \frac{\sigma^2_{V_B} p_2^2}{(1 - V_B)^4}$$

In the same way variances for  $k_2$  and BP can be computed.

Using this method the variances associated to each parameter were derived and GTS was then applied separately for each of the 4 kinetic parameters  $R, k_2, BP, V_B^T$  using the criterion determined in the simulation phase to stop the algorithm (see Chapter 3, relative variation less than 1% for at least 90% of the pixels of the population).

## 4.5 Computational cost

Before starting the analysis of results a word is in order on the computational cost of the different algorithms. With SRTMV and the BFM approach, it took roughly half an hour to generate complete parametric maps for one subject. The obvious idea which allowed to save a lot of computational time was to compute the convolution integrals for the different values of  $\vartheta$  only once at the beginning of the algorithm, as they are not pixel-dependent. LRR with BFM, therefore, required the same identical time to generate the final maps because the same approach was used. The application of GTS itself required less than 5 minutes for slice. It is recalled, however, that if GTS or LRR are to be employed, the computational cost of SRTMV itself is to be taken into account, as this is the preliminary step for both GTS and LRR.

This is not true for NLRR3 approaches, which do not require any preliminary estimation. Although these algorithms are based on appropriate linearizations of non-linear problems, they require a precise computation of the sensitivity matrix at each pixel and iteration, which is very time-consuming. Only one iteration was performed but one hour, nonetheless, was required on average to generate final maps.

Among the Bayesian approaches, therefore, GTS emerged as the fastest one. This is not a negligible issue, especially when generating 3D parametric maps (all the values reported refer to a *single slice* only...).

## 4.6 Examples of parametric maps

In this section examples of parametric maps obtained using SRTMV, ridge regression, in both non linear (NLRR3a, NLRR3b) and linear (LRR) versions, and GTS are displayed for each of the 4 kinetic parameters. These images deserve a few comments.

- Maps obtained with any Bayesian method are, as expected, more regular at visual inspection than SRTMV maps.
- Linear ridge regression maps are the most similar to the original SRTMV maps; they seem to have obtained from SRTMV through a process of smoothing.
- In NLRR3b and GTS maps the employment of the anatomical information of the atlas can clearly be seen. The structure of the cerebellum, for instance, can be easily singled out from any parametric map, and especially from BP and  $V_b$  maps. The possibility to single out anatomical structures from parametric maps clearly depends on the parameter and slice considered. Slice n°21 was selected in order to highlight this interesting aspect, which was evident for all subjects. This effect was present for most slices, although the amount was lower than for the presented slice.
- $V_b$  maps are those which are more heavily modified by the application of ridge regression and GTS: the reason is that these are the noisiest and least regular maps, as can be realized from visual inspection. The application of Bayesian algorithms has, for this parameter, the most significant regularizing effect.

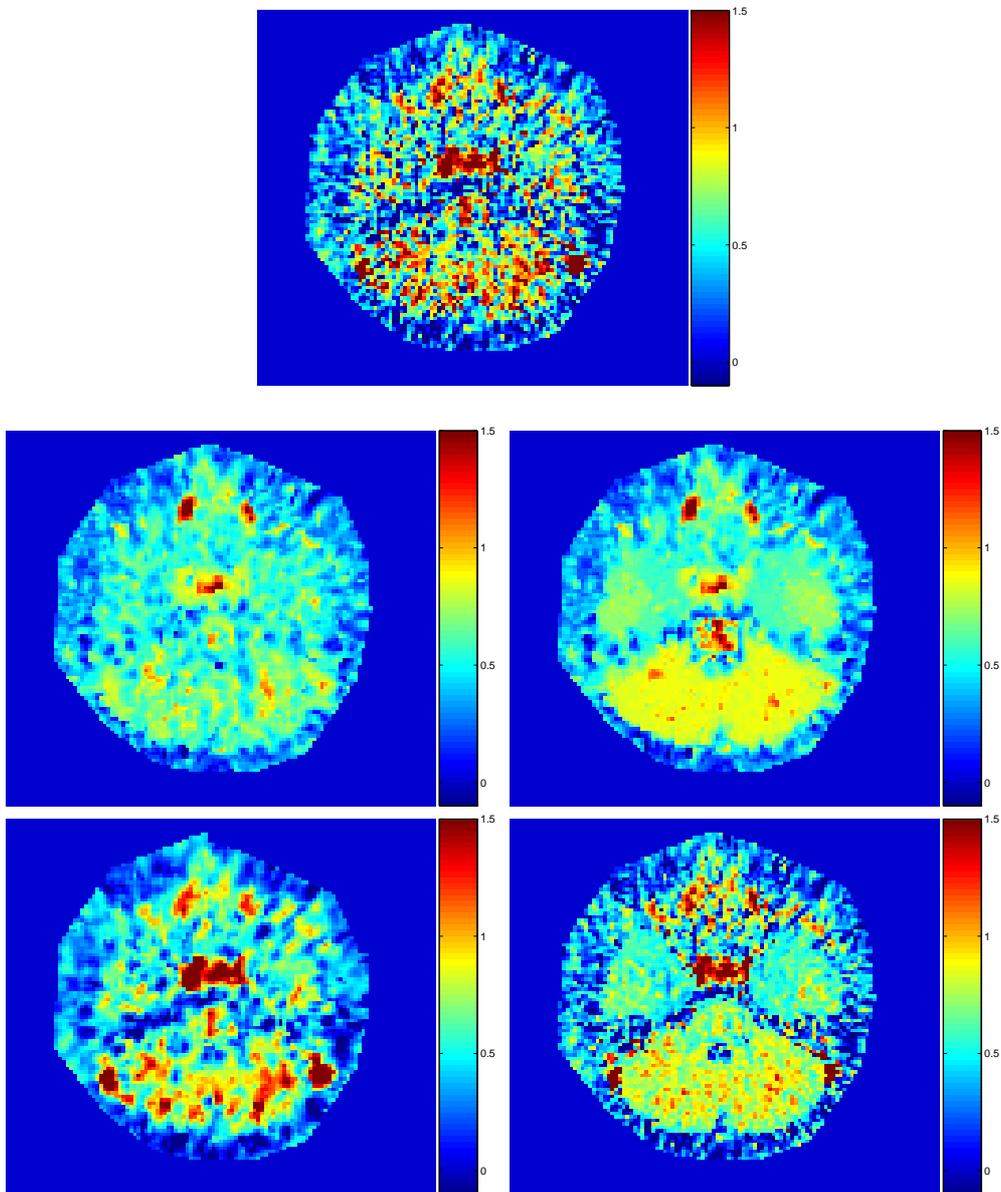


Figure 4.1: *Parametric maps of  $R$  for one slice of a HD patient (slice n° 21). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively).*

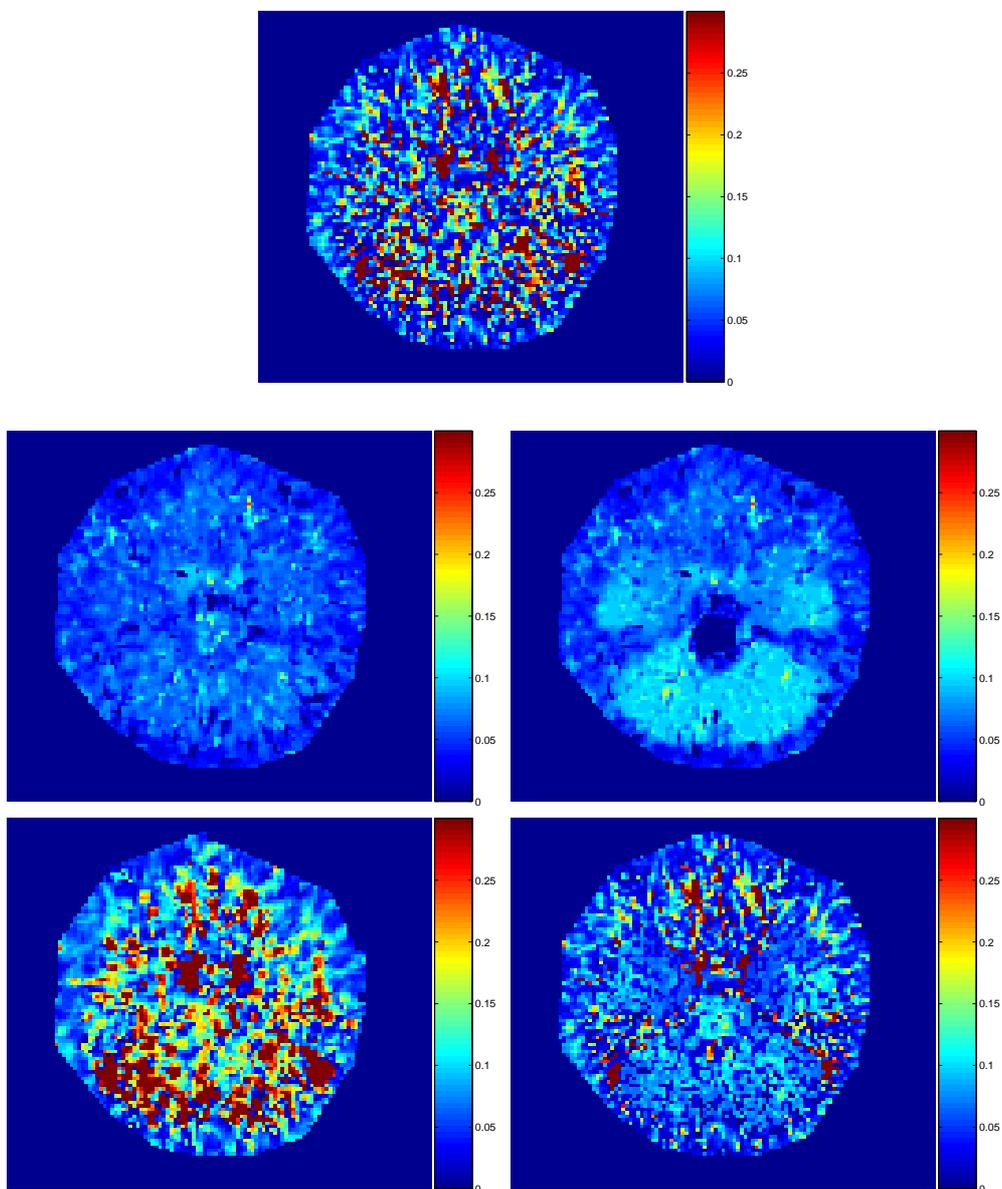


Figure 4.2: *Parametric maps of  $k_2$  for one slice of a HD patient (slice n°21). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively).*

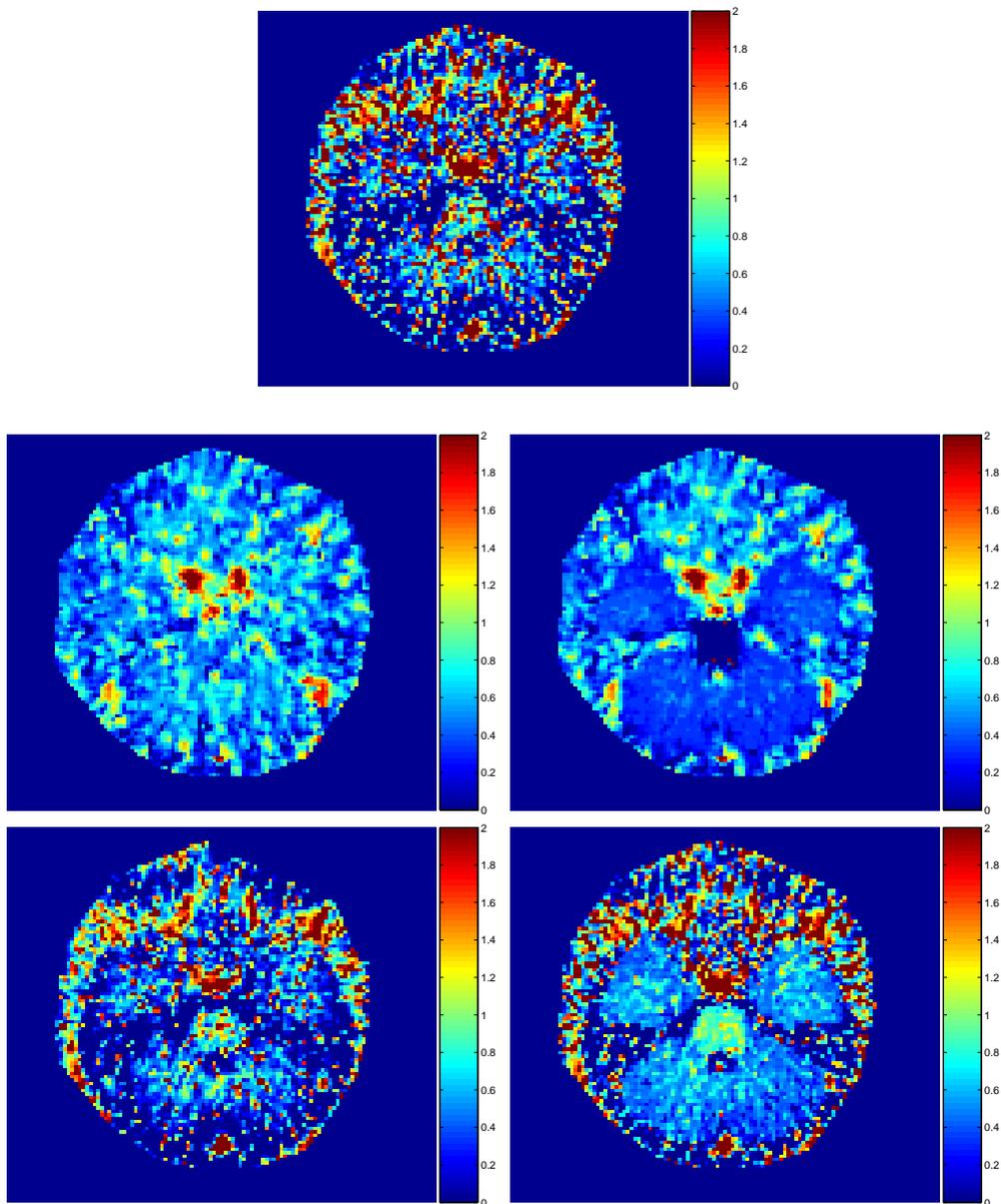


Figure 4.3: *Parametric maps of BP for one slice of a HD patient (slice n°21). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively).*

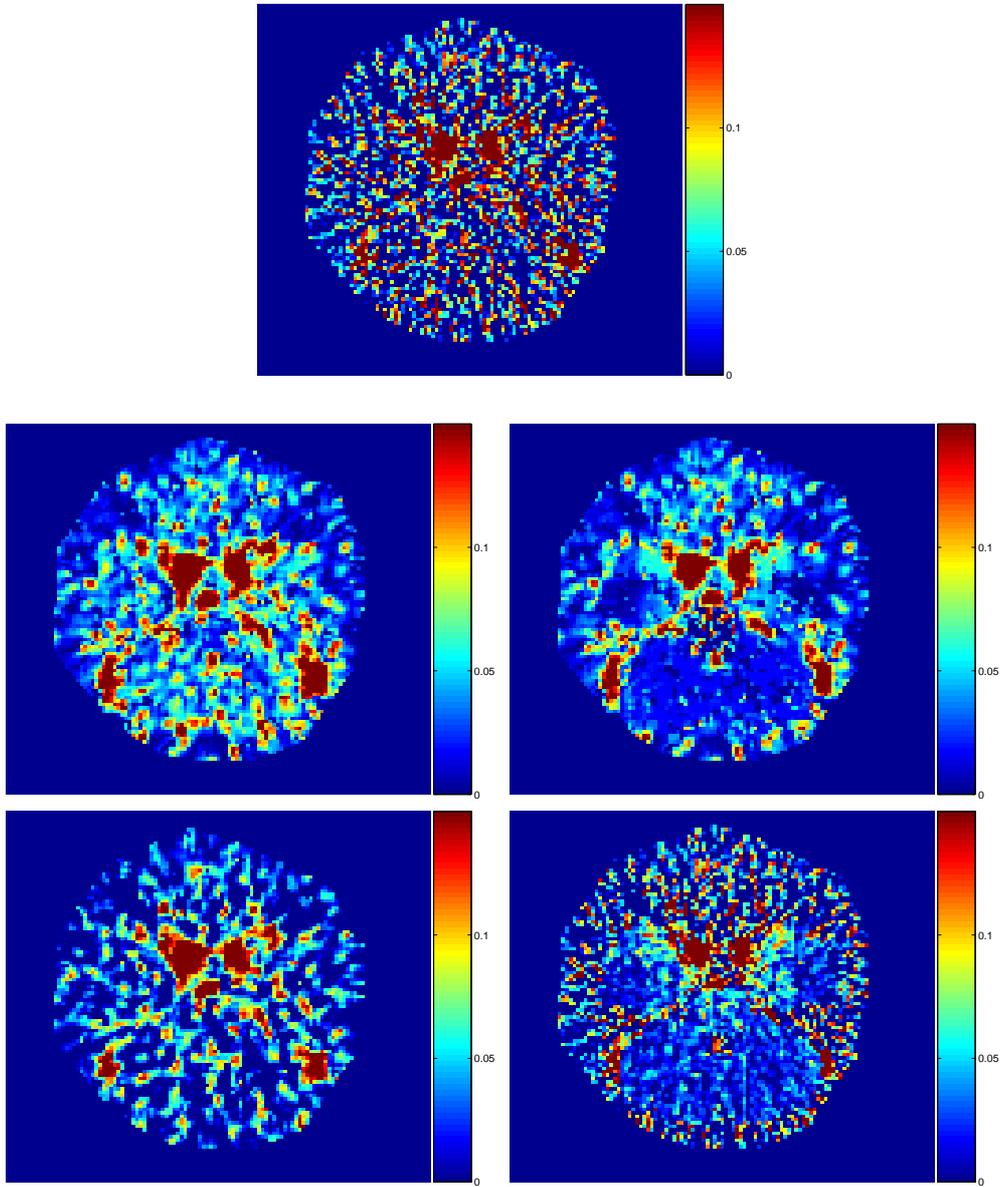


Figure 4.4: *Parametric maps of  $V_b$  for one slice of a HD patient (slice n° 21). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively).*

## 4.7 Correlation between ROI and pixel estimates

If a model is not linear it is not rigorously true, even in absence of noise, that the mean  $p_{pixel}$  of N kinetic parameters computed using N pixel TACs  $C_i(t)$   $i=1,2,\dots,N$ , is equal to the parameter  $p_{ROI}$  computed from the ROI TAC  $C_{ROI}(t)$ , having defined  $C_{ROI}(t)$  as  $\Sigma C_i(t)/N$ . A large correlation between  $p_{ROI}$  and  $p_{pixel}$ , however, is index of reliability of a parametric map.

Using the stereotaxic atlas, 10 ROI TACs were automatically extracted. The 10 ROIs were cerebellum, lateral occipital lobe, anterior cingulate gyrus, posterior cingulate gyrus, frontal lobe-middle frontal gyrus, posterior temporal lobe, parietal lobe, thalamus, occipital lobe-lingual gyrus and occipital lobe-cuneus. For each of the 12 healthy subjects and each ROI, the 4 kinetic parameters were then computed using (4.5), for a total of 120 estimates. Parameters computed from ROI TACs were compared to those obtained averaging the corresponding parameters at pixel level, using the anatomical atlas to associate to each ROI its corresponding pixels. A coefficient  $\eta$  was computed for  $R, k_2, BP, V_b$  and for each of the 5 methods (SRTMV, NLRR3a, NLRR3b, LRR and GTS) assuming the parameters derived from ROI analysis as the noise-free variables.

$\eta$  was defined as:

$$100 \frac{1}{N} \sum \frac{|p_{pixel} - p_{ROI}|}{|p_{ROI}|} \quad (4.16)$$

where  $||$  denotes the absolute value and  $p$  indicates a generic kinetic parameter, with  $p_{ROI}$  and  $p_{pixel}$  as above. The sum was performed over all ROIs and subjects for a total of  $N=120$  elements. Basically  $\eta$  quantifies the distance from the identity line of the element of coordinates  $(p_{ROI}, p_{pixel})$  in the x-y plane. An  $\eta$  of 10, for instance, indicates that, on average, the difference between  $p_{pixel}$  and its corresponding  $p_{ROI}$  is of 10%. This index was preferred to the classic Pearson's correlation coefficient because the latter is related to the existence of a *generic linear correlation* between a set of  $(x_i, y_i)$ , whereas the aim, here, was to quantify how much the set  $(p_{ROI,i}, p_{pixel,i})$  was close to the *identity line*.

Figure (4.5) summarizes the results;  $\eta$  is displayed for the 4 parameters and the 5 methods employed. Values of  $\eta$  bigger than 150 (=150%) were bound to 150 for illustration purposes.

- In SRTMV, as expected, R is the best estimated parameter at pixel level, followed by BP,  $k_2$  and  $V_b$ .
- LRR gave results which were very similar to standard SRTMV- except for the improvement in  $V_b$  correlation- consistently with the fact that LRR maps significantly resembled SRTMV maps.
- Between NLRR methods, NLRR3b outclassed NLRR3a in terms of BPs. This is particularly important as the binding potential is the most important parameter.
- GTS yielded the same results of SRTMV in terms of R and BP, whereas improvements were relevant for  $k_2$  and  $V_b$ .

Globally NLRR3b and GTS provided the best results.

These ideas are further illustrated in Figures (4.6) and (4.7) which show plots of the relative delivery R and of the blood fraction  $V_b$  evaluated at ROI level (x axis) and pixel level (y axis) for SRTMV (left) and GTS (right). Whereas the good correlation

between  $R_{ROI}$  and  $R_{pixel}$  in SRTMV was basically left unchanged by GTS, there was a remarkable improvement when  $V_b$  was considered, consistently with what shown before.

A final comment is in order. The evaluation criterion here was the correlation between estimates at ROI level and those at pixel level, computed averaging estimates of pixels in the same ROI. The use of this criterion implicitly favoured NLRR3b and GTS. In fact, these methods employed the same anatomical information used in the evaluation criterion -the assignation of a given pixel to a specific ROI according to the atlas- to compute priors and construct the "populations", respectively. This explains why these algorithm provided the best performances according to the described criterion.

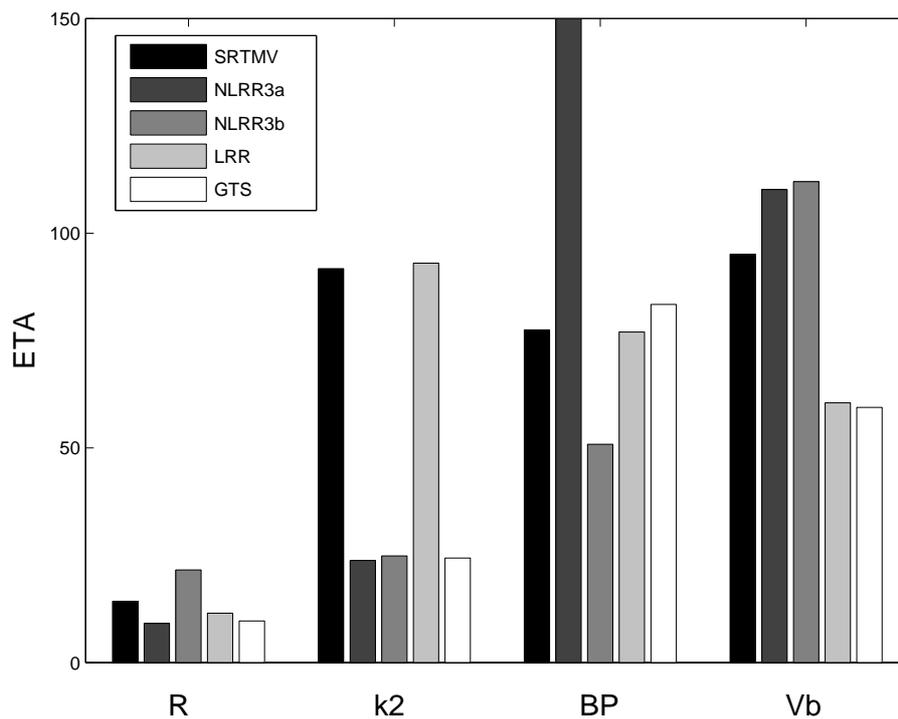


Figure 4.5: The coefficient  $\eta$  is displayed for each of the 4 parameters and each of the 5 methods used to generate parametric maps.

## 4.8 Additional analysis on binding potentials

One of the most important potentialities offered by PET, as stated in the introduction, is its ability to diagnose a certain neuro-degenerative disease with more certainty than with current methods.

Ill subjects have, on average, higher BPs due to the higher binding of  $[^{11}\text{C}]$ -(R)-PK11195, which is caused by the higher amount of activated microglia which is on its turn related to the level of neuro-inflammatory response. In this context the expression "ill subject" is used to refer to a person suffering from HD, Alzheimer's disease, or other neurodegenerative conditions (e.g. dementia with Lewy bodies) characterized by some kind of neuro-inflammatory response, which can be highlighted by  $[^{11}\text{C}]$ -(R)-PK11195. From now on, therefore, the analysis will be focused on BP only as it is the key parameter

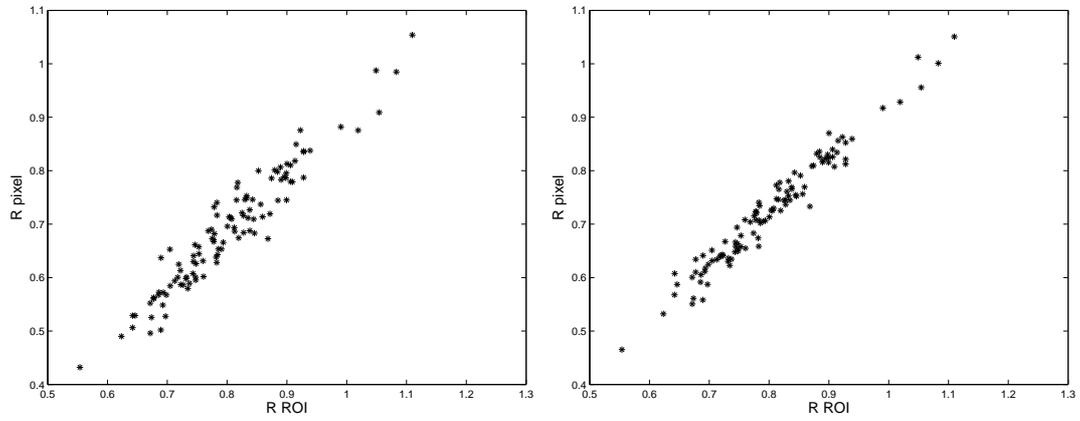


Figure 4.6: Plot of the relative delivery  $R$  evaluated at ROI and pixel level ( $x$  and  $y$  axis respectively) is displayed for SRTMV (left) and GTS (right).

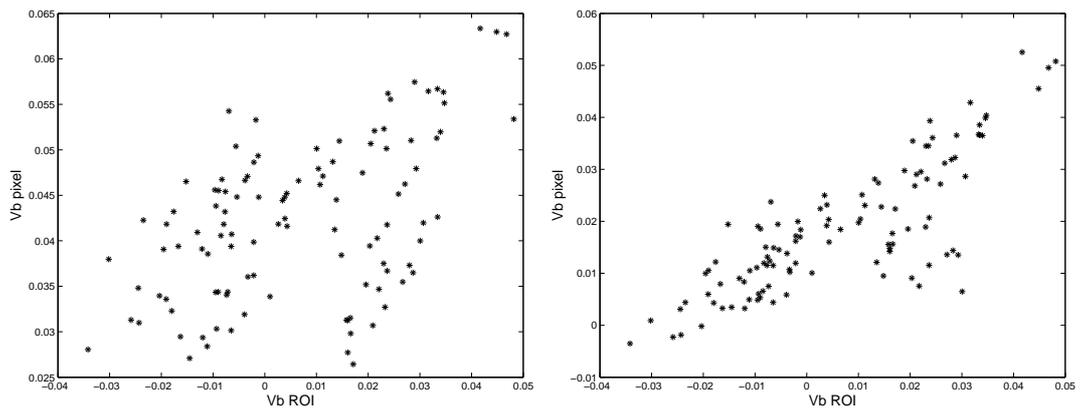


Figure 4.7: Plot of the blood fraction  $V_b$  evaluated at ROI and pixel level ( $x$  and  $y$  axis respectively) is displayed for SRTMV (left) and GTS (right).

in this context.

#### 4.8.1 Correlation with SRTMV BPs

In this subsection the correlation between BP estimated through SRTMV, the method of comparison, and those computed with the different Bayesian approaches will be analyzed. Figure (4.8) displays the correlation between the BPs computed with SRTMV, sorted and reported along the x axis, and those estimated through NLRR3a, NLRR3b, LRR and GTS, respectively. The BPs displayed belong to pixels of the cerebellum of a randomly chosen HD subject. Similar plots for different ROIs and subjects, however, yielded almost identical trends. The identity line is reported in yellow. The same range for the y-axis was employed for 3 out of the 4 graphs; for LRR a wider range (-1:4) was used. The plots highlight interesting aspects of the different methods, some of which were already presented in the previous paragraphs.

It was previously stated that LRR is the least "invasive" approach; consistently, it yielded the best correlation with SRTMV, with BPs aligned along the identity line. This fact can be interpreted positively -the method changes little of the original maps providing at the same time smoother final maps- or negatively, as, for instance, negative and therefore non-physiological values of BPs, were still present.

The two plots of NLRR methods look very similar; negative values were eliminated, whereas high values of SRTMV BPs are significantly reduced. These facts are clearly positive -unreliable values (negative or too high) are corrected- but a tendency of underestimating too much high BPs is visible. NLRR3a performed better than NLRR3b, which provided a lot of outliers.

GTS plot is of particular interest; as in NLRR negative values were corrected and higher values reduced, but the amount of the reduction was lower for high BPs, which is a positive fact. The plots appears of higher quality at visual inspection with respect to NLRR plots; BPs with the most physiological values (range 0.3-1) align almost perfectly along the identity line. Pearson's correlation coefficient, besides, was 0.72 for GTS, whereas it was only 0.26 and 0.1 for NLRR3a and NLR3b respectively.

The conclusion is that, once again, LRR emerged as the least invasive approach. GTS, in this context, performed remarkably better than NLRR algorithms, yielding higher correlations and correcting for non-physiological values without underestimating too much high BPs.

#### 4.8.2 BP variability within a certain ROI

Figure (4.9) displays mean BPs, estimated with SRTMV and the 4 Bayesian approaches for 3 normal controls NC and 3 HD patients. Results refer to the putamen; almost identical trends, however, were found when different regions were studied and results are displayed for putamen only for the sake of space. In the lower part of the figure the standard deviations of BPs for the same ROI and subjects are displayed.

LRR, as usual, gave mean BPs which were always very similar to the corresponding SRTMV BPs; GTS BPs were very related to SRTMV BPs, although slightly lower. NLRR3b BPs were well correlated with SRTMV BPs, but consistently lower. NLRR3a BPs didn't show a clear trend when compared to SRTMV BPs.

The lower section of figure (4.9) unequivocally shows how NLRR approaches and GTS gave an incredible reduction of BP variation. Standard deviation were reduced

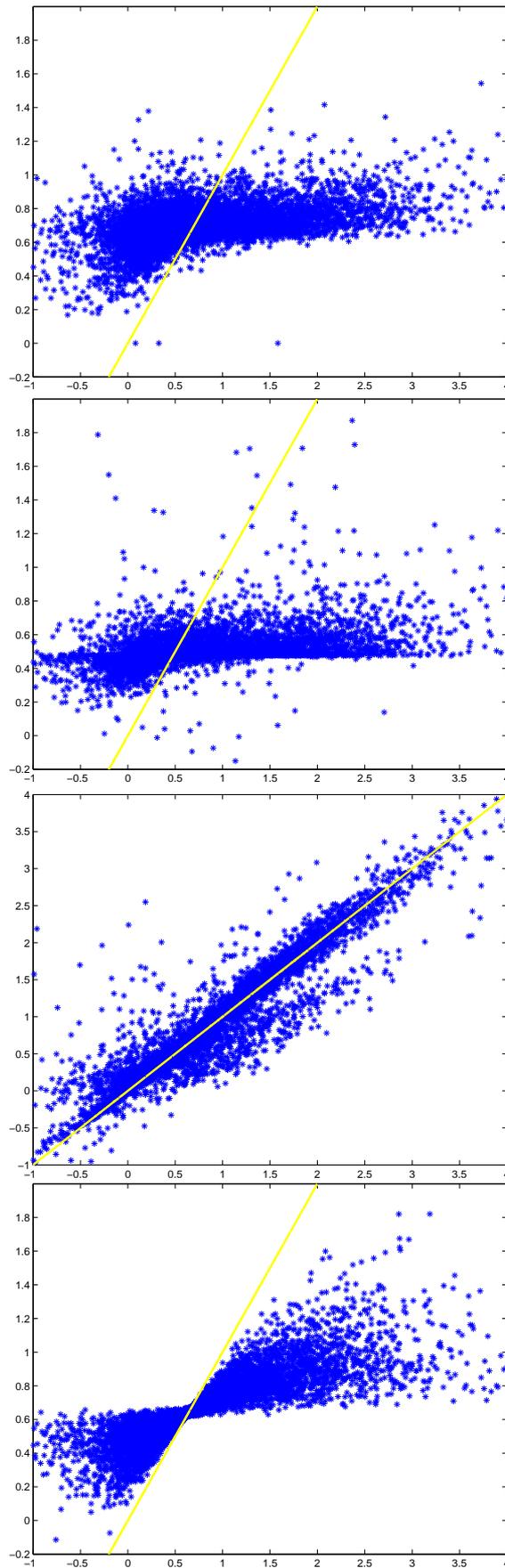


Figure 4.8: Plot of the correlation between SRTMV BPs and the corresponding BPs estimated with the 4 Bayesian approaches analyzed for pixels of cerebellum of a HD patient. The identity line is displayed in yellow.

to one fifth-one sixth, on average. This fact could already be inferred from the visual assessment of parametric maps but here it has been quantitatively assessed. As usual LRR results were almost identical to SRTMV results.

## 4.9 Differentiation between healthy and HD subjects

It is very important, thinking of possible application of PET in a clinical environment, to study what happens to the ability to differentiate between healthy and ill subjects when different methodologies are employed.

When analyzing different images it is not correct, in general, to make pixel-by-pixel comparisons because two pixels in the same position (x,y,z) do not necessarily correspond to the same anatomical component, due to unavoidable differences in the position of the scanned subjects inside the PET scanner; the normalization to the same stereotaxic space, however, made the pixel-by-pixel comparison performed here a more robust and reliable approach.

### 4.9.1 Statistical tests employed

#### T-test

The comparison of HD and NC was performed using first a simple t-test on a pixel basis in order to classify each pixel as "ill" or "healthy". It is recalled here that a t-test assumes that two groups of random samples  $x_1, x_2, \dots, x_N$   $y_1, y_2, \dots, y_M$ , normally distributed with mean  $\bar{x}$  and  $\bar{y}$ , respectively, and the same variance  $\sigma^2$ , are available. The test can assess if  $\bar{x}$  is statistically different from  $\bar{y}$ . A test statistics is computed as

$$\frac{\tilde{x} - \tilde{y}}{\sigma \sqrt{\frac{1}{N} + \frac{1}{M}}} \quad (4.17)$$

where  $\tilde{x}$  and  $\tilde{y}$  are the sample mean of  $x_1, x_2, \dots, x_N$  and  $y_1, y_2, \dots, y_M$  and  $\sigma$  in (4.17) is the sample standard deviation defined as the square root of

$$\frac{(N-1)s_x^2 + (M-1)s_y^2}{N+M-2} \quad (4.18)$$

with  $s_x^2$  sample variance of the  $x_1, x_2, \dots, x_N$  set, simply defined as

$$\frac{\sum_{i=1}^N (x_i - \tilde{x})^2}{N-1} \quad (4.19)$$

This test statistics, under the null hypothesis  $\bar{x} = \bar{y}$  has a t-distribution with N+M-2 degrees of freedom and can be easily employed to compute an appropriate p-value. A one-tailed test  $\bar{x} > \bar{y}$ , where the x variable stands for the BPs of the HD subjects and y for the BPs of the NC subjects, was considered here, and therefore the p-value was simply the area on the right of the test statistics in the t-distribution with N+M-2 degrees of freedom. Low p-values indicate that the alternative hypothesis should be accepted.

A t-test was performed on the BPs of each pixel separately, comparing BPs of HD to BPs of NC. A threshold of 0.01 was employed to declare "activation"; p-values lower than 0.01, in other words, were considered as indicators of a significant difference at any

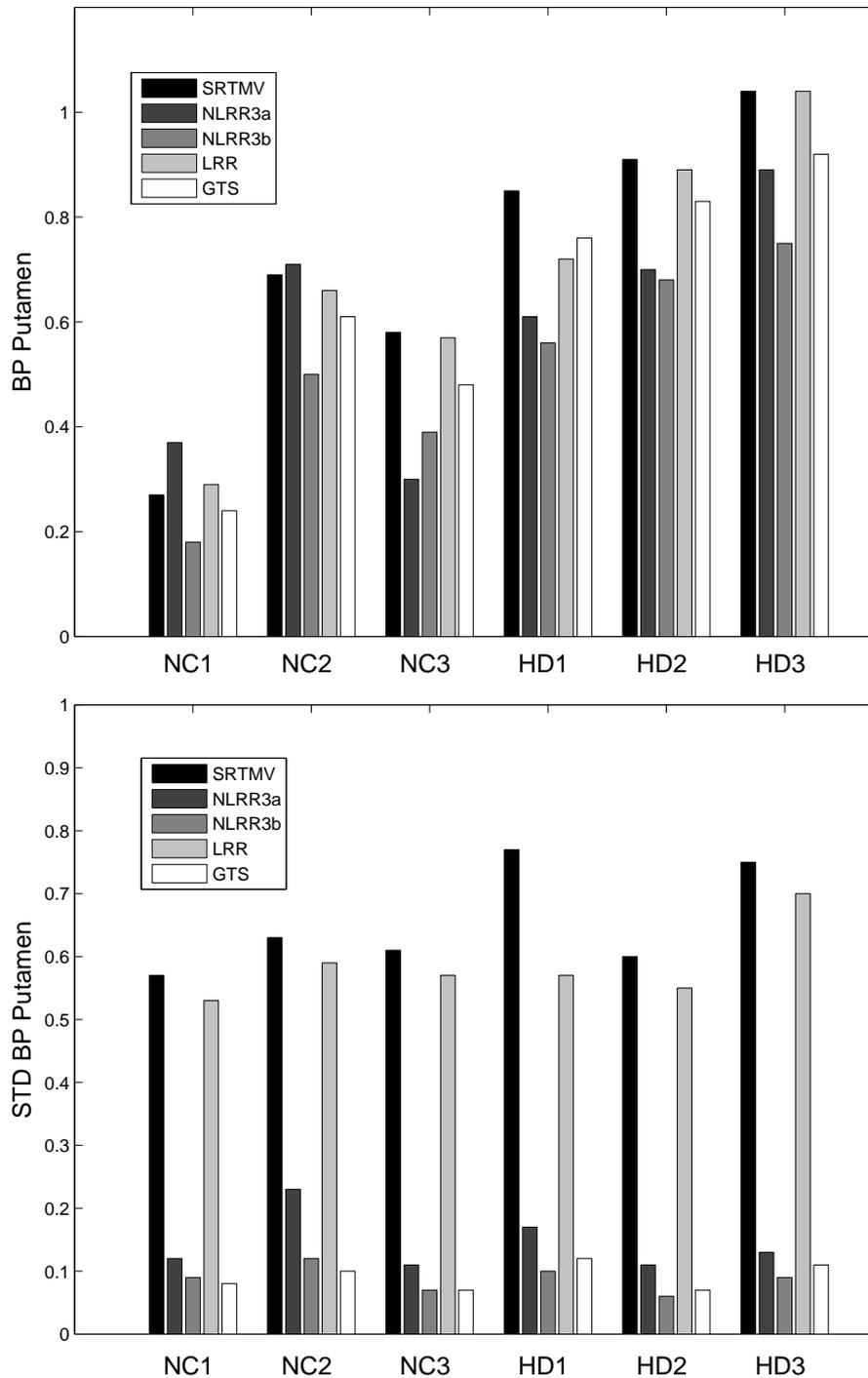


Figure 4.9: Mean BPs, estimated with SRTMV and the 4 Bayesian approaches, and the corresponding standard deviations for 3 normal controls NC and 3 HD patients are displayed in the upper and lower part of the figure respectively. The BPs belong to pixels of putamen.

given pixel between the HD and NC BPs. No Bonferroni correction was performed; the idea was to compare different methods in the same identical conditions rather than keep the total Type 1 error at a fixed threshold using Bonferroni correction, FWE (Family Wise Error) or FDR (False Discovery Rate).

### Wilcoxon rank-sum test

The assumption of normal distribution of binding potentials may be uncorrect. If so, the results provided by the t-test may be not reliable. A well-known non-parametric alternative is provided by the Wilcoxon rank-sum test. The only assumption of the test is that the random samples  $x_1, x_2, \dots, x_N$   $y_1, y_2, \dots, y_M$  are drawn from continuous distributions X and Y with the same shape and spread, the only possible difference between the two being the values of their means  $\bar{x}$  and  $\bar{y}$ ; no normality assumption is required. The (N+M) values  $x_1, x_2, \dots, x_N, y_1, y_2, \dots, y_M$  have to be sorted in increasing order so that ranks from 1 to (N+M) can be assigned to each element. The test statistics value w is computed summing the rank associated to the N  $x_i$  (here the X distribution was chosen, but this clearly quite arbitrary). Clearly a high value of w indicates that it is likely that  $\bar{x}$  is greater than  $\bar{y}$  and viceversa. Under the null hypothesis  $\bar{x} = \bar{y}$  the rank of any  $x_i$  is equally likely to be any of the values 1,2,...(N+M): its mean is therefore (N+M+1)/2, and the mean of W (the random variable of the sum of ranks of  $x_i$ , from which w is drawn) is therefore  $\mu=N(N+M+1)/2$ . It can also be shown that the variance of W is  $\sigma^2= MN(N+M+1)/12$ . This is always true, independently on N and M; if both N and M are greater than 8 ([44]) it is a good approximation to assume W Gaussianly distributed thanks to the central limit theorem. This is true in this case as N=n° of HD subjects=9, M=n° of NC=12.  $(W-\mu)/\sigma$  is therefore normally distributed with mean 0 and variance 1 and p-values can then be easily derived using  $(w-\mu)/\sigma$  as test statistics. In the considered case the Wilcoxon rank-sum test was performed at each pixel under the alternative hypothesis that the mean  $\bar{x}$  of HD BPs was greater than the mean  $\bar{y}$  of NC BPs. A threshold of 0.01 was employed to declare "activation", as before.

### 4.9.2 Sensitivity and specificity

It is essential, when considering a certain estimation method, to analyze its ability to classify correctly ill and healthy subjects in order to provide a reliable diagnosis. This fact can be quantified through the use of the sensitivity Se and specificity Sp. In this context Se can be defined as the ratio of the ill subjects correctly classified as ill to the total number of ill subjects; analogously Sp is the ratio of the healthy subjects correctly classified as healthy to the total number of healthy subjects. These indices are clearly comprised between 0 and 1 and the closer they are to 1, the better the method is. Basically one would like a method with a high sensitivity - a high ability to detect ill subjects, and a high specificity -high probability to classify correctly healthy subjects.

Se and Sp were introduced considering the correct or incorrect classification of individuals. The same identical definitions hold for pixels and, in this context, Se and Sp will be considered as applied to pixels, which were classified as "ill" or "healthy" on the basis of statistical tests described above.

To assess both Se and Sp analysis was addressed to the putamen and the cerebellum.

Huntington's disease is known to severely affect the striatum, of which putamen is a part. The neuro-inflammatory response is clearly present also in other areas of the

brain, but in the striatum the presence of a higher amount of activated microglia in HD patients with respect to NC is a well-known fact. The implicit assumption that ALL pixels of the putamen were "ill" was made; the increase in the number of pixels classified as such, therefore, was considered as an increase of specificity of the analyzed method. The cerebellum was considered in order to study what happens in regions in which there is, on average, a higher pattern of microglial activation in HD but with a lower extent than in striatum.

To quantify Sp a one-sided t-test was performed for each pixel of putamen and cerebellum: the comparison was made between each NC and the group made by the other 11 NC. Under the implicit assumption that all pixels of NC are "healthy", and should be classified as such, the number of "active" pixels was employed to quantify specificity: clearly the lower this number, the higher the specificity and the better the algorithm were.

### 4.9.3 Results: sensitivity

Figures (4.10) and (4.11) display the percentage of active pixels ( $p$  lower than 0.01) of cerebellum and putamen, for SRTMV and the 4 Bayesian approaches, determined using a standard t-test and Wilcoxon rank-sum test, respectively.

The figures highlight a very significant fact: except for LRR, which, as usual, gave results identical to those of SRTMV, the other Bayesian approaches yielded *incredible increases* of sensitivity. The number of "active" pixels in putamen, which as explained before, can be considered as a reliable indicator of sensitivity, became more than 3 times, 5 times and 8 times respectively with NLRR3a, GTS, NLRR3b if compared to SRTMV. Almost identical results were obtained when non-parametric Wilcoxon was used. Similar trends were noticed for the cerebellum. The most relevant increase took place when NLRR3b was employed; despite this fact, GTS emerged as the optimal method. In fact, the high number of "active" pixels detected in cerebellum by NLRR3b was, at least in part, caused by false positives; GTS, besides, gave higher percentage of active pixels for putamen with both tests, whereas NLRR3b gave results of difficult interpretation (cerebellum more active than putamen) when Wilcoxon rank-sum test was employed.

### 4.9.4 Results: specificity

Table (4.1) contains the percentage of pixels in the cerebellum classified as "ill" for each NC computed through a one-tailed t-test. At every pixel, each NC BP was compared to the group made by the other 11 NC: under the assumption that all pixels of NC are "healthy", and should be classified as such, low numbers are indicators of good performances in terms of specificity. As usual LRR gave results almost identical to those of the standard SRTMV; apart the anomalous behaviour of NC12 -all methods gave more than 10% of "ill" pixels according to the described criterion for this subject- GTS and, especially, the two versions of NLRR gave a remarkable increase of specificity which is evident from the reduced number of pixels classified as "ill".

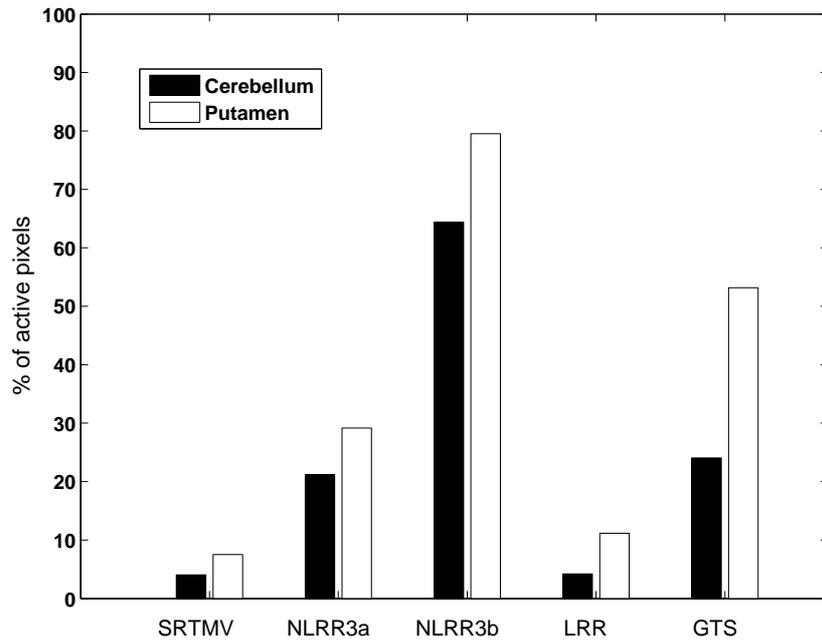


Figure 4.10: Percentage of active pixels ( $p$  lower than 0.01) are displayed for cerebellum and putamen for SRTMV and the 4 Bayesian approaches.  $T$ -test was used to compute  $p$ -values.

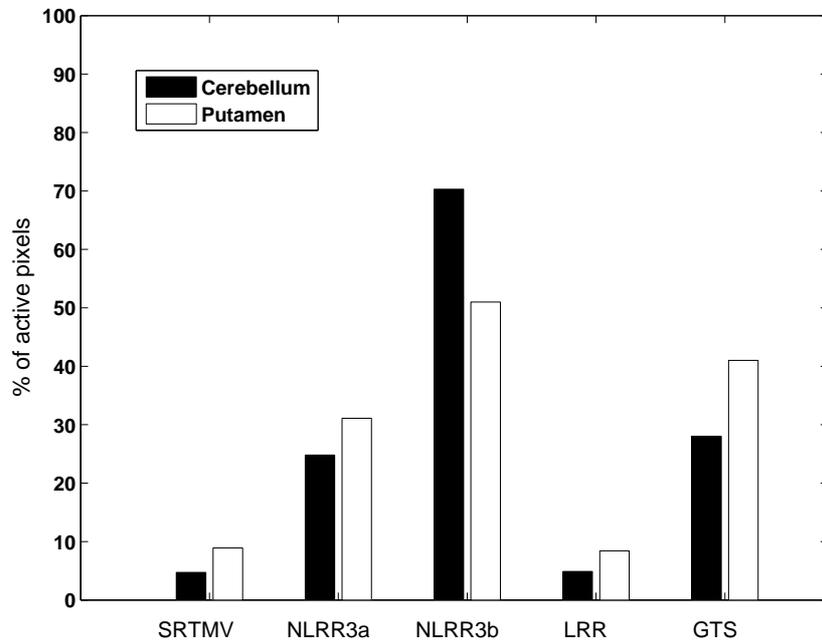


Figure 4.11: Percentage of active pixels ( $p$  lower than 0.01) are displayed for cerebellum and putamen for SRTMV and the 4 Bayesian approaches. Wilcoxon rank-sum test was used to compute  $p$ -values.

	SRTMV	NLRR3a	NLRR3b	LRR	GTS
<i>NC1</i>	1.83	0	0.08	1.79	0.05
<i>NC2</i>	3.41	0.62	1.43	2.84	5.32
<i>NC3</i>	2.54	1.95	0.32	2.49	0.39
<i>NC4</i>	2.26	0	0.06	2.33	0.1
<i>NC5</i>	1.28	0	0.01	1.58	0.05
<i>NC6</i>	1.76	0	0.04	2.06	0.13
<i>NC7</i>	1.37	0.15	0.17	1.53	0.19
<i>NC8</i>	4.6	7.08	1.70	4.73	0.74
<i>NC9</i>	3.07	0.08	2.29	3.02	0.51
<i>NC10</i>	4.73	0	0.04	4.72	0.39
<i>NC11</i>	7.13	4.85	0.81	6.92	8.81
<i>NC12</i>	11.51	22.55	24.35	10.19	22.81

Table 4.1: *Percentage of pixels in the cerebellum classified as "ill" for each NC computed through a one-tailed t-test. Low number are indicators of good performances.*

## 4.10 Summary of main results on [ $^{11}\text{C}$ ]-(**R**)-PK11195

The analysis of a real rich [ $^{11}\text{C}$ ]-(**R**)-PK11195 data set comprising subjects suffering from Huntington's disease HD and age-matched normal controls NC showed how RR and GTS yielded parametric maps much more regular at visual inspection with respect to SRTMV, which is SRTM modified to include the vascular components. NLRR was implemented in two different versions. Clusterization of an appropriate slice, followed by CRM was employed to compute priors in NLRR3a; in NLRR3b clusterization was combined to the use of the anatomical atlas to assign to pixels of a specific ROI the parameters of that ROI as prior. Ridge regression was employed also in a version linearized through the application of the basis function method. GTS was applied separately to each kinetic parameter, with variance estimates obtained applying the linear transfer model to the preliminary SRTMV estimates. In this way a different number of iterations was employed for each parameter to satisfy the 1%-90% criterion, according to the idea that a poorly estimated parameter requires more iterations to stabilize.

Maps obtained with any Bayesian method were, as expected, more regular at visual inspection than SRTMV maps. LRR maps were the most similar to original SRTMV maps and seemed to have obtained from SRTMV maps through a process of smoothing. In NLRR3b and GTS maps the employment of the anatomical information of the atlas could clearly be seen. The structure of the cerebellum, for instance, could be easily singled out from any parametric map, and especially from BP and  $V_b$  maps. The possibility to single out anatomical structures from parametric maps clearly depended on the parameter and slice considered.  $V_b$  maps were those which were modified the most by the application of Bayesian approaches: the reason is that these were the noisiest and least regular maps.

Analysis of the correlation between BPs estimated with SRTMV, on one side, and BPs computed with the various Bayesian methods, on the other, showed once again that LRR gave BPs very well correlated to SRTMV whereas NLRR methods corrected for negative or unreasonably high BPs, despite showing a certain tendency of underestimating high BPs. GTS provided the best results, correcting for non-physiological values without underestimating too much high BP values and keeping the middle values substantially unchanged. The standard deviation of BPs distribution within a given ROI, besides, was dramatically reduced when NLRR approaches and GTS were analyzed.

Voxel-wise differentiation between healthy and ill subjects remarkably increased when NLRR and GTS were applied, which is a very positive fact from a diagnostic point of view. Analysis was focused on putamen, a region in which it is known HD patients have a higher level of activated microglia and in which, consequently, a higher level of [ $^{11}\text{C}$ ]-(**R**)-PK11195 binding is present. In fact, in this region the increment of voxel-wise statistical difference is an equivocal sign of good performance, while in other ROIs this may be related to an increase of the so-called false positives (voxels which are not "active" but which are recognized as such). The number of "active" pixels in putamen, which as explained before, can be considered as a reliable indicator of sensitivity, became more than 3 times, 5 times and 8 times respectively with NLRR3a, GTS, NLRR3b if compared to SRTMV. Almost identical results were obtained when non-parametric Wilcoxon was used and similar trends were noticed for the cerebellum. The most relevant increase took place when NLRR3b was employed; despite this fact, GTS emerged as the optimal method. In fact, the high number of "active" pixels detected in cerebellum by NLRR3b was, at least in part, caused by false positives; GTS,

besides, gave higher percentage of active pixels for putamen with both tests, whereas NLRR3b gave results of difficult interpretation (cerebellum more active than putamen) when Wilcoxon rank-sum test was employed. The increase of sensitivity, fortunately, was not followed by a decrease of specificity.

## 4.11 Discussion

After testing Ridge Regression RR and GTS on simulated data, in this chapter performances of these methods on real data were assessed.

First, an interesting application of NLRR to the 5K model applied to a [ $^{18}\text{F}$ ]FDG data set was presented. NLRR allowed to compute parameter estimates which could not be obtained at all when pixel TACs only were employed, because parameter identification failed to converge. Quality of parametric maps, however, varied according to the kinetic parameter.

The analysis of the [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 data set yielded, as main outcome, that voxel-wise differentiation between healthy and ill subjects remarkably increased when NLRR and GTS were applied. The reason is that Bayesian methods have an impact on parametric maps similar to that of smoothing, which is known to increase statistical difference due to the augmented level of regularity. In fact, smoothing is commonly performed before the application of t-test or non-parametric tests at voxel level to increase statistical power; the level of smoothing, however, is totally arbitrary leading to increases of sensitivity which are consequently highly dependent on the level of smoothing. The application of GTS and NLRR eliminates this element of arbitrariness, leading at the same time to the desired increase of sensitivity.

The analysis described in this chapter showed how LRR was the least invasive Bayesian approach, as all indices computed from LRR maps were similar to those of the original maps. NLRR methods yielded in general good performances, but sometimes their results were of difficult interpretation. GTS globally emerged as the optimal approach: parametric maps appeared smoother and characterized by anatomical information, non-physiological values were corrected while good correlation with original estimates was conserved, and, in the end, a significant and regular increase of sensitivity was noticed. GTS, therefore, outclassed NLRR procedures mainly in that it provided more robust results. The only drawback of GTS, in this context, was that pixels which were not assigned by the atlas to a specific ROI were left unchanged by the algorithm.

The anatomical atlas was heavily used to implement GTS and NLRR3b. What if this information can not be used? The MRI of the scanned subject, for instance, may not be available, or, if high resolution scanners are used, one may not want to perform normalization in order not to lose precious spatial resolution. In this case NLRR3a, with priors derived from clusters' TACs can be used instead of NLRR3b. The segmentation of each slice into a certain number of more homogeneous "populations" before the application of GTS is, on the other side, the advised preliminary step before the application of population approaches. In this case, clearly, no anatomical information will be present in final parametric maps, and results, in general, will depend, at least partially, on the parameter considered and number of clusters used.

## CONCLUSIONS

The generation of parametric maps is an important step in the analysis of dynamic PET data. Pixel-by-pixel maps allow, for instance, to single out pathological areas which may not be characterized if ROI analysis only is performed. This happens because at ROI level only the mean of pixel TACs is considered, and therefore differences due to abnormal conditions may cancel out, or at least be significantly attenuated, when averaging the TACs of pixels of a given ROI. ROI analysis, besides, requires to define a-priori the set of anatomical regions to be analyzed; it may happen that a pathological event (a lesion, an occlusion, the presence of a small tumoral mass) is located in areas outside those comprised in the analysis. Parametric maps, in conclusion, are of paramount importance for their ability to locate regions characterized by pathological or abnormal conditions.

Analysis at pixel level, on the other side, is inevitably associated to a high computational cost; with the performance of current computers and with the development of ad-hoc algorithms (Basis Function Method, Generalized Linear Least Squares), however, this issue seldom constitutes a problem.

A harder challenge is posed by the high noise of pixel TACs which may give rise to unreliable or non-physiological values for the resulting kinetic parameters. To address this issue Bayesian approaches can be used successfully. Ridge regression (RR) has been the only method proposed in PET to overcome this problem in a Bayesian framework. In this thesis a comprehensive comparison between ridge regression methods was performed, which was never done in the past. Besides, several *new* Non Linear Ridge Regression algorithms based on linearization of the non linear problem and/or on the use of cluster analysis to compute priors were proposed and systematically compared to each other. In addition, a novel Two-Stage algorithm (GTS) was proposed and applied *for the first time* to PET parametric imaging. The most appealing feature of GTS is the fact that it can be employed with any tracer and model and has therefore a very wide range of applicability. The method has also a low computational cost which makes it suitable for the application to parametric imaging, even with the current high resolution scanners, for which the number of pixels can reach the number of one million.

Chapters 2 and 3 showed through the use of simulated data that RR and GTS always give rise to the decreases of Root Mean Square Error leaving BIAS substantially unchanged or even reducing it. The improvements may range to low -if the original maps are already of good quality- to very relevant; in any case, no decrease of performances was ever noticed in terms of BIAS and RMSE.

In contrast with linear ridge regression, for which one and only one robust algorithm is available, several versions of non-linear ridge regression were tested. The approach which was selected as the optimal requires to compute priors at pixel level from the preliminary fit of the kinetic model of interest to cluster TACs. In real application to normalized 3D images, clusterization can be combined by the use of anatomical atlases which can be employed to associate to most pixels the parameters of the ROI they belong to.

GTS has two user-dependent settings which can strongly influence final results. Performances of GTS didn't show to be significantly dependent on the method employed to stop iterations; in any case, the criterion to end when the relative difference of the parameter of interest was lower than 1% for at least 90% of the individuals in the con-

sidered population proved to be quite robust in every situation tested. Results, on the other side, appeared to be more dependent on the segmentation technique employed. If the image is normalized, stereotaxic atlases can be employed to create sub-populations made of pixels belonging to the same ROI; in this way, besides, anatomical information is incorporated into parametric maps. Sometimes, however, normalization can not be performed because no MRI is available or because this would cause an unwanted loss of spatial resolution if the image was acquired on a high-resolution scanner. In this case a segmentation of each slice of the image into a small number of clusters, performed on the basis of the parameter(s) of interest, is a good alternative solution which allows to perform GTS on populations relatively homogeneous.

Analysis of a real rich [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 data set showed how RR and GTS yielded parametric maps much more regular at visual inspection with respect to SRTMV, which is SRTM modified to include the vascular components. This was totally expected as Bayesian approaches have a regularizing effect. Linear ridge regression maps significantly resembled original SRTMV maps, showing a higher degree of smoothness. Non-Linear Ridge Regression and GTS maps, on the other side, were remarkably different from original maps. Differences were evident not only in the spatial distribution of kinetic parameters but also in terms of their numerical values: negative or unreasonably high values of binding potentials, for instance, were corrected by the application of these algorithms, even though a certain tendency of underestimating high BPs was noticed.

An important outcome was that voxel-wise differentiation between healthy and ill subjects increased of 3 to 6 times when applying NLRR and GTS, which is a positive fact from a diagnostic point of view. The reason is that Bayesian methods have an impact on parametric maps similar to that of smoothing, which is known to increase statistical difference due to the increased level of regularity. In this context focus was put on putamen, as in Huntington's disease it is well-known that this a region severely affected by the disease and therefore increase of statistical differences in this ROI was considered as an unequivocal sign of increased sensitivity. Fortunately no decrease of specificity took place, making these approaches optimal candidates when interest is addressed in diagnosis from parametric maps.

The comparison between RR and GTS on the basis of RMSE showed that sometimes GTS performed slightly better and sometimes the contrary happened; in any case differences were never relevant. From the analysis of [ $^{11}\text{C}$ ]-( $\text{R}$ )-PK11195 GTS, in general, emerged as the most powerful tool yielding attractive final maps which incorporated the a-priori anatomical information and providing results which appeared more robust than NLRR results. The application of GTS, however, can be sometimes risky as wrong choices regarding the construction of populations may lead to unreliable final parametric maps.

The conclusion is that the application of ridge regression or of GTS is a powerful tool for improving PET parametric maps. These methods can be employed with any tracer and model, and are therefore very appealing for their wide range of applicability. The amount of improvement significantly depends on the tracer, model, noise level, estimation method: the general simple principle is that, within reasonable limits, the worst the initial estimates are, the higher the improvement is.

# Bibliography

- [1] Wu.Y., Carson.R. *Noise reduction in the Simplified Reference Tissue Model neuroreceptor functional imaging* Journal of Cerebral Blood Flow and Metabolism 22(2002),1440-1452
- [2] Lammerstma.AA., Hume.S. *Simplified reference tissue model for PET receptor studies.* Neuroimage 4(1996),153-158
- [3] Irie.T., Fukushi.K, Namba.H., et al. *Brain acetylcholinesterase activity:validation of a PET tracer in a rat model of Alzheimer's disease* Journal of Nuclear Medicine 37(1996),649-655
- [4] Herholz.K., Lercher.M., Wienhard.K., et al. *PET measurement of cerebral acetylcholine esterase activity without blood sampling* European Journal of Nuclear Medicine 28(2001),472-477
- [5] Nagatsuka.S., Fukushi.K., Shinotoh.H., et al. *Kinetic analysis of [11C]MP4A using a high radioactivity brain region that represents an integrated input function for measurement of cerebral acetylcholinesterase activity without arterial activity using N-[11C]methylpiperidin-4-yl acetate without arterial blood sampling.* J Cereb Blood Flow and Metab 24(2001),600-611
- [6] O'Sullivan.F., Saha.A. *Use of ridge regression for improved estimation of kinetic constants from PET data.* IEEE Transactions on Medical Imaging 18(1999),115-125
- [7] Zhou.Y., Huang.S.C., Bergsneider.B., *Linear ridge regression with spatial constraint for generation of parametric images in dynamic positron emission tomography studies* IEEE Transactions on Nuclear Science 48(2001),125-130
- [8] Zhou.Y., Endres C., Brasic. J.R., et al. *Linear regression with spatial constraint to generate parametric images of ligand-receptor dynamic PET studies with a simplified reference tissue model* Neuroimage 18(2003),975-989.
- [9] Zhou.Y., Huang.S.C., Bergsneider.B., Wong.D., *Improved parametric image generation using spatial-temporal analysis of dynamic PET studies* Neuroimage 15(2002),697-707
- [10] Kimura.Y.,Hsu.H.Toyama.H.et al. *Improved signal-to-noise-ratio in parametric images by cluster analysis* Neuroimage 9(1999),554-561.
- [11] Kimura.Y., Senda.M., Alpert.N. *Fast formation of statistically reliable FDG parametric images based on clustering and principal components.* Physics in Medicine and Biology, 47[3](2002),455-458

- [12] Kimura.Y., Naganawa.M., Yamaguchi.J., et al. *MAP-based kinetic analysis for voxel-by-voxel compartment model estimation: detailed imaging of the cerebral glucose metabolism using FDG*. Neuroimage 29(2006),1203-1211.
- [13] Gunn.R., Gunn.S., Cunningham.V. *Positron Emission Tomography Compartmental Models* Journal of Cerebral Blood Flow and Metabolism 21(2001),635-652
- [14] Blomqvist.G. *On the construction of functional maps in positron emission tomography* Journal of Cerebral Blood Flow and Metabolism 4(1984),629-632
- [15] Tomasi.G., Bertoldo.A., Cobelli.C. *Parametric imaging of acetylcholinesterase activity with PET: evaluation of different methods*. Proceedings of IFAC conference of Modelling in Biological Systems, Reims 2006
- [16] Hartley.H. *The modified Gauss-Newton method for the fitting of non-linear regression function by least squares*. Technometric 3(1961),269
- [17] Feng.D., Huang.S.C., Wang.Z.Z., Ho.D. *An unbiased parametric imaging algorithm for nonuniformly sampled biomedical system parameter estimation*. IEEE Transactions on Medical Imaging 15(1996),512-519
- [18] Guo.H., Renaut.R., Chen.K., Reiman.E. *Clustering huge data sets for parametric PET images* BioSystems 71(2003),81-92
- [19] Windischberger.C., Barth.M., Lamm.C., et al. *Fuzzy cluster analysis of high-field functional MRI data* Artificial Intelligence in Medicine 29(2003),203-223
- [20] Pike.V.W., McCarron.J.A., Lammertsma.A.A., Osman.S., et al. *Exquisite delineation of 5-HT<sub>1A</sub> receptors in human brain with PET and [carbonyl-<sup>11</sup>C]WAY-100635*. European Journal of Pharmacology 301(1996),R5-R7.
- [21] Carson.E. and Cobelli.C. *Modelling methodology for physiology and medicine*. Elsevier,2000,77-106
- [22] Bailer.U., Frank.G., Henry.S., Price.J. et al. *Exaggerated 5-HT<sub>1A</sub> but normal 5-HT<sub>2A</sub> receptor activity in individuals ill with anorexia nervosa*. Biol Psychiatry 61(9)(2007),1090-1099
- [23] Steimer.J., Mallet.A., Golmard.JL., Boisvieux.JF. *Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the non linear mixed effect model*. Drug Metabolism Reviews 15(1984),265-292.
- [24] Davidian.M. ,Giltinan.D. *Nonlinear models for repeated measurement data* New York,Chapman&Hall,1995,125-150
- [25] Bertoldo.A., Sparacino.G., Cobelli.C. *"Population approach" improves parameter estimation of kinetic models from dynamic PET data*. IEEE Transactions of Medical Imaging, 23(2004),297-306
- [26] Tomasi.G., Bertoldo.A., Cobelli.C. *PET parametric imaging of improved by Global-Two-Stage method*. "Annals of Biomedical Engineering", Positive first review, July 2007

- [27] Bertoldo.A., Pencek.R.R., Azuma.K., et al. *Interactions between delivery, transport, and phosphorylation of glucose in governing uptake into human skeletal muscle* Diabetes, 55(2006),3028-3037
- [28] Bertoldo.A., Peltoniemi.P., Oikonen.V., et al. *Kinetic modeling of [ $^{18}\text{F}$ ]FDG in skeletal muscle by PET: a four-compartment five-rate-constant model* American Journal of Physiology, 281(2001),E524-E536.
- [29] Hertz L. *Binding characteristics of the receptor and coupling to transport proteins*. In: Giessen-Crouse E, ed. Peripheral benzodiazepine receptors. London: Academic Press; 1993:27-51.
- [30] Gebicke-Haerter P.J. *Microarrays and expression profiling in microglia research and in inflammatory brain disorders*. J Neurosci Res. 81(2005):327-341.
- [31] Banati RB. *Visualising microglial activation in vivo*. Glia. 40(2002):206-217.
- [32] Kropholler.MA., Boellaard.R., Schuitemaker.A., et al. *Evaluation of reference tissue models for the analysis of [ $^{11}\text{C}$ ](R)-PK11195 studies*.Journal of Cerebral Blood Flow and Metabolism, 26[11](2006),1431-1441
- [33] Anderson.AN., Pavese.N., Edison.P., et al. *A systematic comparison of kinetic modelling methods generating parametric maps for [ $^{11}\text{C}$ ](R)-PK11195* Neuroimage. 36[1](2007),28-37
- [34] Kropholler.M., Boellaard.R., Schuitemaker.A., et al. *Development of a tracer kinetic plasma input model for [ $^{11}\text{C}$ ](R)-PK11195 brain studies*. Journal of Cerebral Blood Flow and Metabolism, 25[7](2005),842-851
- [35] Turkheimer.FE., Edison.P., Pavese.N., et al. *Reference and Target Region Modeling of [ $^{11}\text{C}$ ](R)-PK11195 Brain Studies* Journal of Cerebral Blood Flow and Metabolism 48(2007):158-167.
- [36] Pavese.N., Gerhard.A., Tai.YF., et al. *Microglial activation correlates with severity in Huntington disease: a clinical and PET study*. Neurology 66(2006):1638-1643.
- [37] Hammers.A, Allom.R, Koeppe.MJ.,et al. *Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe*. Human Brain Mapping 19(2003):224-247.
- [38] Gunn.R., Lammerstma.A., Hume.P., Cunnigham.V.J. *Parametric imaging of ligand and receptor binding in PET using a simplified reference region model* Neuroimage 6(1997),279-287.
- [39] Bertoldo.A, Pietra.L, Moresco.RM., et al. *Quantitative in vivo imaging of microglia activation using [ $^{11}\text{C}$ ](R)-PK11195 and two reference tissue models*. Neuroimage. (2006) 31:T79.
- [40] Tomasi.G., Edison.P., Bertoldo.A., et al. *Novel reference region modelling reveals increased microglial and reduced vasculature binding of [ $^{11}\text{C}$ ](R)-PK11195 in Alzheimer's disease patients*. submitted to the Journal of Nuclear Medicine, Septmeber 2007

- [41] Liptrot.M., Adams.KH., Martiny.L., et al. *Cluster analysis in kinetic modelling of the brain: a noninvasive alternative to arterial sampling*. Neuroimage 21[2](2004):483-493.
- [42] Van der Weerd AP., Klein.LJ., Boellaard.R., et al. *Image-derived input functions for determination of MRGlu in cardiac [<sup>18</sup>F]FDG PET scans*. Journal of Nuclear Medicine 42[11](2001):1622-1629.
- [43] Wong.KP., Feng.D., Meikle.SR., Fulham. MJ. *Simultaneous estimation of physiological parameters and the input function—in vivo PET data*. IEEE Transactions on Information Technology in Biomedicine 5[1](2001):67-76.
- [44] Jay.L.Devore. *Probability and Statistics for Engineering and the Sciences*, 5th Edition, Pacific Grove(California),Duxbury Editions,2005,659.

# List of Figures

1	<i>Il RMSE per il parametro <math>k_3</math>, mediato su tutti i pixel, ottenuto con RLS, SRRSC e GRRSC è rappresentato per le fette n°27(sinistra) e n°20(destra).</i>	6
2	<i>Mappe di BP per la fetta n°21 di un soggetto HD. La mappa ottenuta con SRTMV (sinistra) è confrontata con quella ottenuta con GTS (destra).</i>	7
3	<i>La percentuale di pixel "malati" di cervelletto e putamen (p-value minore di 0.01) è raffigurata per SRTMV e per 4 approcci Bayesiani. Un one-tailed Wilcoxon rank-sum test è stato qui usato per calcolare i p-values.</i>	7
1.1	<i>The summed image for an healthy subject for a slice with high AChE activity is displayed.</i>	19
1.2	<i><math>k_3</math> parametric map obtained using (1.14) is displayed.</i>	20
1.3	<i><math>k_3</math> parametric map obtained using (1.15) is displayed.</i>	21
2.1	<i><math>k_3</math> BIAS averaged over all pixels, obtained with RLS, SRRSC and GRRSC is displayed for slice n°27(left) and n°20(right).</i>	28
2.2	<i><math>k_3</math> RMSE averaged over all pixels, obtained with RLS, SRRSC and GRRSC is displayed for slice n°27(left) and n°20(right).</i>	28
2.3	<i>The parametric map of <math>k_3</math> of one repetition of the simulated slice n°27 obtained with RLS(left), SRRSC(left) and GRRSC(down) is displayed.</i>	29
2.4	<i>The parametric map of the relative delivery R of one repetition of the simulated slice n°27 obtained with RLS(left)and GRRSC(right) is displayed.</i>	29
2.5	<i>Average RMSE(left) and BIAS(right) of <math>k_3</math> obtained with RLS, SRRSC and GRRSC for the portion of slice n°20 containing the thalamus are displayed.</i>	30
2.6	<i>Example of how a dynamic MP4A image was segmented employing hierarchical cluster analysis with 6 clusters</i>	33
2.7	<i>BIAS averaged over all pixels for the four kinetic parameters of the simulated <math>[^{11}C]WAY100,635</math> data set. Results are displayed respectively for WNLLS, NLRR1, NLRR2 with CRM, and NLRR3 with two different criteria for computing priors (direct attribution and CRM.)</i>	38
2.8	<i>RMSE averaged over all pixels for the four kinetic parameters of the simulated <math>[^{11}C]WAY100,635</math> data set. Results are displayed respectively for WNLLS, NLRR1, NLRR2 with CRM, and NLRR3 with two different criteria for computing priors (direct attribution and CRM.)</i>	38
2.9	<i>"True" <math>K_1</math> parametric map (upper left), NLLS map (upper right), NLRR2-CRM map (lower left), and NLRR3-CRM map(lower right) are displayed for one repetition of the simulated <math>[^{11}C]WAY100,635</math> data set.</i>	39

2.10	<i>A randomly chosen pixel TAC is displayed; notice the high level of noise, which makes the application of the 5K model prohibitive. . . . .</i>	42
2.11	<i>The segmentation of pixels into 3 clusters by means of hierarchical clusterization is displayed. . . . .</i>	42
2.12	<i>The parameter <math>K_1</math> is displayed. On the left is the initial parametric image that was used as prior and on the right the output of the NLRR3. Note that different scales were employed. . . . .</i>	43
2.13	<i>The parameter <math>k_4</math> is displayed. On the left is the initial parametric image that was used as prior and on the right the output of the NLRR3. Note that different scales were employed. . . . .</i>	43
3.1	<i>RMSE for <math>K_i</math> [ml/ml/min], computed first with WNLLS and subsequently with GTS and ITS, is displayed for the pixel level noise simulation. . . .</i>	56
3.2	<i>BIAS for <math>K_i</math> [ml/ml/min], computed first with WNLLS and subsequently with GTS and ITS, is displayed for the pixel level noise simulation. . . .</i>	56
3.3	<i>RMSE for <math>K_1</math> [ml/ml/min], , computed first with WNLLS and subsequently with GTS, is displayed for the high and low levels of population variability. . . . .</i>	57
3.4	<i>RMSE of <math>K_i</math> [ml/ml/min], computed with GTS, is displayed for different population sizes. . . . .</i>	58
3.5	<i>BIAS of <math>K_i</math> [ml/ml/min], computed with GTS, is displayed for different population sizes. . . . .</i>	58
3.6	<i>The percentage reduction of RMSE between GTS and WNLLS, as expressed by (3.17), is displayed for <math>K_1</math>, <math>k_2</math> and <math>k_3</math> for the 3 different scan lengths (90, 60 and 45 minutes) employed to compute WNLLS estimates. . . .</i>	59
3.7	<i>The segmentation of one simulated slice of the insulin data set into 3 clusters is displayed. Above is the segmentation based on the irreversible uptake rate constant, below hierarchical clusterization based on pixel TACs. . . .</i>	62
3.8	<i>BIAS for the simulation in the basal state is displayed. The dotted line refers to Paltak method, dashed lines correspond to GTS-based methods when a convergence criterion was used, and the solid lines represent performances of GTSpop and GTSraw as a function of the iterations. . . .</i>	63
3.9	<i>RMSE for the simulation in the basal state is displayed. The dotted line refers to Paltak method, dashed lines correspond to GTS-based methods when a convergence criterion was used, and the solid lines represent performances of GTSpop and GTSraw as a function of the iterations. . . .</i>	64
3.10	<i>BIAS for the insulin data set is displayed. The dotted line refers to Patlak method, the dashed line corresponds to GTS-based methods when a convergence criterion was used (GTSpop* gave a result equal to GTSraw*), and the solid lines represent performances of GTSpop and GTSraw as a function of the iterations. . . . .</i>	64
3.11	<i>RMSE for the insulin data set is displayed. The dotted line refers to Paltak method, dashed lines correspond to GTS-based methods when a convergence criterion was used, and the solid lines represent performances of GTSpop and GTSraw as a function of the iterations. . . . .</i>	65
3.12	<i>The irreversible rate constant for one simulated slice of the basal data set, obtained respectively with Patlak method (left), GTSpop (right) and linear ridge regression (down), is displayed. . . . .</i>	66

3.13	<i>BIAS for <math>[^{11}\text{C}]\text{WAY100,635}</math> data is displayed. Black bars refer to standard WNLLS, gray to GTSraw and white to GTSpop, with a subdivision of the population into 3 clusters. . . . .</i>	68
3.14	<i>RMSE for <math>[^{11}\text{C}]\text{WAY100,635}</math> data is displayed. Black bars refer to standard WNLLS, gray to GTSraw and white to GTSpop with a subdivision of the population into 3 clusters. . . . .</i>	69
3.15	<i>The "true" <math>k_3</math> parametric map (upper left), and those obtained for one simulated slice using GTSraw (upper right), GTSpop (lower left) and NLLRR3 (lower right) are displayed, using the same color scale for all the figures. . . . .</i>	71
3.16	<i>BIAS and RMSE averaged over the 1000 simulated subjects are displayed for standard WNLLS (black), GRRSC (gray) and GTS (white) with <math>N=10</math> and 3 different reduced scan times. . . . .</i>	74
4.1	<i>Parametric maps of <math>R</math> for one slice of a HD patient (slice <math>n^\circ 21</math>). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively). . . . .</i>	88
4.2	<i>Parametric maps of <math>k_2</math> for one slice of a HD patient (slice <math>n^\circ 21</math>). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively). . . . .</i>	89
4.3	<i>Parametric maps of BP for one slice of a HD patient (slice <math>n^\circ 21</math>). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively). . . . .</i>	90
4.4	<i>Parametric maps of <math>V_b</math> for one slice of a HD patient (slice <math>n^\circ 21</math>). The basic SRTMVb map is compared to NLRRa and NLRRb maps (middle left and right respectively) and to LRR and GTS maps (lower left and lower right respectively). . . . .</i>	91
4.5	<i>The coefficient <math>\eta</math> is displayed for each of the 4 parameters and each of the 5 methods used to generate parametric maps. . . . .</i>	93
4.6	<i>Plot of the relative delivery <math>R</math> evaluated at ROI and pixel level (<math>x</math> and <math>y</math> axis respectively) is displayed for SRTMV (left) and GTS (right). . . . .</i>	94
4.7	<i>Plot of the blood fraction <math>V_b</math> evaluated at ROI and pixel level (<math>x</math> and <math>y</math> axis respectively) is displayed for SRTMV (left) and GTS (right). . . . .</i>	94
4.8	<i>Plot of the correlation between SRTMV BPs and the corresponding BPs estimated with the 4 Bayesian approaches analyzed for pixels of cerebellum of a HD patient. The identity line is displayed in yellow. . . . .</i>	96
4.9	<i>Mean BPs, estimated with SRTMV and the 4 Bayesian approaches, and the corresponding standard deviations for 3 normal controls NC and 3 HD patients are displayed in the upper and lower part of the figure respectively. The BPs belong to pixels of putamen. . . . .</i>	98
4.10	<i>Percentage of active pixels (<math>p</math> lower than 0.01) are displayed for cerebellum and putamen for SRTMV and the 4 Bayesian approaches. T-test was used to compute <math>p</math>-values. . . . .</i>	101

4.11 *Percentage of active pixels ( $p$  lower than 0.01) are displayed for cerebellum and putamen for SRTMV and the 4 Bayesian approaches. Wilcoxon rank-sum test was used to compute  $p$ -values. . . . .* 101

# List of Tables

2.1	<i>RMSE of NLRR3 with direct attribution of the priors for different values of the number of clusters employed for the computation of the priors. . .</i>	39
2.2	<i>BIAS of NLRR3 with direct attribution of the priors for different values of the number of clusters employed for the computation of the priors. . .</i>	39
3.1	<i>Comparison of performances of GTS and linear ridge regression on the simulated [<sup>18</sup>F]FDG data sets . . . . .</i>	65
3.2	<i>Comparison of performances of GTSpop for different number of sub-populations (n=3,6,9) on the simulated [<sup>11</sup>C]WAY100,635 data set: average RMSE is displayed. . . . .</i>	67
3.3	<i>Comparison of performances of GTSpop for different number of sub-populations (n=3,6,9) on the simulated [<sup>11</sup>C]WAY100,635 data set: average BIAS is displayed. . . . .</i>	67
3.4	<i>Comparison of performances of GTS (raw and pop) and non linear ridge regression (version 3 with CRM to compute priors) on the simulated [<sup>11</sup>C]WAY100,635 data set: average RMSE is displayed. . . . .</i>	69
3.5	<i>Comparison of performances of GTS (raw and pop) and non linear ridge regression (version 3 with CRM to compute priors) on the simulated [<sup>11</sup>C]WAY100,635 data set: average BIAS is displayed. . . . .</i>	70
4.1	<i>Percentage of pixels in the cerebellum classified as "ill" for each NC computed through a one-tailed t-test. Low number are indicators of good performances. . . . .</i>	102