Università degli Studi di Padova

Facoltà di Ingegneria

Corso di Laurea Magistrale in Bioingegneria

Tesi di laurea in Bioingegneria

Methods and models for the characterization of arterial input function in dynamic PET studies

Relatore: Alessandra Bertoldo
Correlatori: Gaia Rizzo
Mattia Veronese

Laureando: Matteo Tonietto

15 Ottobre 2012

Contents

Chapter 1

Introduction

Positron Emission Tomography (PET) allows to image the biodistribution and kinetics of radiopharmaceuticals in brain with great accuracy and minimum invasiveness [1]. Interpreting these data by means of specific mathematical models can improve the kind and quality of information that can be extracted allowing the quantification of various physiological parameters as, for example, the penetration of the blood-brain barrier by the ligand; the volume of distribution of the tracer; the binding potential; the receptor occupancy; the receptor affinity and density in a given region [2]. This information can be used to monitor the progress of neurological diseases or to assess the effects of specific drugs, such as their penetration into target sites, their binding to specific receptors and their effects on the metabolic processes [3].

Various models have been proposed in the past 30 years to convert the radioactive tracer concentrations detected by a PET tomograph in an organ or tissue (more precisely, in a region of interest [ROI] or in a unit of image [pixel or voxel]) into measures of physiological parameters [4]. Although each of them starts from different assumptions, most of them require the knowledge of the concentration of the radioligand inside both the tissue and the arterial plasma. The first constitutes the output of the model and it is obtained directly from the PET images while the second represents its input (it is through the arterial blood that the tracer reaches the tissues) and its measurement is more problematic.

The standard approach to obtain the Arterial Input Function (AIF) consists in the cannulation of the radial artery of the patient, through which blood samples are collected during the scan time. Undoubtedly, this is an invasive technique and, although it is considered safe [5], it is

often cause of discomfort in the patient. Furthermore, the handling and analysis of numerous blood samples require adequate facilities, special care and technical expertise by the operator [6].

Given the issues related with arterial sampling, over the years new methods have been developed to eliminate, or at least reduce, this practice. One of the most promising approach consists in replacing the individual arterial measurements with the average of the AIFs of a population of subjects previously acquired. Before its application, the resulting curve has to be calibrated through individual information (venous samples, covariates,...) of the subject under analysis [7].

This method has been validated almost exclusively with [^{18}F]-FDG [7] [8] and only a few works have been reported using other tracers [9]. One of the aims of this study is to identify an optimal protocol for the application of this technique and, subsequently, to assess its performance highlighting the advantages and risks associated with its use.

The second part of the study is focussed on another method, called SIME (SIMultaneous Estimation). Proposed by Feng and colleagues in [10], this technique consists in estimating the input function and the tissue parameters of multiple ROIs simultaneously. This approach requires a parametric description of the AIF and the model most commonly used for this purpose was proposed again by Feng in [11]. In this thesis the ability of Feng's model to describe correctly the AIF of other tracers than [^{18}F]-FDG (for which it was designed) is assessed and new models are proposed which could possibly constitute an improvement.

Chapter 2

Data sets and quantification methods

2.1 Data sets

In dynamic PET studies radioactive tracers are used to obtain kinetic information of the system under analysis. These are chemical compounds in which one or more atoms have been replaced by a radioisotope (all the tracer considered in this study were marked with the ^{11}C isotope). According to the physiological process under investigation, tracer may either be a direct radiolabeled version of a naturally occurring compound, an analog of a natural compound, or a unique compound, perhaps a radiolabeled drug [3]. The tracers available for this study were:

L-[1-^{11}C]leucine: Leucine is one of the 20 standard amino acids and it has a relevant role in the constitution and maintenance of the muscular tissue. It promotes the protein synthesis in muscles and liver, it slows down the breakdown of muscle proteins and it is involved in the regeneration processes [12]. In PET studies, L-[1-^{11}C]leucine is marked with a radioactive isotope in order to tracer its incorporation inside the proteins, allowing thus the measurement of the protein synthesis [13].

[^{11}C]DASB: ^{11}C-3-Amino-4-(2-dimethylaminomethylphenylsulfanyl)benzonitrile, abbreviated as [^{11}C]DASB, is a compound that binds with high affinity and selectivity to the serotonin (5-hydroxytryptamine, or 5-HT) transporter [14]. This transporter is the primary target for selective serotonin reuptake inhibitors (SSRIs), which in turn are an effective treatment for majordepressive disorder, obsessive compulsive disorder, posttraumatic

stress disorder, generalized anxiety disorder, and other mood and anxiety disorders [15].

[^{11}C]PiB: [^{11}C]PiB stands for Pittsburgh Compound-B, an amyloid-binding radiotracer used in PET scans to image beta-amyloid plaques in neuronal tissue. The measurement of amyloid deposition can be used to improve Alzheimer's disease diagnosis and antiamyloid therapy assessment [16].

[^{11}C](R)-rolipram: Rolipram is an inhibitor of Phosphodiesterase-4 (PDE4), an enzyme that metabolizes cAMP which in turn is thought to play an important role in the mechanism of action of antidepressants [17]. The radiolabeled active enantiomer of rolipram, [^{11}C](R)-rolipram, has been used with PET to image and to quantify PDE4 in brain [18].

The data sets were made available by the Section of Neuroadaption and Protein Metabolism of the National Institute of Mental Health, NIH, USA (L-[1-^{11}C]leucine) [13]; the Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York ([^{11}C]DASB) [15] ; the University of Pittsburgh Alzheimer's Disease Research Center, Pittsburgh, USA ([^{11}C]PiB) [16] and by the Molecular Imaging Branch, National Institute of Mental Health, Bethesda, USA ([^{11}C](R)-rolipram) [18].

2.1.1 Arterial data

The measured AIFs as well as the whole blood time activity curves of all subjects related to all the four data sets were available. All the blood samples were manually collected through the radial artery of the patients according to a sampling protocol which envisaged a higher frequency at the beginning of the exam, when the tracer kinetic was faster, followed by a gradually decreased frequency as the exam reached its end.

Tracer	Number of subjects	Number of samples	Infusion time (s)	Exam duration (min)
L-[1-^{11}C]leucine	50	35-41	120	75-90
[^{11}C]DASB	30	31	30	120
[^{11}C]PiB	20	30	20	90-130
[^{11}C](R)-rolipram	10	23	60	90

Table 2.1. Main characteristics of the arterial data sets

All data are expressed in KBq/ml except the [^{11}C]DASB ones because they were normalized by the dose injected (they are expressed in normalized-dose units). Furthermore, as regards the L-[1-^{11}C]leucine dataset, also the $^{11}CO_2$ (main result of the metabolism of the molecule) activity in blood was available. Finally, all data were corrected for the radioactive decay. In table 2.1 the principal characteristics of each dataset are summarized.

2.1.2 PET images

PET images of three L-[1-^{11}C]leucine and three [^{11}C](R)-rolipram subjects were available. All of them were acquired with high resolution tomographs, allowing to obtain a voxel size of $1.2 \times 1.2 \times 1.2\,mm^3$. The full volumes were acquired in 42 times frames for L-[1-^{11}C]leucine subjects and 33 for [^{11}C](R)-rolipram. All images were corrected for attenuation, scatter and radioactive decay.

Furthermore, magnetic resonance images were co-registrated to the PET images in order to define the following ROIs in all subjects: cerebellum, thalamus, putamen, caudate, hippocampus.

2.2 Quantification methods

2.2.1 Compartmental models

Compartmental models describe the system under analysis postulating a structure for the non accessible portion consisting of distinct compartments which are interconnected by pathways representing fluxes of material and/or biochemical conversions [19]. The structure must have solid grounds in biochemistry and physiology since this kind of models are used to define the relationship between the measurable data and the physiological parameters that affect the uptake and metabolism of the tracer.

L-[1-^{11}C]leucine

In figure 2.1 it is shown the compartmental model that was implemented in this study for describing the behaviour of L-[1-^{11}C]leucine in brain. It is a heterogenous model composed by two homogeneous subregions and a compartment associated to metabolites. However, this is a simplification over the heterogenous model of L-[1-^{11}C]leucine presented in [20] since the

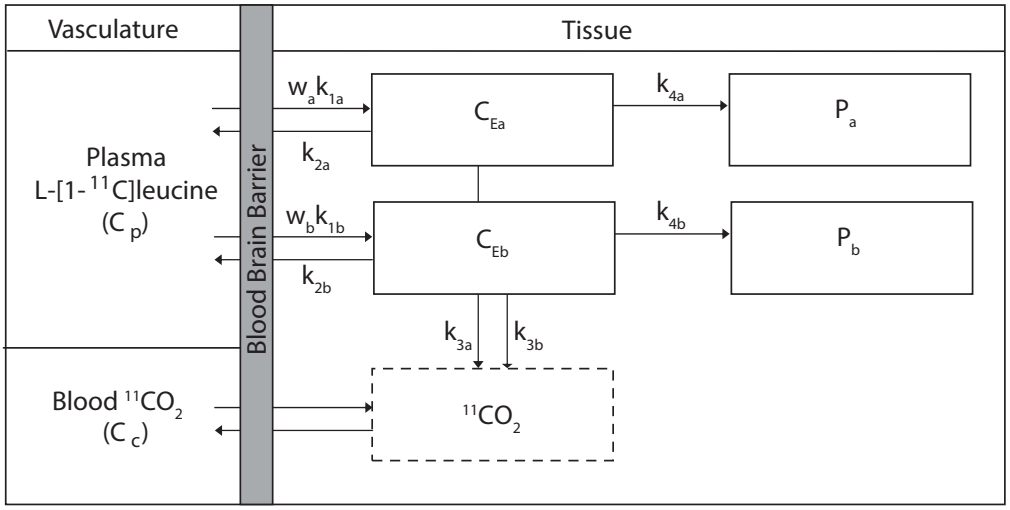


Figure 2.1. Compartmental model of L-[1-^{11}C]leucine: $k_{1(a,b)}$ and $k_{2(a,b)}$ are the rate constants for transport of L-[1-^{11}C]leucine from plasma to tissue and back, respectively; $k_{3(a,b)}$ is the rate constant of the formation process of the metabolite (catabolism, transamination and decarboxylation) and, finally, $k_{4(a,b)}$ is the rate constant of incorporation of L-[1-^{11}C]leucine into proteins.

number of homogeneous subregions is not defined. The choice made reflects the results obtained in [21] where 86% of the ROIs analyzed could be described by two homogeneous subregions (the remaining 14% necessitates of three subregions).

The radiotracer concentration measured by the PET scan (C_m) includes the contribution of both regions of free L-[1-^{11}C]leucine (C_e) and L-[1-^{11}C]leucine incorporated into proteins (P). Each of these homogeneous regions contributes to the heterogeneous tissue for a volume fraction equal, respectively, to w_a and w_b ($w_a \geq 0$, $w_b \geq 0$ and $w_a + w_b = 1$). The PET measure includes also the activity in the blood $V_b \cdot C_b$, where V_b is the portion of the volume occupied by the blood and C_b the concentration of radioligand in whole blood and the main product of the metabolism of the radiotracer, $^{11}CO_2$. Assuming that the metabolite fixing is negligible over the exam time and that its diffusion in brain rapidly equilibrates with arterial blood, its concentration can be approximated by the term $V_d \cdot C_c$, where C_c is the metabolite concentration in whole blood, while V_d is the blood distribution volume in brain, which was set equal to $V_d = 0.41$ [20].

The radiotracer concentration in the tissue is thus given by:

$$C_{tiss}(t) = w_a[C_{Ea}(t) + P_a(t)] + w_b[C_{Eb}(t) + P_b(t)] \tag{2.1}$$

while the measurement equation is:

$$C_m(t) = (1 - V_b) \cdot [C_{tiss}(t) + V_d \cdot C_c(t)] + V_b \cdot C_b(t) \tag{2.2}$$

Unfortunately the w_a and w_b parameters are not uniquely identifiable, i.e. their value can not be determined from the data available. For this reason it was renounced to give a separate description of the two compartments which describe the L-[1-^{11}C]leucine incorporation into proteins in favour of a unique trapping compartment. In fact the model described by equations 2.1 and 2.2 is equivalent to a first-order compartmental model with a trapping component, 2 equilibrating components, one metabolite compartment of known concentration, and one blood component, that is:

$$C_{tiss}(t) = \alpha_0 \int_0^t C_p(z)dz + \alpha_1 \int_0^t C_p(z)e^{-\beta_1(t-z)}dz + \alpha_2 \int_0^t C_p(z)e^{-\beta_2(t-z)}dz \tag{2.3}$$

where the model parameters have been reformulated in the following way:

$$\begin{aligned}
\alpha_0 &= \frac{w_a k_{1a} k_{4a}}{k_{2a} + k_{3a} + k_{4a}} + \frac{w_b k_{1b} k_{4b}}{k_{2b} + k_{3b} + k_{4b}} \\
\alpha_1 &= \frac{w_a k_{1a}(k_{2a} + k_{3a})}{k_{2a} + k_{3a} + k_{4a}} \\
\alpha_2 &= \frac{w_b k_{1b}(k_{2b} + k_{3b})}{k_{2b} + k_{3b} + k_{4b}} \\
\beta_1 &= k_{2a} + k_{3a} + k_{4a} \\
\beta_2 &= k_{2b} + k_{3b} + k_{4b}
\end{aligned} \tag{2.4}$$

As main parameter of interest the trapping rate constant K_I was considered, which is equals to α_0.

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

The [^{11}C](R)-rolipram model is a 2-tissue compartment model and it is depicted in figure 2.2. As for L-[1-^{11}C]leucine, C_p represents the plasmatic tracer concentration, i.e. the input of the system. C_{F+NS} and C_S are the compartments relative to the free and non-specifically bound tracer and the specifically bound tracer respectively [18].

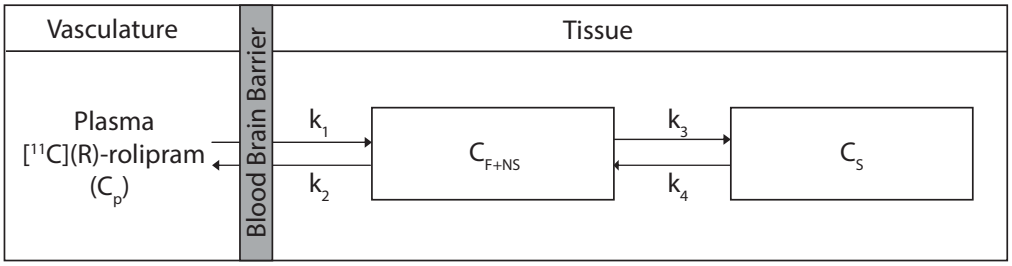


Figure 2.2. 2-tissue compartment model of [^{11}C](R)-rolipram: k_1 and k_2 correspond to the speed of inflow and outflow of radioligand through the blood brain barrier; k_3 and k_4 correspond to the speed of transfer of the radioligand between the compartment of free and non-specifically bound tracer and the compartment relative to the [^{11}C](R)-rolipram bound with specific receptors.

The equations of the system are:

$$\begin{cases} \dot{C}_{F+NS}(t) = k_1 C_p(t) + k_4 C_S(t) - (k_2 + k_3) \cdot C_{F+NS}(t) \\ \dot{C}_S(t) = k_3 C_{F+NS}(t) - k_4 C_S(t) \end{cases} \tag{2.5}$$

while the measurement equation is:

$$C_m(t) = (1 - V_b) \cdot [C_{F+NS}(t) + C_S(t)] + V_b \cdot C_b(t) \tag{2.6}$$

where V_b and C_b were already defined with L-[1-^{11}C]leucine.

As parameter of interest the total distribution volume V_t, which equals the ratio at equilibrium of total radioactivity in brain to the total plasma concentration, was chosen. V_t is obtained as:

$$V_t = \frac{k_1}{k_2}\left(1 + \frac{k_3}{k_4}\right) \tag{2.7}$$

2.2.2 Spectral Analysis

Spectral Analysis (SA) is a technique first introduced by Cunningham and Jones in 1993 [22] for the analysis of dynamic PET data. The strength of this method is that with relatively few assumptions it allows to identify the tissue kinetic parameters without knowing the compartmental structure of the system. One of these assumptions is that the tracer system is linear, time-invariant (LTI) and its eigenvalues and coefficients are all real-valued and non-positive [23]. These conditions are generally met in

dynamic PET studies. For this kind of systems, the impulse response can be written as:

$$h(t) = \sum_{i=1}^{M} \alpha_i \cdot e^{-\beta_i t} \tag{2.8}$$

where α_i and $-\beta_i$ are respectively the coefficients and the eigenvalues of the system and M is unknown.

Since the system is LTI, its response, i.e. the tracer concentration in the tissue, can be calculated through the convolution operator:

$$\begin{aligned} C_{tissue}(t) &= C_p(t) * h(t) \\ &= \sum_{i=1}^{M} \alpha_i \cdot \int_0^t C_p(z) \cdot e^{-\beta_i (z-t)} dz \end{aligned} \tag{2.9}$$

While the measurement equation is:

$$C_m(t) = (1 - V_b) \cdot C_{tissue}(t) + V_b \cdot C_b(t) \tag{2.10}$$

SA requires to define a grid of M fixed βs that covers an appropriate spectral range, where M has to be "high" (≥ 100). Once the βs are fixed, the problem of estimating the remaining parameters becomes linear and it can be easily solved using a Weighted linear Non-Negative Least Squares (WNNLS) estimator, where the weights to be used are the same already defined in subsection 2.2.1. The result of the estimation is called "spectrum": it is important to note that, even if there is a large number of coefficients to be estimated, at most N (i.e. the number of time samples) of them can be nonzero, so there are few positive peaks in the spectrum.

By convention the higher βs are called "high-frequency" components and they should be related to the vascular volume, while βs close to 0 are referred to as "low-frequency" components and they account for irreversible trapping of the tracer. The components in between are called "intermediate frequencies" and they reflect the tissue activity of the tracer (also referred to as equilibrating components) [4].

To clarify, even if the technique is called *spectral* analysis and the βs are referred to as *frequencies*, SA does not operate in the frequency domain but exclusively in the time domain.

When using SA, the two parameters of interest, K_I for L-[1-^{11}C]-leucine and V_t for [^{11}C](R)-rolipram, are given by the following formulas:

$$K_I = \alpha_0 \tag{2.11}$$

$$V_t = \sum_{i=1}^{M} \frac{\alpha_i}{\beta_i} \tag{2.12}$$

2.2.3 Spectral Analysis with Iterative Filtering

Noise in data greatly influences the accuracy with which very low frequency and high frequency components can be detected. For this reason a refinement of the SA technique that incorporates a numerical filtering step was proposed in [21] and called Spectral Analysis with Iterative Filtering (SAIF).

The basic idea of this method is that all the components with exponents greater than zero but less than β_L are assumed to have been shifted from β_0 (see equation 2.9) due to noise in the data and components with exponents greater than β_U are assumed to be connected to the blood volume term. β_L and β_U are called cut off frequencies and need to be fixed *a priori*.

A two steps numerical filter was thus defined: initially the equilibrating components are removed and new values of the trapping component and V_b are estimated; then this two are removed from the data and the equilibrating components are re-estimated. All the process is repeated until reaching the stabilization of the WRSS (Weighted Residuals Sum of Squaresp) yielding to better accuracy in the trapping component α_0 and V_b estimation.

Chapter 3

Average-Based Input Function

Average-Based Input Function (ABIF) method is founded on the hypothesis that different subjects who undergo the same PET study share an identical shape, but different amplitude, of the Arterial Input Function (AIF). Under this assumption it is ideally possible to substitute the individual AIF of the subject under analysis with the one of any other subject, after having properly scaled it. For this purpose only a single arterial sample could be extracted from the patient, greatly reducing his discomfort during the exam and considerably simplifying the exam procedure for clinicians. In practice, since data are affected by measurement errors, it is preferable to average the AIFs of many subjects and scale the result with one or two arterial samples [9].

One condition that this approach requires to be satisfied is that the PET exams have to be consistent between subjects: in particular, the infusion protocol has a significant influence on the AIF shape and it is therefore advisable the use of a computer driven pump in order to eliminate a source of uncertainty. The tracer dose is usually calculated on many factors and it also depends on the patient anthropometric parameters and so it may change from subject to subject. However, provided that it varies within certain limits, the tracer dose should affects the amplitude, but not the shape, of the AIF, leaving the ABIF hypothesis satisfied. Actually in this kind of studies it is often used the normalized AIF instead of the AIF itself. This is calculated dividing the latter by the dose injected so as to make it less variable between different subjects. However this passage is not strictly necessary in the ABIF method since the scaling process overrides the normalization (both scaling and normalizing consist in multiplying the AIF for a constant term).

Another aspect that must be taken into account is the population composition: while it is quite reasonable to assume that a population of healthy patients with comparable ages who undergo the same PET study may share a similar shape of the AIF (the extent to which this assertion is true is actually one of the key point of this chapter), this may not be the case when the population is composed by a mixture of healthy and unhealthy subjects or people with really different ages. Pathologies, in particular cardiovascular diseases (high pressure, ...), or even the state (awake, asleep or anesthetized) of the patient might potentially alter the AIF shape. These are just speculations but before applying the ABIF approach, it is important to asses if the same-shape hypothesis holds true in the population under analysis.

Finally, the tracer itself may constitute a limit to the applicability of this method: different tracers have different kinetics (here with kinetics it is intended the variation of the tracer concentration inside plasma and not the variation inside the tissue under analysis) and there are no guarantees that if the same-shape hypothesis holds for one tracer it will do the same for another. In particular literature lacks of information about the application of the ABIF method in dynamic PET studies with other tracers than 2-deoxy-2-(^{18}F)fluoro-D-glucose (FDG) [7] [8] and only a few works have been reported using non-FDG brain tracers [9][24][25].

ABIF approach is emerging between the minimally invasive techniques because of its relative ease of implementation: it does not require any particular mathematical skill (unlike SIME approach) and it does not present significant issues. Of course it requires a sufficiently rich dataset of subjects whose AIFs have been previously collected with the standard method (arterial sampling).

In the first part of the study it was supposed that the populations under analysis satisfy the same-shape hypothesis and the attention was focused on determining the optimal settings of the ABIF approach. In fact, even if it does not seem to present any particular difficulty, this approach involves a sequence of steps that leaves some kind of arbitrariness to the operator dealing with them. One of the key factors is represented by the number of subjects to be averaged in order to calculate a reliable ABIF. Of course this depends mainly on the noise present in the data and on the variability of the tracer between subjects (the more homogeneous this is, the fewer subjects will be needed). However, if the number of subjects that must be

used is too high, the ABIF methods will lose most of its attractiveness: in fact it would be necessary to directly measure the AIF of tens of subjects with the invasive method before being able to use the ABIF approach and there are not many centers who can perform such a large study (it should be remembered that this approach applies only to consistent exams and the data collected from one study may not be used for another).

The calibration step (formerly called "scaling") is another critical point. It consists on the calculation of a Correcting Factor (CF) that has to be multiplied to the averaged AIFs in order to take account of the amplitude variability between subjects. The CF is usually obtained by the ratio between the population AIF value at a certain time and the concentration of tracer in the subject at the same time acquired from an arterial or venous sample [7]. As suggested in [9], not all the times at which the sample is extracted are equally good. In order to improve the robustness of the method, in this study at least two samples were always used for the calibration step. This allowed some degrees of freedom on the choice of the sampling times and even on the methods used for calculating the CF. Both were examined in detail with the purpose of finding the optimal combination. At the end of this part the entire procedure for the definition of the ABIF was completely and optimally defined.

The successive part of the study is focused on the evaluating the reliability of the ABIF approach itself. Here the key point is whether or not this method can effectively be used to replace the measured AIF. This was tested by comparing the tissues kinetic parameter estimates using as input first the ABIF and then the measured AIF (i.e. Gold standard). Tissue quantification was performed by means of spectral analysis and spectral analysis with iterative filtering (see section 3.5), both applied voxel-wise.

3.1 Materials

The first part of the study was conducted on all the four data sets described in section 2.1 (L-[1-^{11}C]leucine, [^{11}C]DASB, [^{11}C]PiB, [^{11}C](R)-rolipram), while the second part, i.e. the evaluation of the impact of using ABIF on the tissue parameters estimations, was performed only on the L-[1-^{11}C]leucine and [^{11}C](R)-rolipram ones.

About the L-[1-^{11}C]leucine dataset, three subjects were excluded from the study because the rising part of their AIFs was considerably

shorter than normal. Probably different infusion protocol or dose injected were responsible of these effects. The L-[1-^{11}C]leucine dataset was therefore reduced to 47 subjects.

As regards the [^{11}C]PiB dataset, two subjects were eliminated due to their abnormal AIF shapes which indicates that probably something went wrong during theirp acquisition or the preprocessing step. The dataset considered was therefore composed by only 18 subjects.

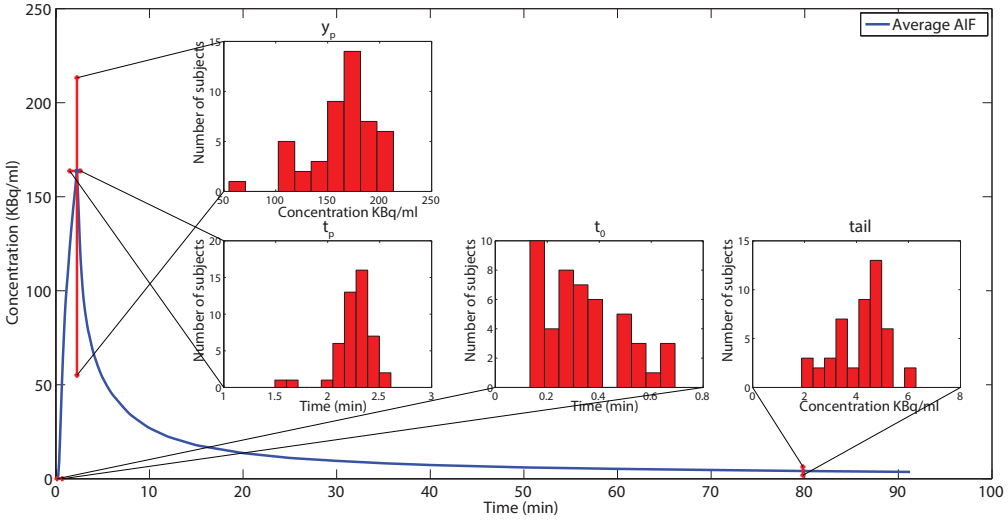
For what concerns the [^{11}C](R)-rolipram and the [^{11}C]DASB data sets, they were used in their entirety.

3.1.1 Dataset variability

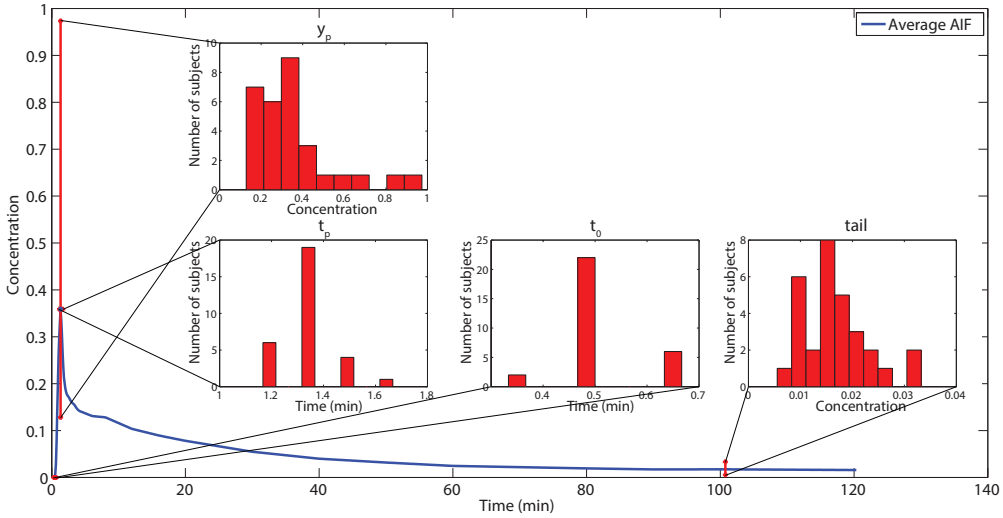
In order to understand the degree to which each dataset approximates the same-shape hypothesis, a few parameters were individuated in each AIF and the range within they vary was evaluated. These parameters are:

- t_0 (min): it represents the instant in which the AIF starts to rise, i.e. when the tracer arrives at the place of measurement;
- t_p (min): Time to Peak, it denotes the time in which the tracer concentration reaches its maximum;
- y_p (KBq/ml for all the subjects except [^{11}C]DASB where it is expressed as dose-normalized unit): it indicates the tracer concentration at the time t_p;
- *tail* (see y_p): it's the average of the tracer concentrations measured in the last three samples;

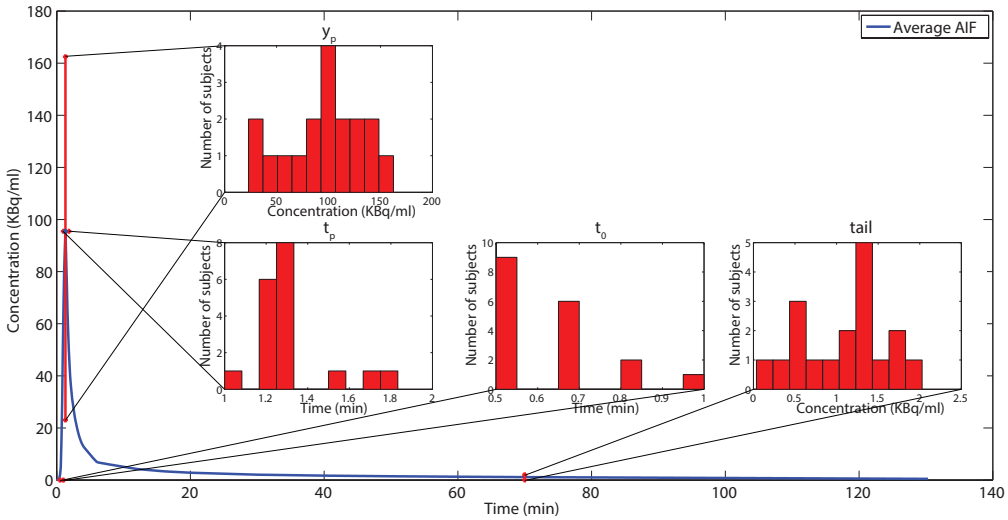
The parameters distributions within each dataset are shown in figure 3.0. y_p is by far the most variable parameter but this is not a problem since it is related with the amplitude, and not the shape, of the AIFs. The range within t_0 and t_p is quite narrow: the largest difference can be found in the L-[1-^{11}C]leucine dataset where t_p shows a minimum-maximum difference of around 1 min (figure 3.0ap). Unlike y_p, these parameters are related to the AIF shape and, more precisely, they indicate that the curves are translated on the temporal axis respect to each other.



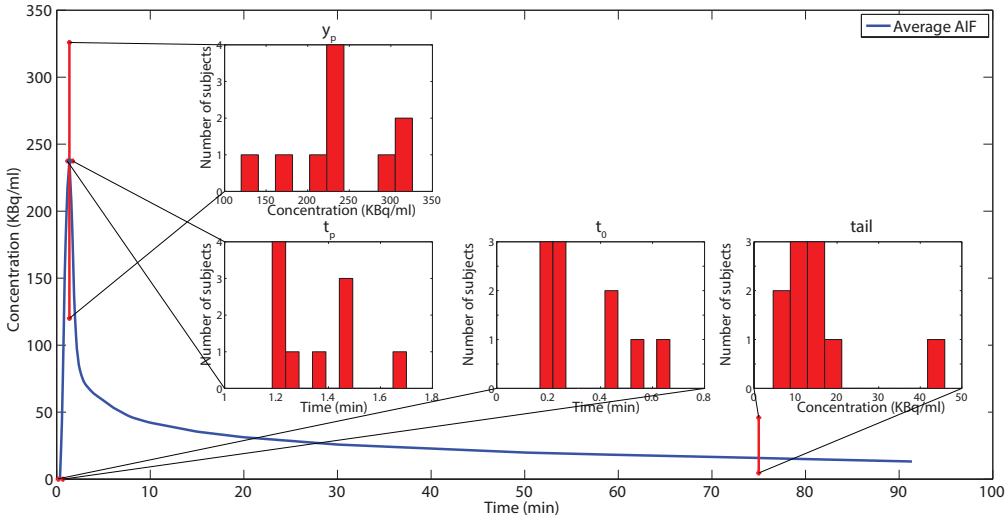
(a) L-[1-^{11}C]leucine



(b) [^{11}C]DASB



(c) [^{11}C]PiB



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

Figure 3.0. AIFs parameters distribution within each data sets

3.2 ABIF construction

In this section the mathematical passages needed to obtain the ABIF are discussed. For the sake of simplicity, the procedure is explained with reference to a single generic dataset. The results are valid for all the data sets considered in this study.

The first step carried out to build an ABIF was the averaging of the individual AIFs available. Since the sampling grid changed from subject to subject and given the temporal misalignment of the AIFs mentioned in the 3.1.1 subsection, this step required a few preliminary substeps:

1. the average time to peak ($\hat{t_p}$) was calculated and all the AIFs were translated so as to make the individual t_p coincide with $\hat{t_p}$;

2. a virtual grid T_v was created. The starting time was taken at $t_{init} = 0$, the distance between two consecutive points was calculated as $step = \hat{t_p}/20$ (so that $\hat{t_p} \in T_v$) and the final time was set at $t_{end} = min\left\{t : t \geq max\left\{t_{end_i} - t_{p_i} + \hat{t_p}\right\} \wedge t \equiv 0 \pmod{step}\right\}$, where i indicates the i-th AIF used. T_v was therefore an equally spaced grid which contained $\hat{t_p}$ and whose final point was greater than the final point of all the aligned AIFs used;

3. the averaged AIF was thus calculated as $\hat{C_p}(t) = \frac{1}{N}\sum_{i=1}^{N} lin[C_{p_i}(t - t_{p_i} + \hat{t_p})]$, where $lin[\cdot]$ represents the linear interpolation operator and

N is the number of AIFs used. When $(t - t_{p_i} + \hat{t_p})$ was greater than t_{end_i}, $lin[\cdot]$ behaved as a linear extrapolation operator, while when $(t - t_{p_i} + \hat{t_p})$ was lower than zero, it returned zero.

3.2.1 Calibration

Once the average AIF was calculated, the next step was to scale it to the proper amplitude. This phase takes the name of "calibration" and it requires one or more blood samples. If only one blood sample ($C_{p_i}(t_1)$) is available this passage is quite simple and the ABIF of the i$^{\text{th}}$ subject is given by:

$$ABIF_i(t) = CF \cdot \hat{C_p}(t) \quad \text{where} \quad CF = \frac{C_{p_i}(t_1)}{\hat{C_p}(t_1)} \tag{3.1}$$

However, in this study at least two samples were always used to scale $\hat{C_p}(t)$. Three different calibration techniques were tested in order to determine the best one:

Calibration 1: mean-based method

The first technique is an extension of the 3.1 equation when M blood samples are available. It simply consists on calculating the mean CF and for this reason it is called mean method.

$$ABIF_i(t) = CF \cdot \hat{C_p}(t) \quad \text{where} \quad CF = \frac{1}{M} \sum_{j=1}^{M} \frac{C_{p_i}(t_j)}{\hat{C_p}(t_j)} \tag{3.2}$$

Calibration 2: linear least squares method

The second technique is called least squares method and it consists in finding the CF that minimize the sum of the squared differences between the blood samples and the ABIF. This can be expressed as:

$$\begin{aligned} &ABIF_i(t) = CF \cdot \hat{C_p}(t) \quad \text{where} \\ &CF = \underset{CF}{\operatorname{argmin}} \left\{ \sum_{j=1}^{M} [CF \cdot \hat{C_p}(t_j) - C_{p_i}(t_j)]^2 \right\} \end{aligned} \tag{3.3}$$

CF can be expressed in closed form as:

$$CF = \frac{\sum_{j=1}^{M} \hat{C}_p(t_j) \cdot C_{p_i}(t_j)}{\sum_{j=1}^{M} \hat{C}_p(t_j)^2} \tag{3.4}$$

Calibration 3: non-linear least squares method

The third and final technique is an attempt to correct not only the amplitude but also the position of the t_p of the ABIF. This is achieved finding both a CF and a time shift t_s that minimize the sum of the squared differences between the blood samples and the ABIF. That is:

$$\begin{aligned} &ABIF_i(t) = CF \cdot \hat{C}_p(t - t_s) \quad \text{where} \\ &[CF, t_s] = \underset{CF,\, t_s}{\operatorname{argmin}} \left\{ \sum_{j=1}^{M} [CF \cdot \hat{C}_p(t_j - t_s) - C_{p_i}(t_j)]^2 \right\} \end{aligned} \tag{3.5}$$

Unfortunately there is no closed form for the equation 3.5 because the problem is non-linear in the parameters (and it is therefore called non-linear least squares method). In its implementation a constrain was set on t_s in order to make the t_p of the ABIF ranging between the minimum and maximum t_p of the AIFs used.

3.2.2 Sampling time

The samples used for the calibration step are crucial for the correct estimation of the ABIF [9]. For this reason three different configurations were tried.

Initially three blood samples, taken at the times $t_1 = 10\,min$, $t_2 = 50\,min$ and $t_3 = 80\,min$, were used for the calibration process. All these samples were chosen after the peak of the curve with the aim of covering the largest portion of the AIF after the t_p.

Then, the number of samples used was reduced to two and only the samples at the times $t_1 = 10\,min$ and $t_2 = 50\,min$ were kept. This was done in order to evaluate how the performance changed when less samples were employed.

Finally, an attempt to calibrate the averaged AIF using the tail, i.e. the last three samples, was made. The reason for this choice was that after a certain period of time the tracer concentration in the venous vessels tends to match the one in the arterial vessels. If the use of these proves to be effective, the arterial samples may be easily replaced by the venous ones, greatly simplifying the exam procedure.

3.2.3 Number of subjects

The last factor analyzed which could have an impact on the performance was the number of AIFs averaged together to obtain the ABIF. Intuitively, the more AIFs are averaged, the better the performance. The aim is to find a compromise between the number of subjects required and the ability to correctly reconstruct the curve. Starting from a minimum of five subjects, the ABIF performance were evaluated increasing this number up to using the entire dataset minus the subject under analysis.

3.3 Assessment of ABIF performance

3.3.1 Performance indices

In order to asses how each of these three factors apffects the performance of the ABIF method, it was necessary to evaluate all the possible combinations. Before doing so, it was essential to define how the performance was quantified. This was done using the following performance indices:

- MSE: Mean Squared Error, calculated between the estimated ABIF and the measured AIF of the subject under analysis. Given $C_{p_i}(t)$ defined for $t \in T_i$, where $T_i = \{t_{1_i}, t_{2_i}, \ldots, t_{j_i}, \ldots, t_{M_i}\}$ and $ABIF_i(t)$ defined for $t \in T_v$, the MSE between $ABIF_i(t)$ and $C_{p_i}(t)$ is calculated as:

$$MSE_i = \frac{1}{M_i} \sum_{t \in T_i} \{lin[ABIF_i(t)] - C_{p_i}(t)\}^2 \tag{3.6}$$

  where $lin[\cdot]$ represents the linear interpolation operator defined in section 3.2.

- ΔAUC: Delta Area Under the Curves, intended as the relative percentage difference between the AUC_{ABIF_i} and AUC_i. That is:

$$\Delta AUC_i = 100 \cdot \frac{AUC_{ABIF_i} - AUC_i}{AUC_i} \tag{3.7}$$

Unlike the MSE index, the ΔAUC is insensitive to errors of opposite sign: if the ABIF underestimates the AUC of the first part of the curve and overestimates the second part, the resulting ΔAUC will not reflect the error. In order to address, at least in part, this issue, the index was split in two: ΔAUC_{peak} and ΔAUC_{tail}, the first considering the area from $t = 0$ to $t = 10\,min$, while the second considering the remaining part. In order to avoid encumbering the reading, hereinafter only the ΔAUC will be mentioned, but every time this is done, all the three indices will be meant. For this study, all the AUCs were computed using the trapezoidal method.

3.3.2 Leave-one-out approach

A leave-one-out scheme was used to assess how the three factors - calibration method, sampling times and number of subjects - affects the two performance indices - MSE and ΔAUC. Fpor each subject all the factors were made to vary one by one, generating the corresponding ABIF by using the rest of the dataset and calculating the two performance indices at each variation. More in detail, the adopted scheme consisted of the following steps:

1. the i-th subject was extracted from the dataset and it was considered as the subject under analysis. This means that this subject's AIF was supposed to be unknown and thus it had to be estimated through the ABIF method;

2. 1000 different subsets composed by j randomly chosen AIFs were generated from the rest of the dataset. Each subject could be part of more than one subsets but two different subsets must differ for at least one subject (see figure 3.1);

3. the AIFs of each of these subsets were averaged together in the way discussed at the beginning of the section 3.2. This passage led to obtain 1000 different averaged AIF.

4. each of this averaged AIFs was then calibrated with all the possible combination of the three calibration methods with the three sampling times. This means that for each averaged AIF, nine different ABIFs were generated (see table 3.1);

5. MSE and ΔAUC were computed between each of the 9000 ABIFs and the measured AIF of the subject under analysis.

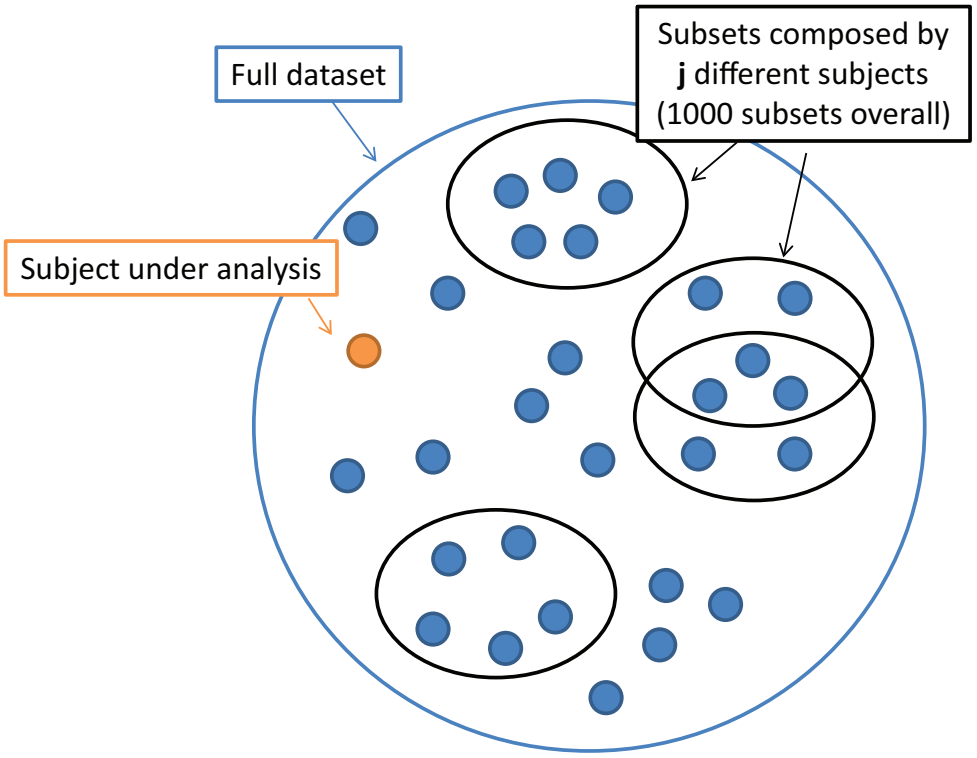


Figure 3.1. Exemplification of the subsets generated from the dataset for each subject under analysis.

6. for each of the nine combinations of the calibration settings, 1000 values of both the performance indices were thus available. These represented the ABIF capacity of approximating the measured AIF of the ith subject given a particular fixed setting: a fixed calibration method applied with samples taken at fixed times and using a curve obtained by averaging j subjects. What changed in the computation of these 1000 indices were just the subjects composing the subset used for creating the ABIF. The discussion on how the choice of the subjects affected the method performance are deferred to the second part of this study (section 3.5). For this part the 1000 values were considered as repeated measurements of the same quantity and thus only their averages were kept: $\hat{MSE}$ and $\Delta\hat{AUC}$

7. the whole process was repeated for all the subjects composing the dataset (N subjects). This led to N estimates of $\hat{MSE}$ and $\Delta\hat{AUC}$, one for each subject. The mean and the standard deviation of these quantities were computed so as to have only one value (and its stan-

	Sampling times (min)		
Calibration methods	Mean t=10,50,80	Mean t=10,50	Mean t=tail
	Least squares t=10,50,80	Least squares t=10,50	Least squares t=tail
	Non-linear t=10,50,80	Non-linear t=10,50	Non-linear t=tail

Table 3.1. Combining the calibration methods proposed in subsection 3.2.1 with the sampling times discussed in subsections 3.2.2 gives nine alternative ways to perform calibration.

dard deviation) for each of the two indices for a given calibration setting (intended both as the method and as the sampling times) using a generic population of j subject to calculate the ABIF. In other words, the two values $Mean\,\hat{MSE}$ and $Mean\,\Delta\hat{AUC}$ and their respective standard deviations indicate how good the ABIF approximates the measured AIF on average and how this performance may change from subject to subject.

3.4 Results

The four figures 3.2 - 3.5 allow to have a straightforward vision of how all the variables involved in the ABIF method impact on the performance indices and they lend themselves to an easy comparison of all the possible settings considered in this study.

3.4.1 Sampling times for calibration

Comparing the three columns, which represents the sampling times used for calibration, it emerges that calibrating using the last three samples always performs worse than the other possibilities. In fact, observing the $Mean\,\hat{MSE}$ and the $Mean\,\Delta\hat{AUC}_{peak}$ indices in the right columns, it can be seen that their values are always the highest. For some particular settings the $Mean\,\Delta\hat{AUC}_{tail}$ values are actually smaller than when other sampling times are employed. The reason is soon explained: the sampling

schedule employed for the measured AIF is such as to have less and less frequent samples as the exam reaches its end. Since the $Mean\,\Delta\hat{AUC}_{tail}$ considers only the final portion of the AIF, its value is calculated on relatively few samples, three of which are actually known. Since these three samples are the last three, they are also the most spaced and thus the corresponding area they cover is high when considering only the final portion of the AIF. This implies that when used in the calibration method, independently on which of the three techniques is implemented, they provides a lot of information to correctly estimate the final portion of the AIF, strongly staggering the $Mean\,\Delta\hat{AUC}_{tail}$ value. Its main contribution is in fact given by known information and the index no more indicates the ability of the ABIF method to estimate the unknown part of the AIF. Thus, since this index is not reliable and given the results of the other two indices, it can be concluded that late samples, even if attractive (they easily could have been substituted with venous samples), may not be the best choice for the calibration step.

Comparing the first two columns of the figures emerges that all the indices are practically identical with a slightly lower indices in the central column ($t_{1,2} = 10,50\,min$). It was just seen that the calibration using the last three samples always performed poorly, so probably even in this case the use of the sample at the time $t = 80\,min$ should be avoided.

3.4.2 Calibration method

As concern the calibration methods, the non-linear least squares one can be immediately excluded from the comparison: looking at the L-[1-^{11}C]leucine (3.2) and the [^{11}C]PiB (3.4) figures, it can be noted that the $Mean\,\hat{MSE}$ index (first row) with the non-linear method (green) is by far the highest. Moreover, it increases as more subjects are used.

This behaviour can be explained recalling how the method works: in addition to a constant CF, the algorithm tries to recover the peak position estimating the delay t_s between the ABIF and the measured AIF (see the specific paragraph in subsection 3.2.1). Since the method is indeed non-linear, a constraint was set on t_s in order to make the t_p of the ABIF ranging between the minimum and maximum t_p of the AIFs used. As the number of AIFs used increases, the constraint becomes less and less rigid allowing more freedom to the t_s parameter and since this method clearly fails to recover the correct peak position, the results tend to be more and

more unreliable.

The [^{11}C](R)-rolipram dataset is too poor to make this kind of reasoning, but still it can be see that the $Mean\,\hat{MSE}$ index get higher instead of decreasing.

Why this does not happen with [^{11}C]DASB is soon explained: looking at the figure 3.0 about the data sets variability it can be noted how most of the [^{11}C]DASB subjects (19/30) share an identical t_p. This implies that the constrain on t_s remains rigid even if the number of subjects increases.

The other two indices, $Mean\,\Delta\hat{AUC}_{peak}$ and $Mean\,\Delta\hat{AUC}_{tail}$, are not affected by the time shift as the area under the curve remains unchanged but since the aim of this part is to determine the optimal way of recovering the AIF, it is of secondary importance that the area of the estimated AIF is quite correct if the curve profile is completely wrong. Thus, given its not reliable performance, this method was definitely rejected.

Comparing the other two calibration techniques, the following pattern emerges: the linear least squares method (red) is almost always associated with better $Mean\,\hat{MSE}$ and $Mean\,\Delta\hat{AUC}_{tail}$ values while the mean-based method (blue) generates better $Mean\,\Delta\hat{AUC}_{tail}$ values.

The two techniques are based on two different concepts: the mean-based method minimizes the relative difference between the ABIF and the measured samples; the linear least squares method minimizes the absolute difference instead. In fact, looking at the equation 3.2 relative to the mean-based method, it can be noted that the ratios, and not the differences, between the two curves are considered. The linear last squares method, instead, considers the square of the differences, an absolute value.

Now, since all the sampling times considered are after the t_p, they lay in the decreasing part of the curve. With the mean method the higher samples "count" the same as the lower because they are divided by the curve $\hat{C}_p(t)$ which theoretically has the same shape but a different amplitude of the measured AIF. However, if the curves are subtracted as in the least squares method, the higher samples will "count" the most because at those times the curves are more distant and the algorithm tries to minimize the (square of the) distance.

Considering the performance indices, the $Mean\,\hat{MSE}$ indicates the absolute difference between the estimated and the measured curves, thus the least squares method performs better; the $Mean\,\Delta\hat{AUC}_{peak}$ measures the relative difference of the first part of the curve, precisely the part that is favorite by the last squares algorithm, so again the least squares method performs better; the $Mean\,\Delta\hat{AUC}_{tail}$, instead, is related to the remaining part of the area under the curves which is reconstructed worse by this method and so its values are greater.

Since $Mean\,\Delta\hat{AUC}_{peak}$ is usually grater than $Mean\,\Delta\hat{AUC}_{tail}$ (i.e. the first part of the curve is reconstructed worse than the second with both method) and taken into account the $Mean\,\hat{MSE}$ values, the linear least squares method was chosen as the optimal method.

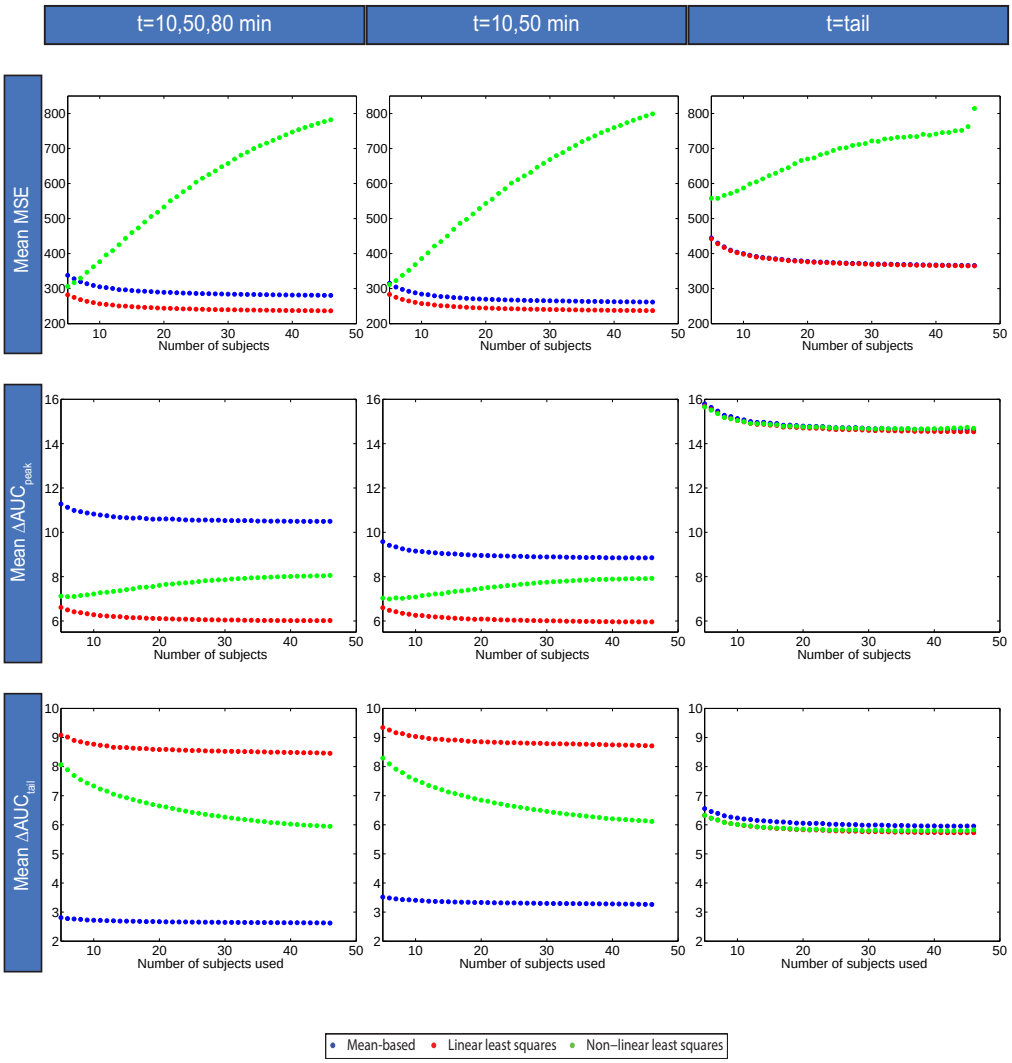


Figure 3.2. L-[1-^{11}C]leucine: this figure shows how the three performance indices, $Mean\,\hat{MSE}$ (first row), $Mean\,\Delta\hat{AUC}_{peak}$ (second row) and $Mean\,\Delta\hat{AUC}_{tail}$ (third row), vary when the calibration is performed using the samples at the times $t_{1,2,3} = 10, 50, 80\,min$ (first column), at the times $t_{1,2} = 10, 50\,min$ (second column) and at the last three times (third column). For each panel, blue line represents the mean-based calibration method, red line represents linear last squares method and green line the non linear least-squares method. Each line represents the behaviour of the index as function of the number of subjects used for defining the ABIF

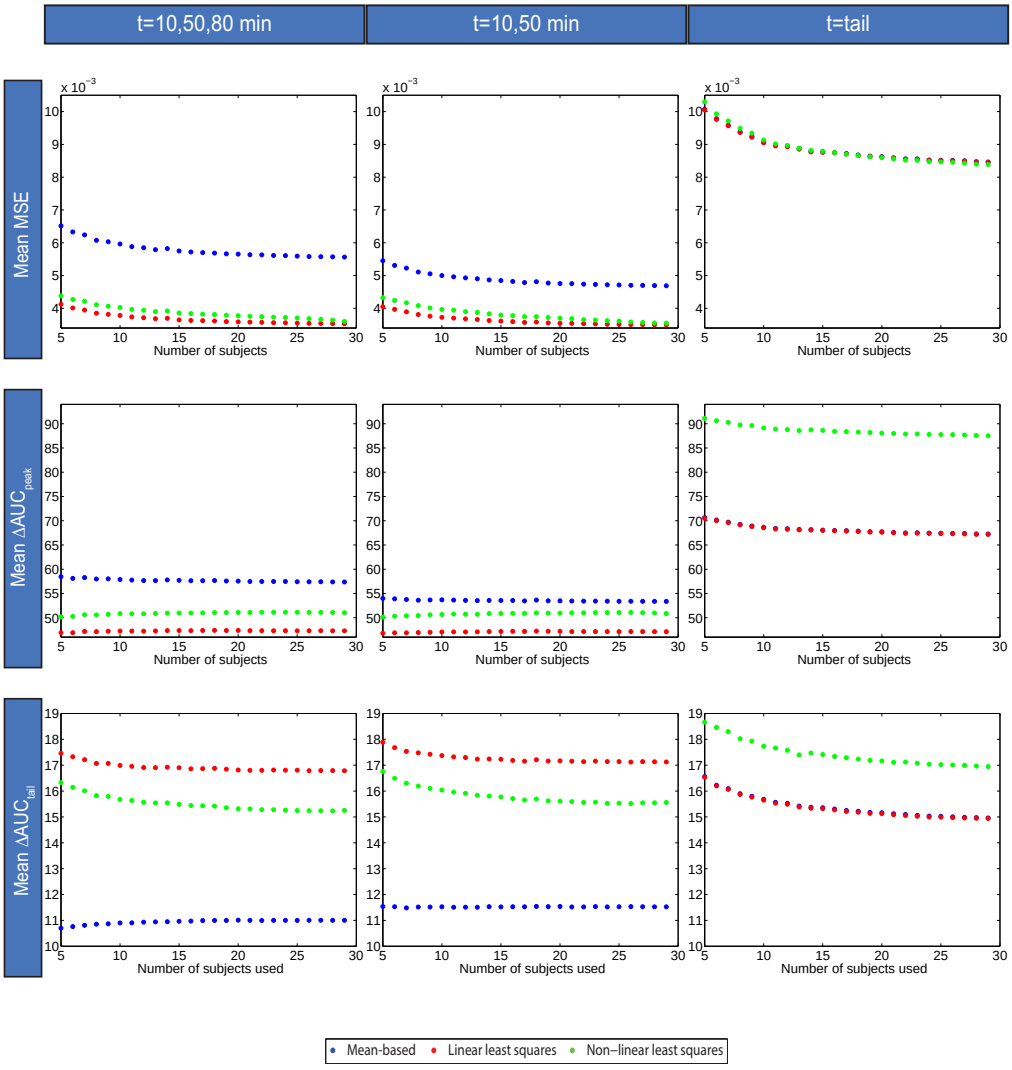


Figure 3.3. [^{11}C]DASB: see description in figure 3.2.

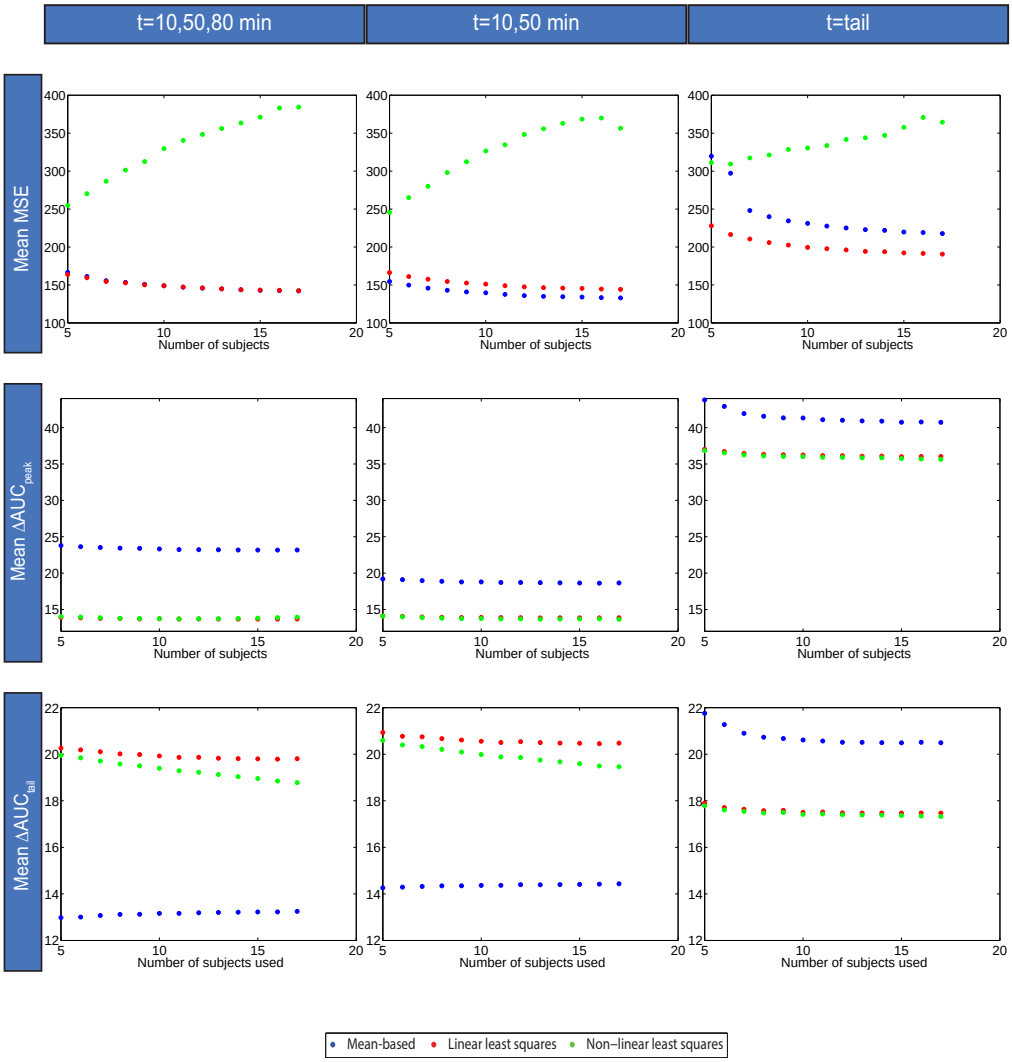


Figure 3.4. [^{11}C]PiB: see description in figure 3.2.

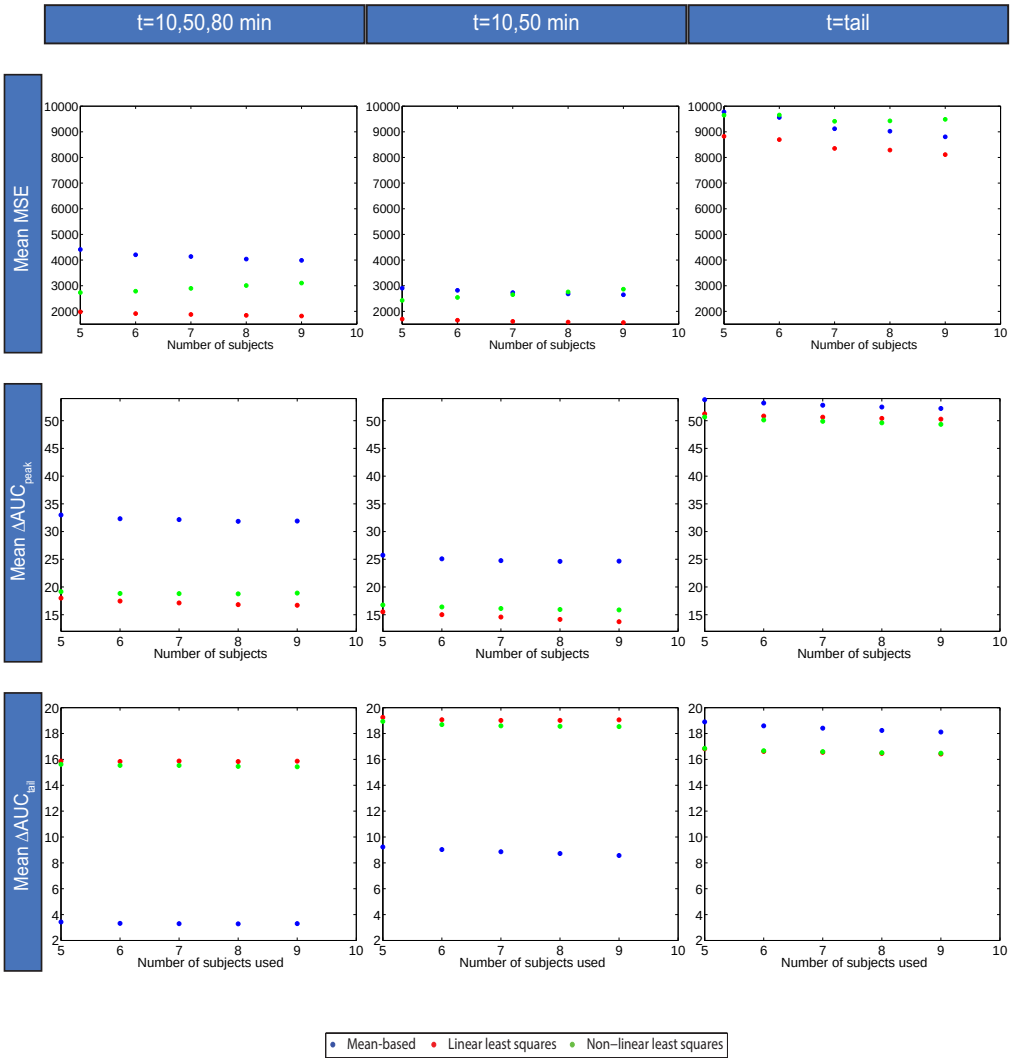


Figure 3.5. [^{11}C](R)-rolipram: see description in figure 3.2.

3.4.3 Number of subjects

Calibration performed with the linear least squares method, applied using the samples at the times $t_{1,2} = 10, 50\,min$ was identified as optimal. In this subsection all the results were obtained using that configuration and the attention was focused on the impact that the number of subjects averaged together have on the AIF reconstruction. Again, the aim is to identify the optimal number.

L-[1-^{11}C]leucine

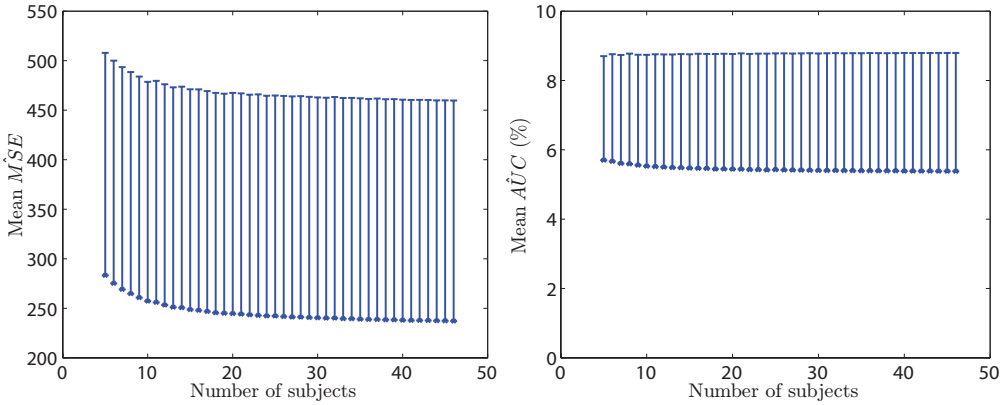


Figure 3.6. L-[1-^{11}C]leucine: $Mean\,\hat{MSE}$ with its standard deviation (left) and $Mean\,\Delta\hat{AUC}$ with its standard deviation (right) plotted versus the number of subjects. The linear least squares method applied using the samples taken at the times $t_{1,2} = 10, 50\,min$ is used for the calibration step.

From the figure 3.6 it can be seen that both indices get lower when the number of subjects averaged increases. This is quite evident for the $Mean\,\hat{MSE}$, that falls by around 16.3% when all the dataset (minus the subject under analysis), instead of only 5 subjects, is used. The improvement of the $Mean\,\Delta\hat{AUC}$ index is less appreciable: the index value varies between 5.70% (5 subjects used) and 5.38% (all the dataset used). The standard deviations of both indices seem not to benefit from the raised number of subjects used: for the $Mean\,\hat{MSE}$ the improvement is less than 1% (0.89%), while the standard deviation of the $Mean\,\Delta\hat{AUC}$ index actually gets worse, passing from ±3.0% to ±3.4%.

Assuming the $Mean\,\hat{MSE}$ index reached a plateau at the end of the graph, a threshold of 5% was chosen to determine the optimal number of subjects to be used. In other words, the optimal number was identified as the lower number whose corresponding $Mean\,\hat{MSE}$ value is at most 5% grater than the minimum $Mean\,\hat{MSE}$. This number is 15 and the corresponding value is 4.86% grater than the minimum.

[^{11}C]DASB

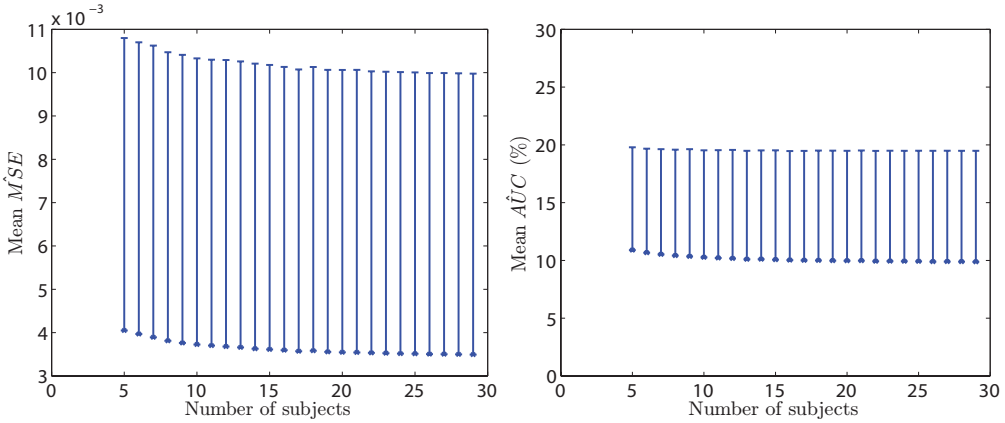


Figure 3.7. [^{11}C]DASB: see description in figure 3.6.

Similarly to L-[1-^{11}C]leucine, also with [^{11}C]DASB both indices decrease as the number of subjects used increases (figure 3.7). Again the $Mean\,\hat{MSE}$ index experiences a more marked reduction (13.7%) than the $Mean\,\Delta\hat{AUC}$ does (1.01%) while the standard deviation remains almost constant: it slightly decreases (3.9%) in the $Mean\,\hat{MSE}$ graph and it widens from $\pm$8.9% to $\pm$9.6% in the $Mean\,\Delta\hat{AUC}$ graph.

As regards [^{11}C]DASB, the optimal number of subjects to be used identified with the 5% threshold is 13. The corresponding $Mean\,\hat{MSE}$ value is 4.7% grater than the minimum.

[^{11}C]PiB

Even with [^{11}C]PiB the pattern remains the same (figure 3.8): the difference between the $Mean\,\hat{MSE}$ value when using only 5 subjects and

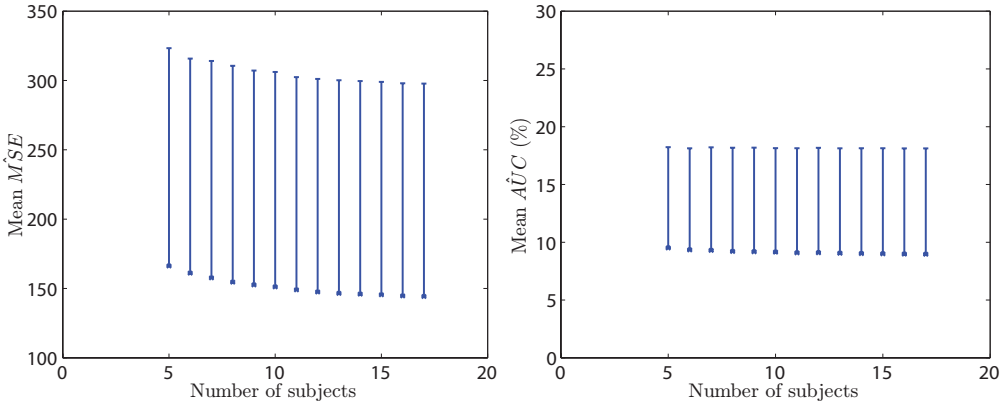


Figure 3.8. [^{11}C]PiB: see description in figure 3.6.

the value when using all the dataset is about 13.2%, while the same difference, but with regard to the $Mean\,\Delta\hat{AUC}$ index, is just about 0.55%. Again, the $Mean\,\hat{MSE}$ standard deviation decreases (-2.2%) while the $Mean\,\Delta\hat{AUC}$ standard deviation rises from $\pm 8.7\%$ to $\pm 9.2\%$.

For this tracer, the resulting optimal number of subjects is 10 (whose corresponding $Mean\,\hat{MSE}$ is 4.83% greater than the one obtained using the entire dataset). However, in this case, since the number of subjects available were not very high (18), the assumption that the $Mean\,\hat{MSE}$ index reached a plateau at the end of the graph is not properly verified. Hence, the optimal number may be higher (and a richer dataset should be used to determine its correct value).

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

As regards the [^{11}C](R)-rolipram dataset, only 10 subjects were available but the considerations made for the other tracers are still valid. Both the indices decrease as the number of subjects used increases. These improvements are of the order of 8% for the $Mean\,\hat{MSE}$ index and 0.4% for the $Mean\,\Delta\hat{AUC}$ index. The standard deviations of these indices narrow in the case of the former index (2.3%) and expand in the second case (from $\pm 5.9\%$ to $\pm 6.5\%$).

Similarly to what happened with the [^{11}C]PiB one, the [^{11}C](R)-rolipram dataset available was too poor to allow the identification of the optimal number of subjects to be used in the ABIF method.

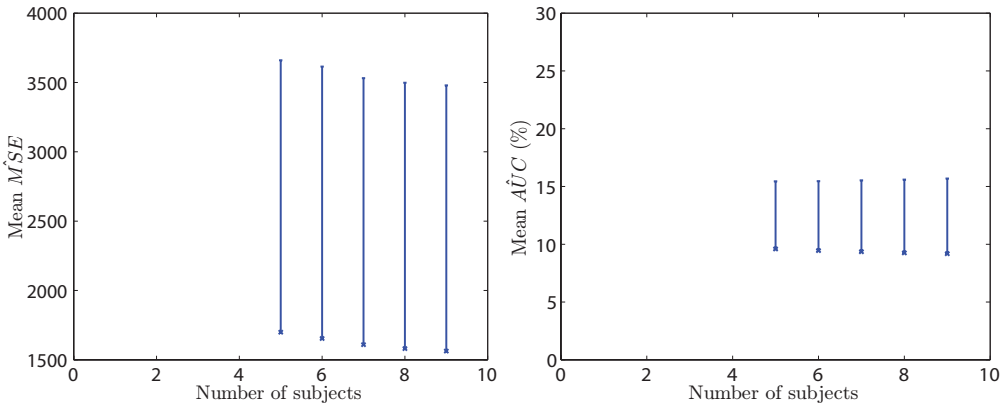


Figure 3.9. [^{11}C](R)-rolipram: see description in figure 3.6.

3.5 Impact on quantification

As result of the first part of this study, an optimal procedure for the application of the ABIF method was individuated and fully defined. The best calibration technique identified was the least squares method while for the optimal sampling times, $t_1 = 10\,min$ and $t_2 = 50\,min$ were chosen among the possibilities tested. Finally, the number of AIFs to be used depended on the specific tracer.

The aim of this second part was to asses if this method could be applied with success in replacement of the full arterial sampling and with what degree of confidence the results on the tissue parameters estimations could be accepted. Another key point was to understand what relevance the subjects used have, i.e. if two ABIFs generated with the same settings but with different subjects give similar results or not. In order to do so, the tissue parameters estimation was performed on the L-[1-^{11}C]leucine and the [^{11}C](R)-rolipram data sets. For the L-[1-^{11}C]leucine dataset, three subjects whose PET images were available were chosen and the trapping rate constant K_I was estimated using SAIF, applied voxelwise. This analysis was performed several times using as input functions ABIFs generated from different subsets of 15 (i.e. the optimal number identified in the first part) subjects each. Finally, SAIF was applied on the three subjects using their measured AIF. The K_Is so obtained were considered as the gold standard against which to compare the ones obtained with the ABIFs. As regard the [^{11}C](R)-rolipram dataset, the parameter of interest is the distribution volume V_t. Since this tracer had reversible kinetics, SAIF could not

be used and thus standard SA, still applied voxelwise, was used instead. Even for this dataset, the analysis were performed on three subjects using several different ABIFs and the measured AIF.

3.5.1 Implementation

As concern the implementation of the SAIF and SA methods, the software SAKE [26] was used. As βs grids, and for the SAIF also the cutoff frequencies, the default values were kept: about the grid the number of components were set to 100, starting from $\beta_{Min} = 0.01\,min^{-1}$ to $\beta_{Max} = 1\,min^{-1}$; with SAIF, the passband filter had the two cutoff frequencies at $\beta_L = 0.1\,min^{-1}$ and $\beta_U = 0.5\,min^{-1}$.

As regards the error law that describes the measurement error of the L-[1-^{11}C]leucine PET data, this was assumed to be [27]:

$$SD(t_i) \propto \sqrt{\frac{z_i \cdot e^{k \cdot t_i}}{\Delta t_i}} \tag{3.8}$$

where z_i is the measured datum at the time t_i; k is the decay constant of the ^{11}C isotope and it is equal to $k = 0.034\,min^{-1}$ and Δt_i is the time interval of the PET scan during the acquisition of the i^{th} datum.

The error law used with [^{11}C](R)-rolipram was assumed to be [18]:

$$SD(t_i) \propto \sqrt{\frac{z_i}{\Delta t_i}} \tag{3.9}$$

All the input functions used (measured and estimated with ABIF) were assumed to be error free.

3.5.2 Results

In this section a comparison between measured AIF and two ABIFs generated using different subsets of subjects was made. This provided an idea of the performance variability of the method depending on the population of subjects used for ABIF definition.

L-[1-^{11}C]leucine

From the L-[1-^{11}C]leucine dataset the subject 6 was selected since it exhibited the widest difference between the best and the worst case. In figure 3.10 the two ABIF used are shown: ABIF 1 gave the best results while ABIF 2 the worst ones.

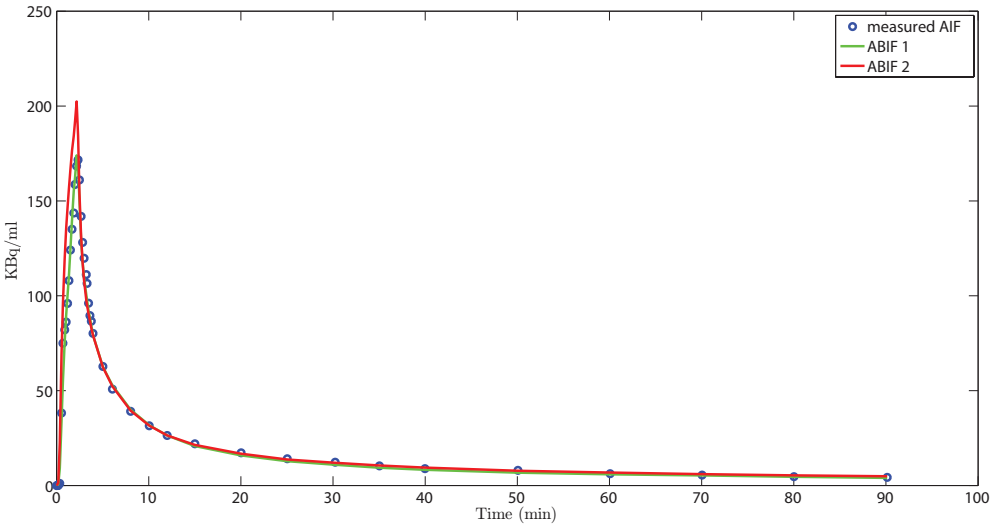


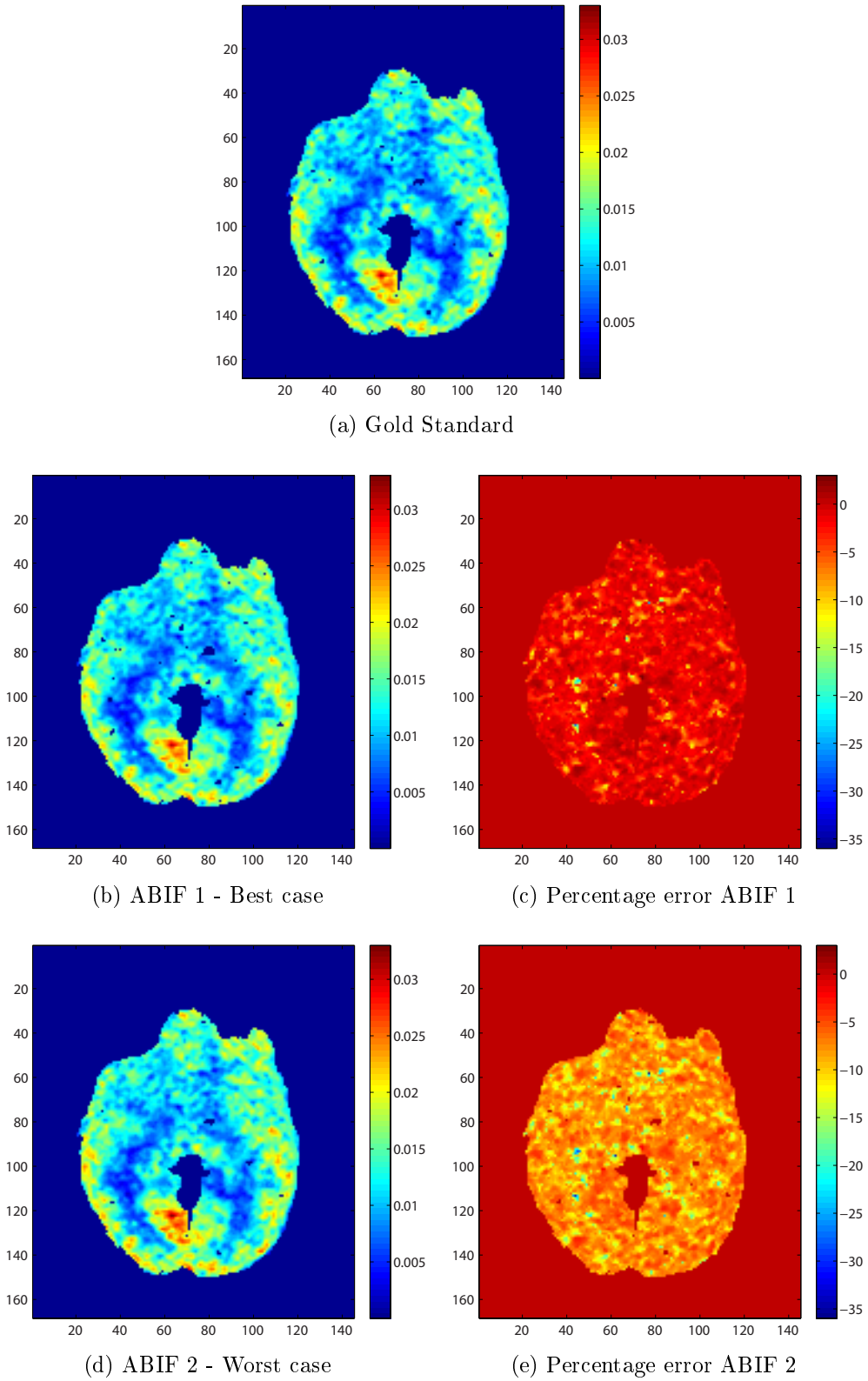
Figure 3.10. Two ABIFs generated with the optimal settings using different subsets of subjects (15 each). The blue circles represent the measured AIF.

The K_I maps of the 100^{th} slice generated with these two ABIFs and with the measured AIF are shown in figure 3.11. Visually the three maps look almost identical, so the percentage error maps were computed as well (right column). The percentage error e was defined as:

$$e(x,y) = 100 \cdot \frac{K_I(x,y) - \tilde{K_I}(x,y)}{K_I(x,y)} \tag{3.10}$$

where $K_I(x,y)$ represents the value of K_I at the coordinates (x,y) calculated with the measured AIF (Gold Standard) and $\tilde{K_I}(x,y)$ represents the same value but calculated with the ABIF.

The histograms of e computed with both ABIF for the slice 100 are shown in figure 3.12. The mean and the standard deviation of the error committed using ABIF 1 (best case) are $-2.01 \pm 2.23\%$, while they are $-8.02 \pm 2.48\%$ when ABIF 2 (worst case) is used.



(a) Gold Standard

(b) ABIF 1 - Best case

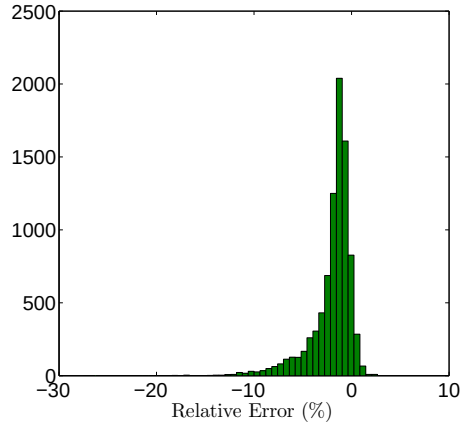
(c) Percentage error ABIF 1

(d) ABIF 2 - Worst case

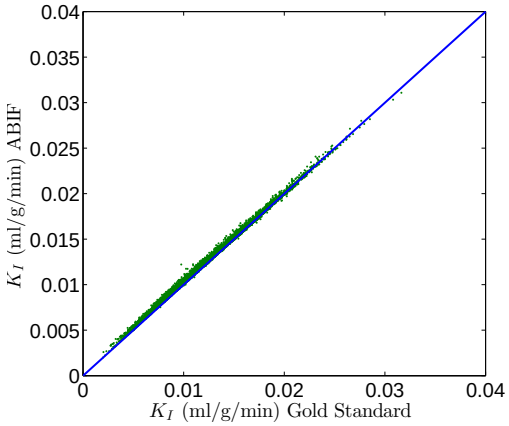
(e) Percentage error ABIF 2

Figure 3.11. K_I maps of one representative slice generated using as input function the measured AIF (a), the ABIF 1 (b) and the ABIF 2 (d). Maps of the percentage errors are shown as well (c)(e).

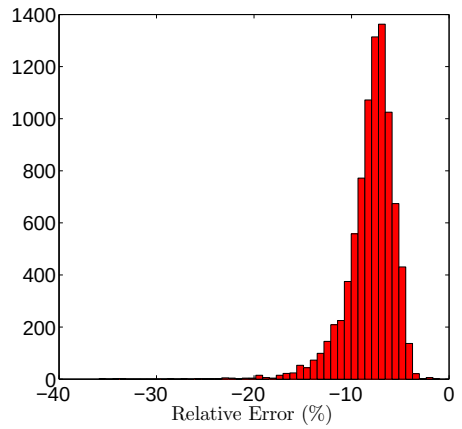
Always in figure 3.12, the scatter plots are shown as well. The regression line for the best case has a slope of $m = 0.988$ and a y-intercept of $q = 0.003$ (in case of perfect match the previous values would have been $m = 1$ and $q = 0$). The Pearson's coefficient R^2 is instead equal to 0.998 (1 in case of perfect matching). As concern the worst case, $m = 1.0423$, $q = 0.0004$ and $R^2 = 0.997$.



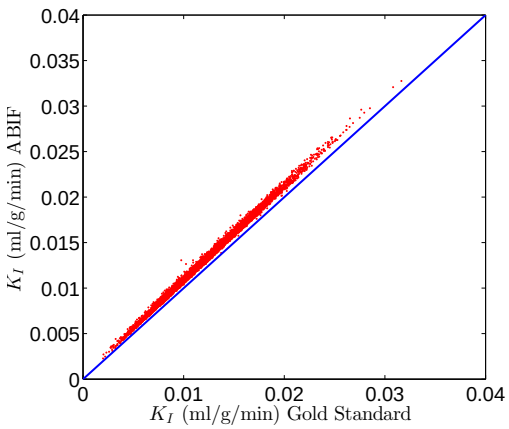
(a) Error histogram ABIF 1 - Best case



(b) Scatter plot ABIF 1 - Best case



(c) Error histogram ABIF 2 - Worst case



(d) Scatter plot ABIF 2 - Worst case

Figure 3.12. Histograms of the errors (a, c) and scatter plots (b, d) of the K_I parameter of the same slice shown in figure 3.11 when using the two ABIFs compared to the gold standard

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

As regards the [^{11}C](R)-rolipram dataset, the subject who gave the most different results when two different ABIF were used was the number 18. In figure 3.13 the measured AIF, the best ABIF (green) and the worst (red) are shown. Since the optimal number of subjects to be used was not defined for this tracer, it was decided to fix this number as 5, corresponding to half of the dataset available.

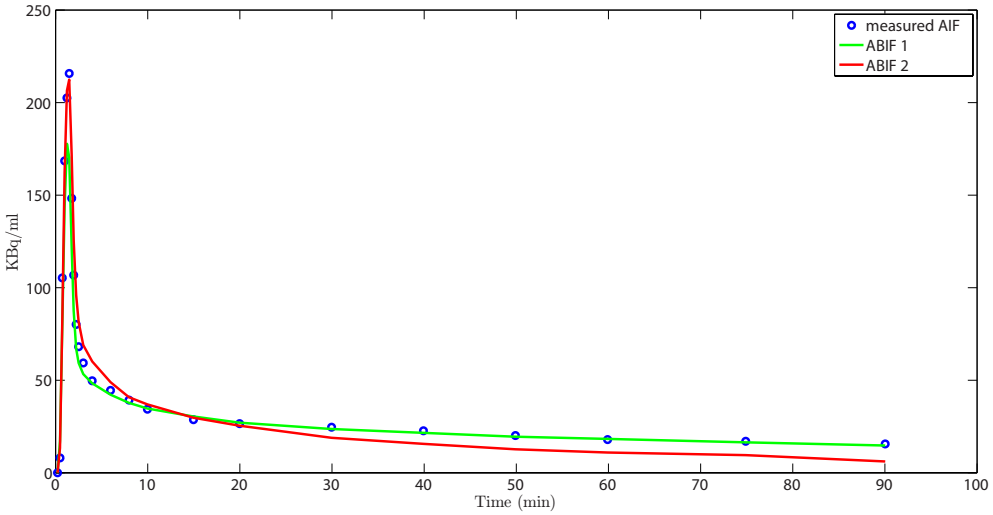
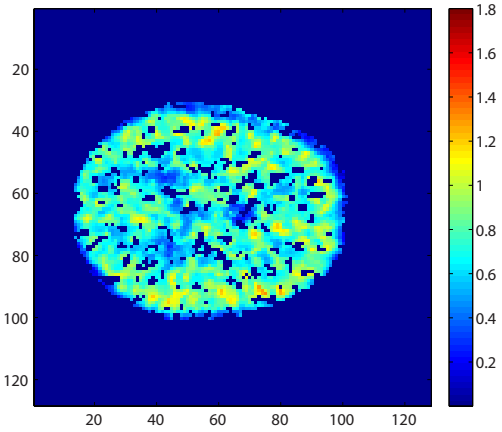


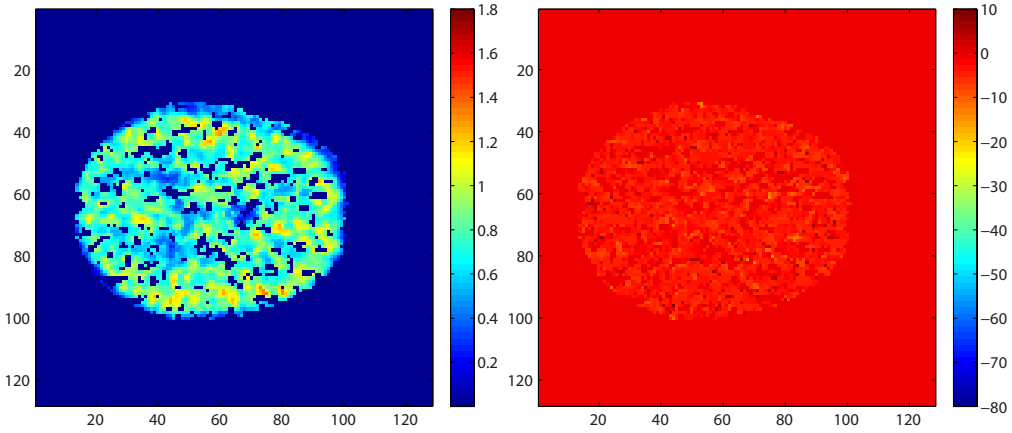
Figure 3.13. Two ABIFs generated with the optimal settings using different subsets of subjects (5 each). The blue circles represent the measured AIF.

In figure 3.14, the three V_t maps, one for each input functions, of the 19^{th} slice are shown as well as the corresponding error maps (see equation 3.10). The mean error and its standard deviation are $-3.89 \pm 2.40\%$ in the best case and $-48.09 \pm 11.63\%$ in the worst.

As regards the scatter plots shown in figure 3.15, for the best case (green) $m = 1.034$, $q = 0.004$ and $R^2 = 0.99$, while for the worst (red) $m = 1.44$, $q = 0.023$ and $R^2 = 0.92$.

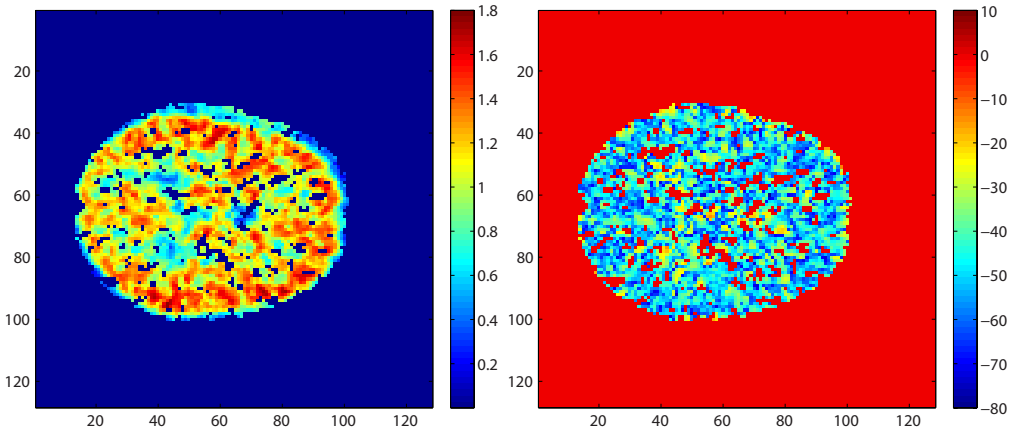


(a) Gold Standard



(b) ABIF 1 - Best case

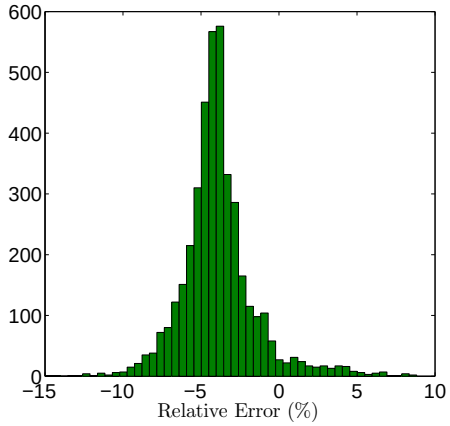
(c) Percentage error ABIF 1



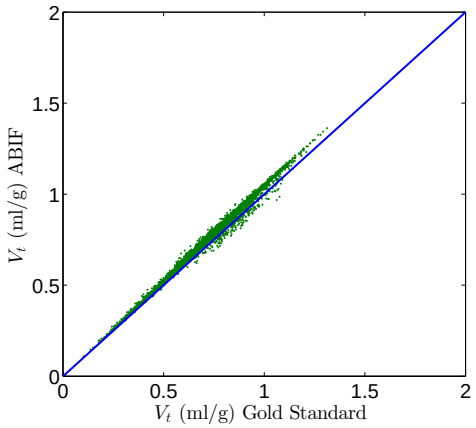
(d) ABIF 2 - Worst case

(e) Percentage error ABIF 1

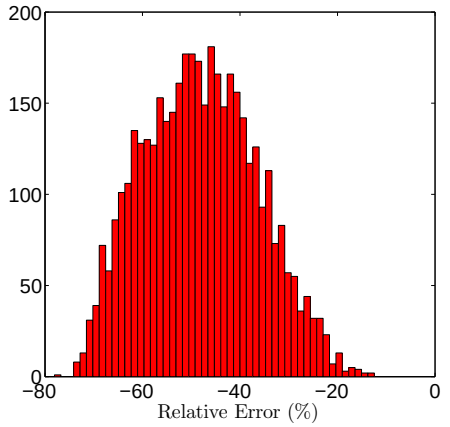
Figure 3.14. V_t maps of one representative slice generated using as input function the measured AIF (a), the ABIF 1 (b) and the ABIF 2 (d). Maps of the percentage errors are shown as well(c)(e).



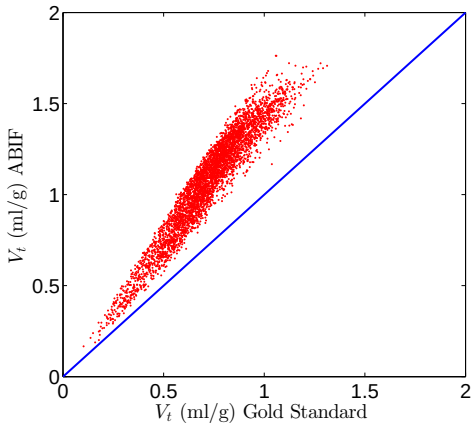
(a) Error histogram ABIF 1 - Best case



(b) Scatter plot ABIF 1 - Best case



(c) Error histogram ABIF 2 - Worst case



(d) Scatter plot ABIF 2 - Worst case

Figure 3.15. Histograms of the errors (a, c) and scatter plots (b, d) of the V_t parameter of the same slice shown in figure 3.14 when using the two ABIFs.

3.5.3 Discussion

According to the tracer considered, this second part yielded to considerably different results, almost of opposite sign: with L-[1-^{11}C]leucine the maximum mean percentage error of the K_I parameter encountered was just around -8% and even in the worst case the correlation coefficient R^2 kept over 0.99; with [^{11}C](R)-rolipram the mean percentage error of the V_t parameter reached and overcame the value of -48%, that is definitely too high to be accepted. What is really stunning is that the performance of the method are highly dependant on the subjects used to define the ABIF.

How to determine if the subjects used are suitable or not is actually a big issue. In fact, it is not properly correct to speak about "good" subjects in absolute terms: their goodness is relative to the subject under analysis, they are good if their average shares the same shape of the one of the AIF to be estimated. Since in practice the only information known about this AIF are just two samples, the problem seems impractical.

One factor that could have influenced the results obtained with the [^{11}C](R)-rolipram dataset is the low number of subjects used for the ABIF definition (5 subjects). In fact with L-[1-^{11}C]leucine, where 15 subjects were used, the performance were much better. Increasing the size of the subset used would likely mitigate the presence of outliers but this does not solve the problem when the outlier is the subject under analysis himself.

Another difference occurred between the two analysis performed is the tissue parameter chosen: as regards the L-[1-^{11}C]leucine one, the parameter chosen was K_I; as concern the [^{11}C](R)-rolipram analysis, the distribution volume V_t was selected. K_I is a direct parameter, this means that its value is calculated directly from the SAIF (it is the trapping rate constant, it correspond to α_0). On the other side, V_t is a derived parameter, calculated from all the αs and βs obtained from the SA (see equation 2.7). The errors associated which each of these values propagates resulting in an higher uncertainty of the V_t value. Unfortunately, the use of derived parameters is necessary for clinical application, even with L-[1-^{11}C]leucine.

However, although all these problems just stressed are not trivial, the real issue with the ABIF method is another: the results can not be trusted. The error committed can be around 4% as well as around 50% without any known possibility to discern between the two cases. No clue if the parameters are over or underestimated. Even with L-[1-^{11}C]leucine there are no guarantees that the error committed could not overcome the 8% threshold and be much higher and this eventuality would be completely invisible to the operator. In substance the method lacks a way of assessing the accuracy of estimations.

3.6 Conclusion

ABIF approach is a relative simple technique that can be used when the full arterial samples are not available. Despite its ease of implementation, the method leaves some kind of arbitrariness to the operator and thus the

first step carried out was to define an optimal procedure. This was done checking all the possible variables that might affect the results, that are the calibration method, the times of the samples used for the calibration step and the number of subjects used for the ABIF definition. The goal was to determine the optimal combination of these factors. This means finding a trade-off between the performance reached with a particular configuration and the amount of data necessary to that particular combination (number of subjects, number of samples used, computation time, ...).

The optimal procedure was identified pas the one which involves the use of the samples at the times $t_{1,2} = 10, 50\,min$ in combination with the least squares calibration method. The number of subjects to be used was dependant on the tracer considered.

Nevertheless, even when all the optimal conditions are met, the performance of this method strongly depends on the subjects used for the ABIF construction and the subject under analysis itself. This was particularly evident with the [^{11}C](R)-rolipram dataset as the mean error committed in the V_t estimation raised from around 4% to more than 48% just changing the subjects averaged together.

With the L-[1-^{11}C]leucine dataset performance was better and the results seemed more robust. However, the big unsolved issue that affect the ABIF method is that there is no way to assess the goodness of the results: the tissue parameters estimations can be completely wrong but still looking as perfectly plausible.

In conclusion, the use of this approach should be limited to those arterial data sets with homogeneous kinetics when the full arterial sampling is impractical due to the patient condition and should be avoided when the purpose of the study involve statistical comparisons between different data sets.

Chapter 4

Arterial Input Function Modelling

In dynamic PET studies it may happen that a mathematical description of the AIF is required. For example when the samples are collected by an automatic machine, the data are too noisy to be used directly and they need to be fitted with an appropriate model. A model is also useful for estimating the measurement error associated with the samples and it can also be applied in simulation studies.

In this chapter one of the most commonly used models, Feng's model, was analyzed in detail and particular attention was paid on evaluating the best error law for describing the measurement error. The results shows that Feng's model is affected by some issues. For this reason, three new models were proposed and compared in order to overcome Feng's limitations.

4.1 State of the art: Feng's model

The model proposed by Feng in [11] is the solution of a LTI system composed by two simple and one repeated eigenvalues. Its expression is given by:

$$y(t) = \begin{cases} 0 & t \leq t_0 \\ [A(t-t_0) - B - C]e^{-\alpha(t-t_0)} + B \cdot e^{-\beta(t-t_0)} + C \cdot e^{-\gamma(t-t_0)} & t > t_0 \end{cases} \tag{4.1}$$

where β and γ are the simple eigenvalues while α is the repeated one, i.e. the α exponential has two coefficients: $A(t-t_0)$ and $-(B+C)$. The reason of the presence of the repeated eigenvalue is, according to Feng, due

to the cardiac output and blood flows. In fact the tracer arrives at and leaves from the measurement site transported in these ways. Both phases (arrival and leaving) are "fast" if compared to normal diffusion processes and thus the two eigenvalues associated should be high and close to each other. For this reason, they are assumed to be equal (α) and greater than β and γ.

As regards the t_0 term, this is a delay constant included to take into account the time between the tracer injection (in the venous system) and its arrival in the measurement site (artery). Obviously the tracer concentration before its arrival ($t \leq t_0$) is equal to 0 and thus the coefficients of the system are calculated so as to satisfy the initial conditions of tracer absence (their sum is equal to 0 when $t = t_0$). Furthermore, all parameters are assumed to be real-valued and positive.

4.1.1 Error Law

Before estimating the parameters of Feng's model, it was necessary to decide how to handle the measurement errors. However literature lacks of information about the error law to be associated with the arterial samples. For this reason four different laws were proposed and tested in order to determine the most correct. A few assumptions were common to all of them: the error term is a random variable v_i, additive, with zero mean and uncorrelated. This means that each datum z_i at sample time t_i can be expressed as:

$$z_i = y_i + v_i \tag{4.2}$$

where y_i is the value of the Feng's model at the time t_i.

In order to give a complete characterization of the error term v_i, its variance needs to be defined. The variance of the error that affects radioactive measurements is usually assumed to have a Poisson statistic [28]. That is:

$$Var(v_i) \propto z_i \tag{4.3}$$

Since the data available were corrected for the radioactivity decay, it was hypothesized that the error should reflect this correction too. For this reason a new variance was proposed, that is:

$$Var(v_i) \propto z_i \cdot e^{k \cdot t_i} \tag{4.4}$$

where $k = 0.034\,min^{-1}$ and it is the decay constant of the ^{11}C isotope. This second error law was called "exponential". Finally, other two error laws were proposed:

$$Var(v_i) \propto 1 \tag{4.5}$$
$$Var(v_i) \propto z_i^2 \tag{4.6}$$

called "constant SD" and "constant CV" error laws respectively. These two were introduced as they are really common in physiological models.

These four error laws were used to determine the weights to be associated with each datum in a WNLS (Weighted Non-linear Least Squares) estimator. More in detail, the relative weights assigned to each datum correspond to the inverse of the variance of the error associated with that datum. The term "relative" is used because the variance is supposed to be known up to a proportionally constant σ^2, which is estimated *a posteriori* as:

$$\sigma^2 = \frac{WRSS}{N - M} \tag{4.7}$$

where WRSS is the Weighted Residual Sum of Squares obtained with the estimated parameters, N the number of data and M the number of parameters.

4.1.2 Initial parameters

Since Feng's model is not linear in its parameters and due to the sensitivity of the estimates to the initial values, the fitting process required to define an approximate solution. An initial value for each parameter had to be estimated, from where the algorithm could start its research of the correct solution. The parameter t_0 was fixed as the time of the sample preceding the first sample whose value was more than 10% of the peak value. As regards the other parameters, the initial values where obtained implementing a modified version of the curve peeling method described in [19].

The assumption of this method is that the first two exponentials (α and β) do not give a significant contribute to the final portion of the curve, but this is determined almost exclusively by the last exponential. Thus,

the initial parameters C_0 and γ_0 can be estimated imposing the passage of that exponential for the last two samples. That is:

$$\gamma_0 = \frac{log(z_{end-1}) - log(z_{end})}{t_{end} - t_{end-1}} \tag{4.8}$$

$$C_0 = z_{end} e^{\gamma_0 (t_{end} - t_0)} \tag{4.9}$$

Once the last exponential is estimated, it can be subtracted from the data in order to obtain:

$$z'_i = z_i - C_0 e^{\gamma_0 (t_i - t_0)} \tag{4.10}$$

The previous step is repeated imposing the passage of the second exponential for two consecutive samples taken around the 10^{th} minute of the modified data z'. That is:

$$\beta_0 = \frac{log(z'_{10m}) - log(z'_{10m+1})}{t_{10m+1} - t_{10m}} \tag{4.11}$$

$$B_0 = z'_{10m} e^{\beta_0 (t_{10m} - t_0)} \tag{4.12}$$

The last two parameters are obtained imposing the passage to the peak (t_p, z_p). This results in really long calculus and the final formulation was reached with the aid of the software Wolfram Mathematica 8 [29]:

$$\begin{aligned} \alpha_0 = & \frac{1}{t_{p0} \left(C_0 e^{t_{p0}\beta_0} + e^{t_{p0}\gamma_0} \left(B_0 - e^{t_{p0}\beta_0} z_p\right)\right)} \cdot \\ & \left[e^{t_{p0}\gamma_0} \left(B_0 - e^{t_{p0}\beta_0} z_p + B_0 t_{p0} \beta_0\right) + C_0 e^{t_{p0}\beta_0} (1 + t_{p0}\gamma_0) + \right. \\ & \left. \left(C_0 e^{t_{p0}\beta_0} + e^{t_{p0}\gamma_0} \left(B_0 - e^{t_{p0}\beta_0} z_p\right)\right) W(tmp)\right] \end{aligned} \tag{4.13}$$

$$A_0 = \frac{B_0 \left(\alpha_0 - e^{t_{p0}(\alpha_0 - \beta_0)} \beta_0\right) + C_0 \left(\alpha_0 - e^{t_{p0}(\alpha_0 - \gamma_0)} \gamma_0\right)}{-1 + t_{p0}\alpha_0} \tag{4.14}$$

where

$$t_{p0} = t_p - t_0 \tag{4.15}$$

$$\begin{aligned} tmp = & - \frac{(B_0 + C_0)}{C_0 e^{t_{p0}\beta_0} + e^{t_{p0}\gamma_0} \left(B_0 - e^{t_{p0}\beta_0} z_p\right)} \cdot \\ & e^{\frac{C_0 e^{t_{p0}\beta_0} (-1 + t_{p0}\beta_0) + e^{t_{p0}\gamma_0} \left(B_0 (-1 + t_{p0}\gamma_0) - e^{t_{p0}\beta_0} z_p (-1 + t_{p0}(\beta_0 + \gamma_0))\right)}{C_0 e^{t_{p0}\beta_0} + e^{t_{p0}\gamma_0} \left(B_0 - e^{t_{p0}\beta_0} z_p\right)}} \end{aligned} \tag{4.16}$$

and where $W(\cdot)$ is the Lambert W function.

4.1.3 Evaluation of error law

Parameters were estimated with a WNLS estimator and the weights were determined by the four error laws saw in subsection 4.1.1. In order to choose the best error law, the following factors were considered:

1. weighted residuals: they have to be uncorrelated, with zero mean and variance equal to 1. Unfortunately there were too few samples to implement any statistical test, so the correlation was assess by visually inspecting the weighted residuals plot;

2. Residual Sum of Squares: RSS gives an indication of how good the model fits the data. The lower, the better.

	Error law	RSS	Weighted residuals		
			mean	Var	Correlation
L-[1-^{11}C]leucine	Constant SD	5260	-0.0273	0.8616	High
	Constant CV	7375	0.0427	0.1656	Low
	Poisson	5708	0.0243	0.7708	Low
	Exponential	5669	0.0253	0.7654	Medium
[^{11}C]DASB	Constant SD	0.026	-0.0464	0.8238	High
	Constant CV	0.057	0.0585	0.1430	Low
	Poisson	0.033	0.0359	0.7656	Low
	Exponential	0.032	0.0334	0.7615	Medium
[^{11}C]PiB	Constant SD	1587	-0.0522	0.8139	High
	Constant CV	3867	0.0600	0.1382	Low
	Poisson	2090	0.0269	0.7317	Low
	Exponential	2091	0.0270	0.7299	Medium
[^{11}C](R)-rolipram	Constant SD	7291	-0.0354	0.7582	High
	Constant CV	11599	0.0625	0.2621	Low
	Poisson	8340	0.0297	0.6956	Low
	Exponential	8288	0.0307	0.6922	Medium

Table 4.1. Impact of the different error laws on the Feng's model fit performance

From the table 4.1 emerges that the Poisson and constant CV error laws gives the less correlated residuals. Between the two, Poisson is associated with lower RSS, weighted residuals averages closer to zero and weighted residuals variances closer to one. For these reasons, Poisson error law was selected as the best for all data sets.

4.1.4 Feng's model limitations

Even if with the Poisson error law Feng's model gave the best results, its overall performances were poor. As can be seen in figure 4.1, the model is not able to describe correctly the curve. The peak is clearly underestimated and the weighted residuals are still correlated. This behaviour is common to all the four tracers considered.

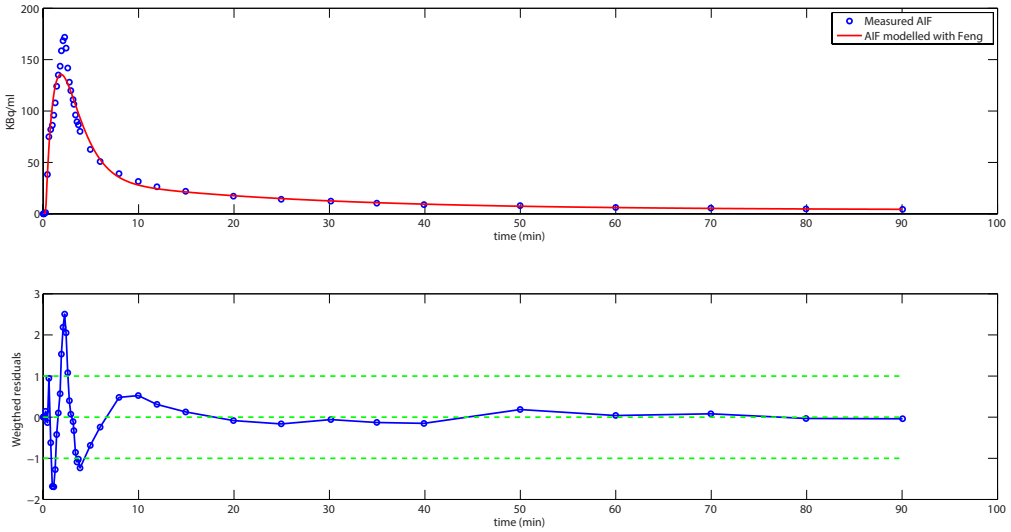


Figure 4.1. Example of Feng's model fit with a L-[1-^{11}C]leucine plasmatic concentration curve (above) and weighted residuals (below).

It has been speculated that this behavior may be due to the inability of the model to describe correctly the rising and the decreasing part at the same time. To verify this, it was tried to fit only the wash-out phase of the curve with Feng's model. Then, the resulting weighted residuals were aligned in order to make the t_p of the curve equal to zero. Subsequently, the residuals corresponding to the same times were averaged together (using nearest interpolation when the times did not coincide perfectly). The resulting curve is shown in figure 4.2.

If Feng's model were adequate to describe the wash-out phase, the residuals would have been random and their average would have been close to zero. The curve in figure 4.2 is referred to the L-[1-^{11}C]leucine dataset and it shows a complete different result: the residuals are still correlated and they follow an oscillating pattern. Similar results were found with the other data sets.

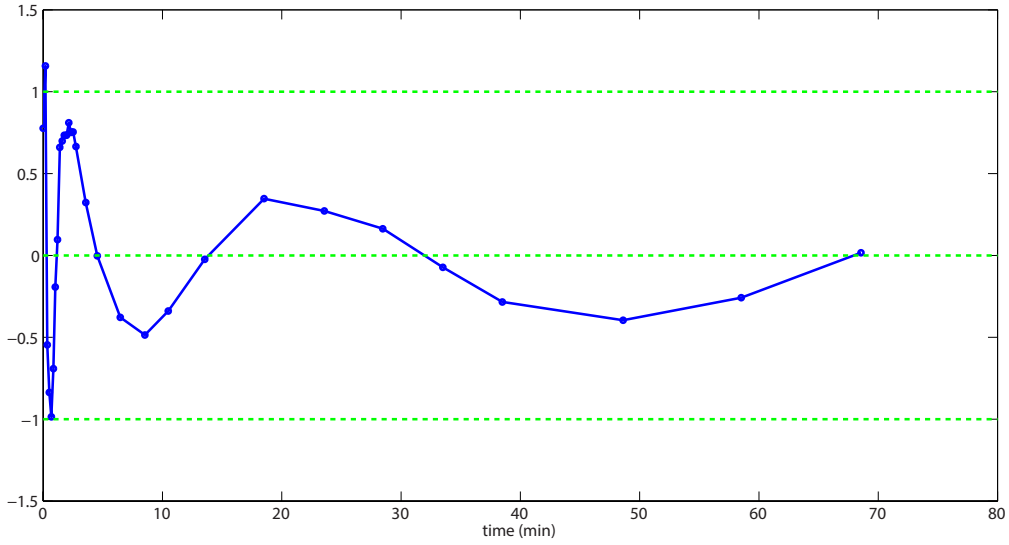


Figure 4.2. The curve shows the average weighted residuals obtained fitting the AIF curves after the peak with Feng's model. The dataset used for this picture is the L-[1-^{11}C]leucine one.

4.2 New Models

Three hypothesis were made regarding the inability of Feng's model to correctly describe the data, each of which gave rise to a new model:

1. Feng's model underestimates the peak because of limitations in the algorithm In order to take into account this possibility, the estimation was performed constraining two parameters so as to make the model passing for the peak. This new model was called Feng Constrained and it has the same formulation of Feng (equation 4.1) with the following constrains:

$$A = \frac{B\left(\alpha - e^{t_p(\alpha-\beta)}\beta\right) + C\left(\alpha - e^{t_p(\alpha-\gamma)}\gamma\right)}{-1 + t_p\alpha} \tag{4.17}$$

$$B = \frac{e^{t_p(\beta-\gamma)}\left(-Ce^{t_p\gamma} + e^{t_p(\alpha+\gamma)}z_p(-1+t_p\alpha) + Ce^{t_p\alpha}(1+t_p(-\alpha+\gamma))\right)}{e^{t_p\beta} + e^{t_p\alpha}(-1+t_p(\alpha-\beta))} \tag{4.18}$$

2. Considering the infusion time to be negligible is a wrong assumption. This possibility was tested modeling the infusion as a trans-

lated rect function of duration equal to T and applying it as input of the Feng's model. Since it is derived from a LTI system, the output is given by the convolution operator and thus this new model was called Feng*rect. Only for this model, the parameter t_0 was not fixed before the estimating step. The model expression for $t > t_0$ is:

$$\begin{aligned} y(t) = \int_{t_0}^{t} ([A(\tau - t_0) - B - C]e^{-a(\tau - t_0)} + Be^{-b(\tau - t_0)} + Ce^{-c(\tau - t_0)}) \cdot \\ [1(t - \tau) - 1(t - \tau - T)]d\tau \end{aligned} \tag{4.19}$$

3. The oscillations are due to the presence of a complex eigenvalue in the system. Complex eigenvalues in physiological systems are known to exist [30] and they manifest with a dumped oscillating mode. This can be modelled as a sin function modulated by an exponential function. The new model proposed consists of adding to the Feng's model this oscillating term. The new model was called FengComplex and its expression for $t > t_0$ is:

$$\begin{aligned} y(t) = &[A(t - t_0) - B - C - D\sin(\phi)]e^{-\alpha(t - t_0)} + Be^{-\beta(t - t_0)} + \\ &Ce^{-\gamma(t - t_0)} + De^{-\delta(t - t_0)}\sin[\omega(t - t_0) + \phi] \end{aligned} \tag{4.20}$$

Given the high number of parameters of this model (10), the fitting process was divided in three steps:

1. data were fitted with Feng's model and the residual were obtained;
2. residuals were fitted with the dumped sin function: $De^{-\delta(t-t_0)}\sin[\omega(t-t_0)+\phi]$;
3. the parameters just obtained were used as initial parameters for the whole model.

All the parameters of all the three new models were estimated assuming the measurements error could be described by the Poisson error law.

4.2.1 New models evaluation

In the evaluation of these three models, Feng Constrained was almost immediately discarded since its performance were even worse than the

original Feng's model. Figure 4.3 represents one of the best fit obtained with this model and it clearly shows that it is not adequate to describe correctly the data.

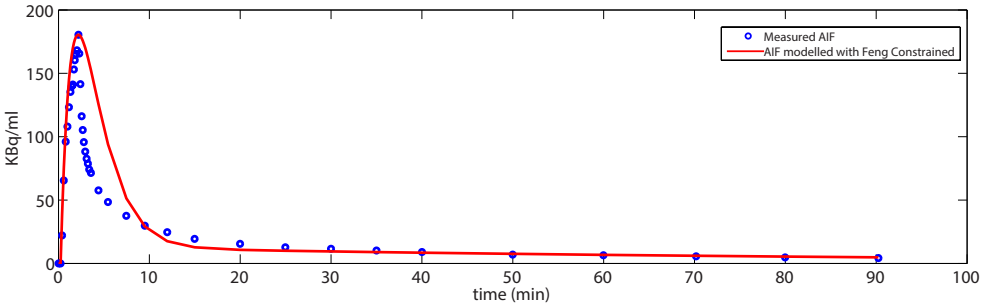


Figure 4.3. Example of Feng Constrained model with a L-[1-^{11}C]leucine plasmatic concentration curve.

In order to choose the best model between the two remaining, the principle of parsimony was followed. The aim was to find the model which is best able to fit the data with the minimum number of parameters. The Akaike information criterion (AIC) and the Schwarz criterion were used. The model which had the smallest criterion was the best. This two criteria are calculated in the following way:

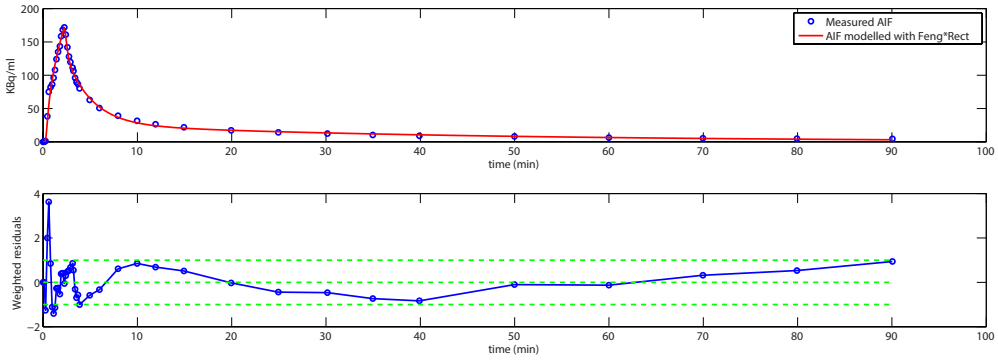
$$AIC = N \ln(\text{WRSS}) + 2P \tag{4.21}$$

$$SC = N \ln(\text{WRSS}) + P \ln(N) \tag{4.22}$$

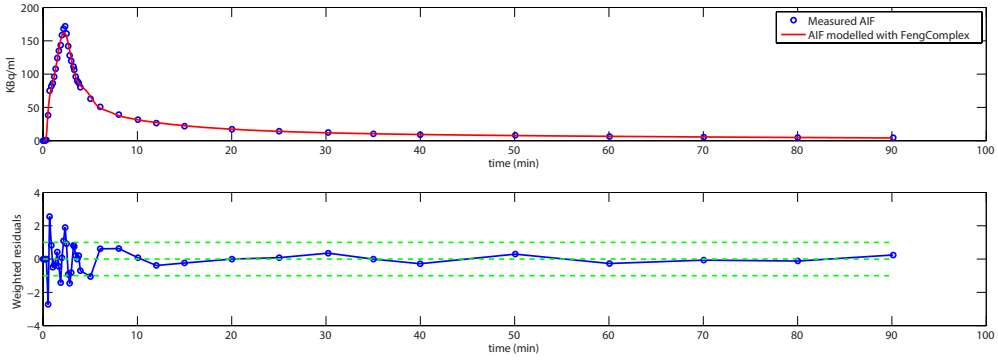
where N represents the number of samples, P the number of parameters of the model and WRSS is the Weighted Sum of Squared Residuals.

In table 4.2 the average of the two criteria are reported. Feng*rect always has the lowest value (except for the SC criterion in [^{11}C](R)-rolipram). For this reason it was chosen as the best model.

An example of two fits are shown in figure 4.4. As it can be seen, the Feng*rect model describes quite correctly the data even if the residuals are not properly uncorrelated. Plotting together the residuals obtained fitting the data with the Feng*rect model (figure 4.5) shows that the oscillating pattern is still present. Even with FengComplex the residuals maintain this behaviour.



(a) Feng*rect



(b) FengComplex

Figure 4.4. Example of Feng*rect and FengComplex models with a L-[1-^{11}C]-leucine plasmatic concentration curve.

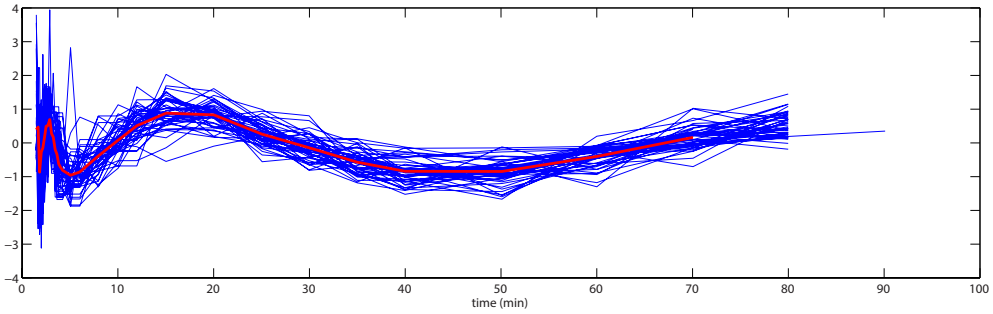


Figure 4.5. The weighted residuals obtained fitting the L-[1-^{11}C]leucine AIFs with Feng*rect model exhibit an oscillating behaviour

	Model	WRSS	AIC	SC
L-[1-^{11}C]leucine	FengComplex	9.61	57.35	53.15
	Feng*rect	8.67	49.65	46.71
[^{11}C]DASB	FengComplex	3.60	37.24	32.15
	Feng*rect	2.89	28.31	24.75
[^{11}C]PiB	FengComplex	4.07	38.29	33.06
	Feng*rect	5.19	35.46	31.80
[^{11}C](R)-rolipram	FengComplex	4.21	34.37	27.98
	Feng*rect	6.90	33.29	28.82

Table 4.2. WRSS, Akaike (AIC) and Schwarz (SC) indices of the FengComplex and Feng*rect models

4.3 Conclusion

In this study Feng's model was analyzed in detail. First, four possible error laws were compared in order to determine the best possible error law for the data sets considered, which resulted to be the Poisson error law. Second, the critical step of determining the initial values for the minimization algorithm was addressed implementing an automatic method for their estimation from the data.

However, Feng's model was designed to be used with [^{18}F]FDG [11] and when the tracer injection was a bolus. When used with other tracers or when the infusion protocol was different, the model proved to be not adequate to describe the data. For this reason new models were proposed.

Among all, the model called Feng*rect was chosen as the best. In fact it was able to describe rather correctly both the rising and the wash-out phase of the AIFs using the same number of parameters as Feng's model. Moreover, it estimated correctly the peaks and it did not require the t_0 parameter to be fixed before the estimation phase.

As regards the residuals, even if improved, they still were correlated. In particular they exhibit an oscillating behaviour whose cause is not known.

Chapter 5

SIMultaneous Estimation

SIMultaneous Estimation (SIME) is a technique that consists in estimating the tissue parameters of multiple Regions Of Interest (ROI) and the input function parameters simultaneously. This approach was mainly validated in simulation studies [10] [31] [32] or with [^{18}F]FDG data [33], while relatively little information can be found regarding other tracers [34].

In this final part of the study, the parametric model proposed in chapter 4 to describe the input function (Feng*rect) was used in the implementation of SIME method.

5.1 Methods

The compartmental models seen in subsection 2.2.1 describe the transfer function of a generic region in brain for a given tracer. This generic region can be seen as a LTI system, whose input is given by the plasmatic concentration of tracer and whose output is measured by the PET scan. If input and output functions are known, the estimation of the transfer function parameters can be done applying non-linear regression analysis. Nevertheless, it is not possible to obtain the input function and the system transfer function simultaneously from the single output function. However, from PET images multiple output functions of multiple ROI are available and, even if each of them is described by a different transfer function, they all share the same input function. It is therefore possible to estimate both the input and the transfer functions taking advantage of the uniqueness of the first.

5.1.1 Mathematical formulation

As parametric expression of the input function, the Feng*rect model proposed in chapter 4 was used. Its full expression is given by:

$$C_p(\theta, T, t) = \begin{cases} 0 & t \leq t_0 \\ \int_{t_0}^{t}([A(\tau - t_0) - B - C]e^{-a(\tau - t_0)} + Be^{-b(\tau - t_0)} + Ce^{-c(\tau - t_0)}) \times \\ [1(t - \tau) - 1(t - \tau - T)]d\tau & t > t_0 \end{cases} \tag{5.1}$$

where $\theta = [A, B, C, \alpha, \beta, \gamma, t_0]^T$ and T is the infusion time, supposed to be known.

The transfer function of a generic ith ROI is indicated as $h_i(K_i, t)$, where K_i represents the vector of parameters associated with that ROI. The output of the ith system, i.e. the tracer concentration C_{t_i}in the ith ROI, is thus given by:

$$C_{t_i}(\theta, K_i, T, t) = h_i(K_i, t) * C_p(\theta, T, t) \tag{5.2}$$

However, in the PET measures there is also a component related to the blood volume as well as the eventual presence of metabolites. This can be written as:

$$C_{m_i}(\theta, K_i, V_{b_i}, C_{other}, T, t) = f[C_{t_i}(\theta, K_i, T, t), V_{b_i}, C_{other}] \tag{5.3}$$

where V_{b_i} is the percentage of the ith ROI volume occupied by blood, C_{other} represents all the information about the tracer activity in whole blood including metabolites and $f[\cdot]$ is the function that combines all these factors. In this study C_{other} was supposed to be known.

The objective function related to the tissue to be minimize is thus given by:

$$\phi(\theta, K_1, \ldots, K_M, V_{b_1}, \ldots, V_{b_M}) = \tag{5.4}$$

$$\sum_{i=1}^{M}\sum_{j=1}^{N} w_i(t_j)[z_i(t_j) - C_{m_i}(\theta, K_i, V_{b_i}, C_{other}, T, t_j)]^2 \tag{5.5}$$

where M is the number of ROI, N is the number of data available for each ROI, $z_i(t_j)$ is the PET measure of the ith ROI at the time t_j and $w_i(t_j)$ is the weight to be associated with it, calculated as the inverse of its variance.

If nC arterial samples P_c are available the following term accounting for the plasmatic modelp must be added:

$$\sum_{c=1}^{nC} v_c[P_c - C_p(\theta, T, t_{sc})]^2 \tag{5.6}$$

where v_c represents the weight to be associated with the arterial samples (usually high) and t_{sc} is the sampling time of the c^{th} arterial sample.

5.1.2 Estimation

For each subject, five ROIs and the last three arterial samples were used. As initial parameters of the input function the values obtained fitting all the measured samples with the Feng*rect model were used. As regards the initial parameters of the ROI transfer functions, the value obtained with the Gold standard were used.

In order to further improve the parameter estimation, a bayesian prior was assigned to the input function parameters. This prior was calculated fitting the entire dataset (excluding the subject under analysis) with the Feng*rect model and calculating, for each parameter, mean and variance of the values estimated.

All the analysis were performed using the software SAAM II.

5.2 Materials

Initially this approach was thought to be applied to both the L-[1-^{11}C]-leucine and [^{11}C](R)-rolipram images dataset (see section 3.1). However, the objective function obtained with L-[1-^{11}C]leucine data proved to be particulary challenging to minimize and the algorithms used never converged to an acceptable solution. For this reason, only the results obtained with the [^{11}C](R)-rolipram dataset are presented.

5.3 Results

An example of the results that can be obtained are shown in table 5.1. As regards the transfer function, the average percentage error committed in the five ROIs respect to the values obtained with the gold standard is

below 20% for all parameters except V_b. Actually, V_b does not represent a kinetic parameter as the others, but a volume and it is known to be more problematic to estimate [34]. As concern the parameter of interest V_t, the error committed is around 4% in all ROIs.

Input function parameters	Error %	Transfer function parameters	Average error %	SD error %
A	13.28	k_1	-20	1.7
B	20.60	k_2	-4	2.0
C	-6.97	k_3	20	6.1
α	-1.00	k_4	-12	1.1
β	6.31	V_b	37	5.5
γ	-1.68	V_t	4	0.8
t_0	65.04			

Table 5.1. Percentage errors between the parameters estimated with the SIME method and the gold standard method of one [^{11}C](R)-rolipram subject. As regards the transfer function parameters, with average error is intended the mean error between the five ROIs.

The input parameters do not have a physiological meaning, however looking at figure 5.1 it can be seen that the input function is estimated quite well. Of course having chosen as initial parameters the value obtained fitting the Feng*rect model to all the measured samples and the use of a bayesian prior have undoubtedly contributed to reach this good performance. Nevertheless, the new model implemented in the SIME method have proved to be adequate to describe the data.

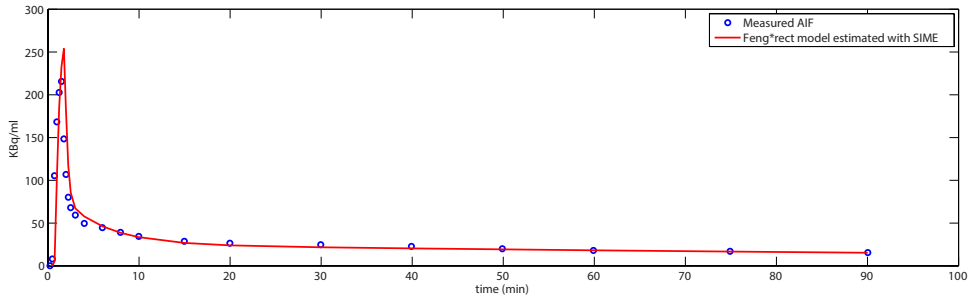


Figure 5.1. Input function as estimated with SIME of a [^{11}C](R)-rolipram subject compared to the measured input function.

5.4 Conclusion

The aim of this chapter was not to reach an optimal implementation of the SIME method and neither to evaluate its performance rigorously. The aim was to obtain some preliminary results about the robustness and flexibility of the new proposed model for the description of the AIF, the Feng*rect model.

Further studies are of course necessary to assess the model performance when compared to other models and when it is applied in less ideal condition. Moreover, the SIME method can be improved by implementing more sophisticated algorithm to minimize the objective function (simulated annealing was proposed in [31] and [34]); assigning optimal weights to each ROI; using prior on the V_b parameter; ...

Nevertheless, despite all the limitations of this study, the Feng*rect model proved to have a great potential and to give reliable estimation of the AIF when used with SIME.

Chapter 6

Conclusion

Dynamic PET quantification often involves the cannulation of the radial artery of the patient, a practice considered invasive and cause of discomfort for the patient. Furthermore, the handling and analysis of numerous blood samples require adequate facilities, special care and technical expertise by the operator. For these reasons in this study two alternatives to the arterial sampling, called ABIF and SIME respectively, were analyzed.

As regards the first, two goals were pursued: determining an optimal protocol evaluating the main factors that could affect ABIF performance and validating this approach with other tracers than [^{18}F]FDG. It has been found that despite the use of an adequate number of subjects, this method is not able to completely recover the intra-subjects variability. Moreover the ABIF definition is heavily conditioned by the characteristics of the individuals composing the population. With this method, the performance are therefore unpredictable.

For all these reasons, the use of this approach should be limited to those arterial data sets with homogeneous kinetics when the full arterial sampling is impractical due to the patient condition and should be avoided when the purpose of the study involve statistical comparisons between different data sets.

As concern SIME, the study focused on the parametric description of the arterial input function. The most commonly employed model found in literature, Feng's model, was originally designed for [^{18}F]FDG and when tested with the tracers considered in this study, it performed poorly. Among the new models proposed, the modification of Feng's model (taking into

account the tracer injection time) proved to be adequate to describe correctly both the rising and the wash-out phase of the AIFs. This new model does not introduce new parameters to the Feng's original model, thus its complexity was maintained unaltered.

Preliminary studies on the implementation of this new solution to the SIME method shown that it has a great potential and it can achieve reliable description of the AIF.

Bibliography

[1] M. Reivich and A. Alavi, *Positron emission tomography.* New York: A.R. Liss, 1985.

[2] T. V. B. Greitz, D. H. Ingvar, and L. Wide, *The metabolism of the human brain studied with positron emission tomography.* New York: Raven Press, 1985.

[3] P. E. Valk, *Positron emission tomography : basic science and clinical practice.* London: Springer, 2003.

[4] E. R. Carson and C. Cobelli, *Modelling methodology for physiology and medicine.* San Diego, Calif. ; London: Academic, 2001.

[5] B. A. Everett, M. A. Oquendo, A. Abi-Dargham, M. S. Nobler, D. P. Devanand, S. H. Lisanby, J. J. Mann, and R. V. Parsey, "Safety of radial arterial catheterization in pet research subjects," *J Nucl Med*, vol. 50, no. 10, p. 1742, 2009.

[6] G. Tomasi, C. B. Smith, and K. C. Schmidt, "Determination of regional rates of cerebral protein synthesis with the l-[1-11c]leucine pet method using a population-derived input function and venous sampling." Xth International Conference on Quantification of Brain Function with PET, May 2011.

[7] S. Takikawa, V. Dhawan, P. Spetsieris, W. Robeson, T. Chaly, R. Dahl, D. Margouleff, and D. Eidelberg, "Noninvasive quantitative fluorodeoxyglucose PET studies with an estimated input function derived from a population-based arterial blood curve," *Radiology*, vol. 188, pp. 131–136, Jul 1993.

[8] K. Wakita, Y. Imahori, T. Ido, R. Fujii, H. Horii, M. Shimizu, S. Nakajima, K. Mineura, T. Nakamura, and T. Kanatsuna, "Simplification for measuring input function of FDG PET: investigation of 1-point

blood sampling method," *J. Nucl. Med.*, vol. 41, pp. 1484–1490, Sep 2000.

[9] P. Zanotti-Fregonara, C. S. Hines, S. S. Zoghbi, J. S. Liow, Y. Zhang, V. W. Pike, W. C. Drevets, A. G. Mallinger, C. A. Zarate, M. Fujita, and R. B. Innis, "Population-based input function and image-derived input function for [(11)C](R)-rolipram PET imaging: Methodology, validation and application to the study of major depressive disorder," *Neuroimage*, Aug 2012.

[10] D. Feng, K. P. Wong, C. M. Wu, and W. C. Siu, "A technique for extracting physiological parameters and the required input function simultaneously from pet image measurements: theory and simulation study," *IEEE Trans Inf Technol Biomed*, vol. 1, no. 4, pp. 243–54, 1997.

[11] D. Feng, S. C. Huang, and X. Wang, "Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography," *Int J Biomed Comput*, vol. 32, no. 2, pp. 95–110, 1993.

[12] J. Rosenthal, A. Angel, and J. Farkas, "Metabolic fate of leucine: a significant sterol precursor in adipose tissue and muscle," *Am J Physiol*, vol. 226, no. 2, pp. 411–8, 1974.

[13] S. Bishu, K. C. Schmidt, T. Burlin, M. Channing, S. Conant, T. Huang, Z. H. Liu, M. Qin, A. Unterman, Z. Xia, A. Zametkin, P. Herscovitch, and C. B. Smith, "Regional rates of cerebral protein synthesis measured with l-[1-11c]leucine and pet in conscious, young adult men: normal values, variability, and reproducibility," *J Cereb Blood Flow Metab*, vol. 28, no. 8, pp. 1502–13, 2008.

[14] S. Houle, N. Ginovart, D. Hussey, J. H. Meyer, and A. A. Wilson, "Imaging the serotonin transporter with positron emission tomography: initial human studies with [11c]dapp and [11c]dasb," *Eur J Nucl Med*, vol. 27, no. 11, pp. 1719–22, 2000.

[15] R. V. Parsey, A. Ojha, R. T. Ogden, K. Erlandsson, D. Kumar, M. Landgrebe, R. Van Heertum, and J. J. Mann, "Metabolite considerations in the in vivo quantification of serotonin transporters using 11c-dasb and pet in humans," *J Nucl Med*, vol. 47, no. 11, pp. 1796–802, 2006.

[16] J. C. Price, W. E. Klunk, B. J. Lopresti, X. Lu, J. A. Hoge, S. K. Ziolko, D. P. Holt, C. C. Meltzer, S. T. DeKosky, and C. A. Mathis, "Kinetic modeling of amyloid binding in humans using pet imaging and pittsburgh compound-b," *J Cereb Blood Flow Metab*, vol. 25, no. 11, pp. 1528–47, 2005.

[17] H. Wachtel, "Potential antidepressant activity of rolipram and other selective cyclic adenosine 3',5'-monophosphate phosphodiesterase inhibitors," *Neuropharmacology*, vol. 22, no. 3, pp. 267–72, 1983.

[18] P. Zanotti-Fregonara, S. S. Zoghbi, J. S. Liow, E. Luong, R. Boellaard, R. L. Gladding, V. W. Pike, R. B. Innis, and M. Fujita, "Kinetic analysis in human brain of [11c](r)-rolipram, a positron emission tomographic radioligand to image phosphodiesterase 4: a retest study and use of an image-derived input function," *Neuroimage*, vol. 54, no. 3, pp. 1903–9, 2011.

[19] C. Cobelli, D. Foster, and G. Toffolo, *Tracer kinetics in biomedical research*. Kluwer Academic, 2000.

[20] C. B. Smith, K. C. Schmidt, M. Qin, T. V. Burlin, M. P. Cook, J. Kang, R. C. Saunders, J. D. Bacher, R. E. Carson, M. A. Channing, W. C. Eckelman, P. Herscovitch, P. Laverman, and B. K. Vuong, "Measurement of regional rates of cerebral protein synthesis with l-[1-11c]leucine and pet with correction for recycling of tissue amino acids: Ii. validation in rhesus monkeys," *J Cereb Blood Flow Metab*, vol. 25, no. 5, pp. 629–40, 2005.

[21] M. Veronese, A. Bertoldo, S. Bishu, A. Unterman, G. Tomasi, C. B. Smith, and K. C. Schmidt, "A spectral analysis approach for determination of regional rates of cerebral protein synthesis with the l-[1-(11)c]leucine pet method," *J Cereb Blood Flow Metab*, vol. 30, no. 8, pp. 1460–76, 2010.

[22] V. J. Cunningham and T. Jones, "Spectral analysis of dynamic pet studies," *J Cereb Blood Flow Metab*, vol. 13, no. 1, pp. 15–23, 1993.

[23] K. Schmidt, "Which linear compartmental systems can be analyzed by spectral analysis of pet output data summed over all compartments?," *J Cereb Blood Flow Metab*, vol. 19, no. 5, pp. 560–9, 1999.

[24] K. B. Contractor, L. M. Kenny, C. R. Coombes, F. E. Turkheimer, E. O. Aboagye, and L. Rosso, "Evaluation of limited blood sam-

pling population input approaches for kinetic quantification of [18F]fluorothymidine PET data," *EJNMMI Res*, vol. 2, p. 11, 2012.

[25] P. Zanotti-Fregonara, R. Maroy, M. A. Peyronneau, R. Trebossen, and M. Bottlaender, "Minimally invasive input function for 2-18F-fluoro-A-85380 brain PET studies," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 39, pp. 651–659, Apr 2012.

[26] A. B. G. Rizzo, M. Veronese, "Spectral analysis kinetic estimation: an integrated tool for quantification of pet data." IXth International Symposium on Functional Neuroreceptor Mapping of the Living Brain, August 2012.

[27] Y. Wu and R. E. Carson, "Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging," *J Cereb Blood Flow Metab*, vol. 22, no. 12, pp. 1440–52, 2002.

[28] E. R. Carson and C. Cobelli, *Modelling methodology for physiology and medicine*. San Diego, Calif. ; London: Academic, 2001.

[29] "http://www.wolfram.com/mathematica."

[30] J. T. Fagarasan and r. Distefano, J. J., "Hidden oscillations in generalized linear mammaillary compartmental models," *Math Biosci*, vol. 93, no. 1, pp. 79–95, 1989.

[31] W. Koon-Pong, S. R. Meikle, F. Dagan, and M. J. Fulham, "Estimation of input function and kinetic parameters using simulated annealing: application in a flow model," *Nuclear Science, IEEE Transactions on*, vol. 49, no. 3, pp. 707–713, 2002.

[32] D. Y. Riabkov and E. V. Di Bella, "Estimation of kinetic parameters without input functions: analysis of three methods for multichannel blind identification," *IEEE Trans Biomed Eng*, vol. 49, no. 11, pp. 1318–27, 2002.

[33] K. P. Wong, D. Feng, S. R. Meikle, and M. J. Fulham, "Simultaneous estimation of physiological parameters and the input function–in vivo pet data," *IEEE Trans Inf Technol Biomed*, vol. 5, no. 1, pp. 67–76, 2001.

[34] R. T. Ogden, F. Zanderigo, S. Choy, J. J. Mann, and R. V. Parsey, "Simultaneous estimation of input functions: an empirical study," *J Cereb Blood Flow Metab*, vol. 30, no. 4, pp. 816–26, 2010.

List of Tables

List of Figures