



UNIVERSITY OF CLAUDE BERNARD LYON 1  
Polytech Lyon

---

**Estimation of Image-Derived Arterial Input Function in  
Brain PET Imaging:  
Application to Modeling PET Dynamics of Glucose  
Metabolism in Patients with Impaired Consciousness**

---

By SEPAND ALI MADAD SOLTANI

Master 2 in Medical Device Engineering  
Internship Report

Supervised by INÉS MÉRIDA and NICOLAS COSTES

Academic Advisor: KEVIN TSE VE KOON

2024-2025



# Abstract

[TODO]

**Keywords:** Dynamic FDG-PET, Image-Derived Input Function, Hybrid PET/MRI, Bayesian Framework

# Contents

<b>Abstract</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Positron Emission Tomography . . . . .	1
1.2 Kinetic Modeling . . . . .	2
1.3 Input Function . . . . .	3
1.3.1 Arterial Input Function . . . . .	3
1.3.2 Population-Based Input Function . . . . .	3
1.3.3 Image Derived Input Function . . . . .	3
1.4 Partial Volume Effect . . . . .	4
1.5 Background . . . . .	4
<b>2 Materials and Methods</b>	<b>6</b>
2.1 Dataset Description . . . . .	6
2.2 Pre-processing . . . . .	6
2.3 Carotid Segmentation . . . . .	7
2.4 Partial Volume Correction . . . . .	8
2.4.1 Geometric Transfer Matrix . . . . .	8
2.4.2 Bayesian Geometric Transfer Matrix . . . . .	9
2.5 Plasma Fraction Correction . . . . .	11
2.6 Evaluation . . . . .	12
2.6.1 IF Curves . . . . .	12
2.6.2 Quantification . . . . .	12
2.7 Simulation . . . . .	13
<b>3 Results</b>	<b>15</b>
3.1 Carotid Segmentation from MRA-TOF . . . . .	15
3.2 Hyperparameter Tuning . . . . .	16
3.3 Simulation . . . . .	17
3.4 IDIF . . . . .	19
3.4.1 $^{18}\text{F}$ FDG Dataset . . . . .	19
3.4.2 $^{11}\text{C}$ Yohimbine Dataset . . . . .	20
3.4.3 Simulated Dataset . . . . .	21
<b>4 Discussion</b>	<b>27</b>
<b>5 Conclusion</b>	<b>28</b>

**References****29**



# Introduction

## 1.1 Positron Emission Tomography

Positron Emission Tomography (PET) is an in vivo functional imaging technique widely used in clinical and research settings to monitor physiological and biochemical processes. In PET, a biologically active molecule is labeled with a positron-emitting radioisotope, serving as a radiotracer, and then injected into the body. As the radiotracer accumulates in target tissues, its radioactive decay produces positrons, which interact with electrons to emit pairs of gamma photons in nearly opposite directions. These photons are detected by the PET scanner, and image reconstruction algorithms generate a three-dimensional representation of the tracer distribution. This imaging modality allows for the investigation of metabolic changes, receptor binding, and other biochemical processes, providing invaluable information in oncology, neurology, cardiology, and other fields.

There are two main categories in PET image acquisition: static imaging and dynamic imaging. Static PET involves acquiring a single scan after the radiotracer injection. This single snapshot offers a powerful yet simplified view of tracer distribution. The common quantification metric in static imaging is the Standardized Uptake Value (SUV), which normalizes tissue uptake by the injected dose and weight of the subject, allowing for a semi-quantitative comparison of tracer accumulation across different tissues or over time [1]. Due to its simplicity, static PET is widely used in clinical settings; however, it also has limitations. Because it reflects only one time point, the SUV cannot capture the temporal dynamics of tracer uptake and clearance, and various physiological factors may influence its measurements, thereby reducing its accuracy.

Dynamic PET imaging provides a more comprehensive view of radiotracer kinetics by acquiring a series of images over a period ranging from a few minutes to more than an hour post-injection, depending on the tracer type. Instead of a single snapshot, dynamic imaging produces time-activity curves (TAC) that illustrate how tracer concentration in each tissue changes throughout the scanning period. This approach enables the measurement of physiological parameters such as the tracer rate of influx ( $K_i$ ), volume of distribution ( $V_T$ ), Binding Potential (BP), and the rates of phosphorylation and dephosphorylation.

## 1.2 Kinetic Modeling

To quantify pharmacokinetic parameters, kinetic modeling is employed. Compartmental modeling is the most popular and is considered the most accurate approach in kinetic modeling. In compartmental modeling, the distribution and kinetics of a radiotracer are described by dividing the system into distinct compartments, each representing a pool of tracer that behaves uniformly. Interactions between compartments can be unidirectional or bidirectional, meaning the tracer may either move in and out or only enter a compartment. Various graphical models (e.g., the Logan [2] and Patlak [3] methods), as well as classical compartmental model fitting approaches, are used to analyze tracer kinetics.

Figure 1.1 shows the two-tissue compartment model (2TCM), also known as the three-compartment model, in series mode. This model comprises one tissue compartment for the free tracer,  $C_F(t)$ , and another for the receptor-bound tracer,  $C_B(t)$ , in addition to an external compartment representing the tracer concentration in the plasma or blood, denoted as the input function  $C_P(t)$ .

The tracer kinetics are governed by a series of first-order differential equations, in which the exchange rates between the compartments are considered constant:

$$\frac{dC_F(t)}{dt} = K_1 C_P(t) - (k_2 + k_3)C_F(t) + k_4 C_B(t), \quad (1.1)$$

$$\frac{dC_B(t)}{dt} = k_3 C_F(t) - k_4 C_B(t), \quad (1.2)$$

where  $K_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$  are the constant rate parameters.

The total radiotracer tissular kinetic measured by PET (the PET data),  $C_T(t)$ , is given by

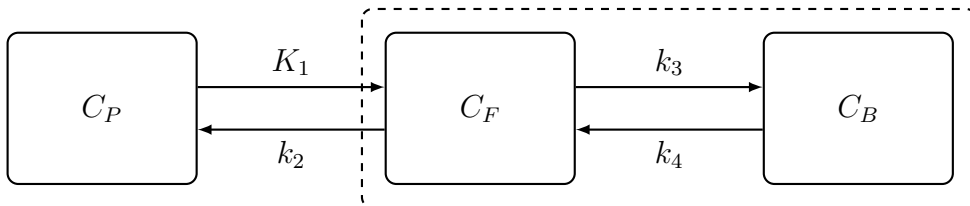
$$C_T(t) = C_F(t) + C_B(t) + C_P(t). \quad (1.3)$$

Thus to solve this system of equations and to estimate  $K_1$ ,  $k_2$ , and  $k_3$  parameters, we must fit the model using the measured PET TACs ( $C_T$ ) and the input function ( $C_P$ ).

For [ $^{18}\text{F}$ ]FDG quantification, the metabolic rate of glucose ( $\text{MR}_{\text{glu}}$ ) is calculated as

$$\text{MR}_{\text{glu}} (\mu\text{mol}/\text{min}/100\text{g}) = \frac{[C]}{LC} \cdot \frac{K_1 \times k_3}{k_2 + k_3}. \quad (1.4)$$

where  $[C]$  denotes the glucose concentration, and  $LC$  is the lumped constant.



**Figure 1.1:** Schematic of the two-tissue compartment model (2TCM)



For [ $^{11}\text{C}$ ]Yohimbine quantification, we utilize the Volume of Distribution ( $V_T$ ) which is the ratio of radiotracer concentration in the target tissue ( $C_T$ ) to the plasma ( $C_P$ ):

$$V_T = \frac{C_T}{C_P}$$

Using the Logan plot method this can be directly estimated from these two values or to be fitted to a compartment model. In the latter case,  $V_T$  can be calculated as

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

## 1.3 Input Function

### 1.3.1 Arterial Input Function

The arterial input function (AIF) is considered the gold standard for obtaining the input function. It is determined by inserting an arterial catheter into the patient and continuously drawing blood samples to measure the radiotracer concentration, thereby obtaining the blood activity curve used in the quantification model. However, this procedure is invasive and can cause discomfort, potentially discouraging patients from undergoing future examinations. Furthermore, this method is labor-intensive and requires trained personnel to manage both the subject and the measurement devices.

### 1.3.2 Population-Based Input Function

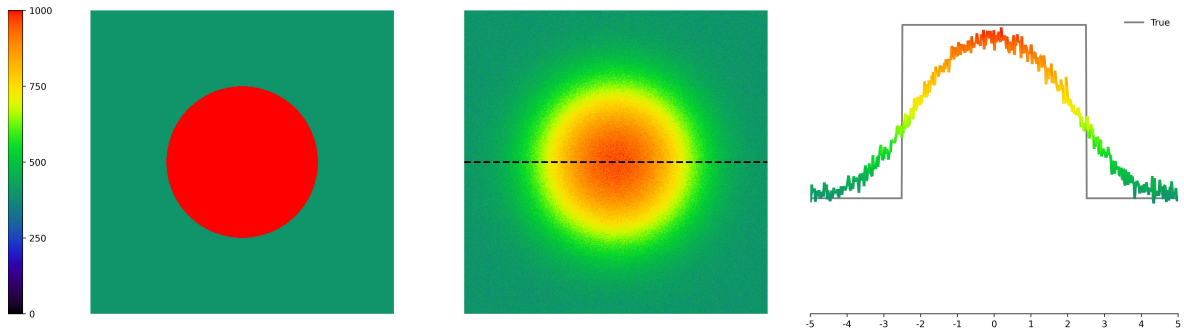
Population-Based Input Function (PBIF) is a method for replacing the subject specific AIF with the average AIFs of a population of other subjects. In practice, an average curve is derived from a representative cohort and then temporally aligned and scaled to the subject using one or more calibration points (e.g., early blood samples or image-derived peaks). PBIF can reduce invasiveness and acquisition burden, but it may introduce bias if the cohort is not well matched to the subject or if calibration is suboptimal.

### 1.3.3 Image Derived Input Function

The image-derived input function (IDIF) has been proposed as a non-invasive alternative for obtaining the input function. IDIF techniques typically involve identifying vascular structures or regions with high blood pool activity within the imaging field and extracting the input function directly from the PET images. For example, in whole-body, cardiac or small animal PET studies the aorta is visible in the in the Field of View (FOV) of the PET camera and is used as the source for IDIF. In brain PET imaging, the Internal Carotid Arteries (ICA) are the largest vessels present in the FOV and have a diameter of approximately 5 mm much smaller than of the aorta. Their smaller size makes IDIF more challenging due to Partial Volume Effects

## 1.4 Partial Volume Effect

Partial Volume Effect (PVE) is the loss and mixing of signal that occurs when structures are small relative to the system’s spatial resolution, causing activity to “spill out” of small objects and “spill in” from neighbors. In PET, this leads to systematic underestimation of arterial activity and contamination from surrounding tissue, especially for vessels like the ICA that have similar size as the effective resolution of PET camera (both around 5mm). In Figure ??, this effect is shown [TODO] Correcting PVE is essential for accurate IDIF estimation and downstream kinetic modeling.



## 1.5 Background

Many methods have utilized the first few frames of the dynamic PET, when the tracer has not yet reached tissue and predominantly resides in the arteries, to obtain a mask of the carotids [4]. These implementations range from fully manual approaches [TODO], to semi-automatic methods that use a few seed voxels for region growing or to initialize morphological operations [TODO], and to fully automatic pipelines based on deep learning such as custom Convolutional Neural Networks (CNNs) [TODO] and U-Net architectures [TODO].

However, even with sophisticated approaches, ICA segmentation directly from PET suffers from strong PVE. With the emergence of hybrid PET/MRI systems, it has become feasible to acquire both functional and anatomical data simultaneously. MRI provides high-resolution soft-tissue contrast, while PET captures metabolic activity. For instance, time-of-flight MR angiography (TOF-MRA) delivers excellent arterial contrast. Many studies have leveraged this to achieve accurate ICA segmentation using manual or semi-automatic procedures, typically relying on high-intensity thresholding and seeded or unseeded region growing [TODO].

Unlike T1-weighted MRI, TOF-MRA is not always included in research and clinical protocols. Meanwhile, T1-weighted scans are routinely acquired for anatomical localization and image registration, among other purposes, which makes them a practical source for arterial segmentation when TOF-MRA is not part of the protocol. In this context, **TODO** [TODO] proposed a deep learning model that infers the ICA directly from T1-weighted images. Similarly, the caliPER software [TODO] enables semi-automatic extraction of arterial masks from either TOF-MRA or T1-weighted data.

As noted above, even with a high-resolution anatomical arterial mask, PVE remains and must be corrected. Partial Volume Correction (PVC) methods estimate spill-in and spill-out coefficients between the ICA and surrounding tissue. Recovery Coefficients (RC) are commonly used, obtained by scanning cylindrical phantoms of various diameters and matching the ICA diameter to the closest phantom [TODO, TODO, TODO, TODO]. The Geometric Transfer Matrix (GTM) method generalizes this by estimating regional spillover based on volume geometry and an explicit point spread function (PSF) model [TODO]. However, these linear mixing models do not account for additional effects such as time-varying noise or motion, and they typically treat the problem as a fixed linear unmixing ( $SP \cdot C_{in} - SO \cdot C_{out}$ ) without propagating uncertainty, which can limit recovery in challenging conditions.

More advanced approaches include model-based matrix factorization that jointly estimates the input function and tissue activity [TODO], as well as deep learning methods trained to infer the IDIF from PET and MRI inputs [TODO, TODO].

In this work, we propose a fully automatic pipeline for extracting the ICA mask from TOF-MRA. We then employ a Bayesian framework that couples GTM-based mixing with priors on the input function and tissue kinetics, explicitly modeling noise to improve IDIF estimation [5].

# Materials and Methods

## 2.1 Dataset Description

Two experimental datasets were available. In the first study, 59 acute comatose patients were included between 7 days and 30 days after coma onset ( $46 \pm 16$  years old; 21 females). PET data were acquired in list mode for 90 minutes following an intravenous bolus injection of [ $^{18}\text{F}$ ]FDG. The second study included 7 healthy subjects ( $25 \pm 3$  years old; all male) who received an intravenous bolus of [ $^{11}\text{C}$ ]Yohimbine.

Both datasets used the same imaging protocol. Using a Siemens Healthineers Biograph mMR simultaneous MR–PET system, an arterial time-of-flight MR angiography (TOF-MRA) was acquired in axial orientation with a voxel size of  $0.3 \times 0.3 \times 0.7$  mm, and a T1-weighted MRI was acquired in axial orientation with isotropic 1 mm voxels, during the 90-minute dynamic PET. Raw PET data were rebinned into 24 time frames (variable-length frames:  $8 \times 15$  s,  $3 \times 60$  s,  $5 \times 120$  s,  $1 \times 300$  s,  $7 \times 600$  s) sinograms for dynamic reconstruction. Reconstruction yielded a voxel size of  $1.04 \times 1.04 \times 2.08$  mm<sup>3</sup> in a matrix of  $344 \times 344 \times 127$  voxels.

In the [ $^{18}\text{F}$ ]FDG study, whole-blood and plasma AIFs were measured from 26 manually collected arterial samples (timing: every 5 s for the first minute, every 15 s until the second minute, and at 3, 5, 10, 20, 30, 45, 60, 75, 80, 85, and 90 minutes post-injection) and counted in a gamma counter. In the [ $^{11}\text{C}$ ]Yohimbine study, 25 arterial blood samples were manually collected (timing: every 5 s for the first minute, every 10 s until the second minute, and at 5, 10, 30, 45, 60, and 90 minutes post-injection). The blood samples were counted in a gamma counter, then centrifuged to separate plasma, and the plasma activity was measured to compute the plasma fraction as the ratio of plasma to whole-blood activity at each time point.

## 2.2 Pre-processing

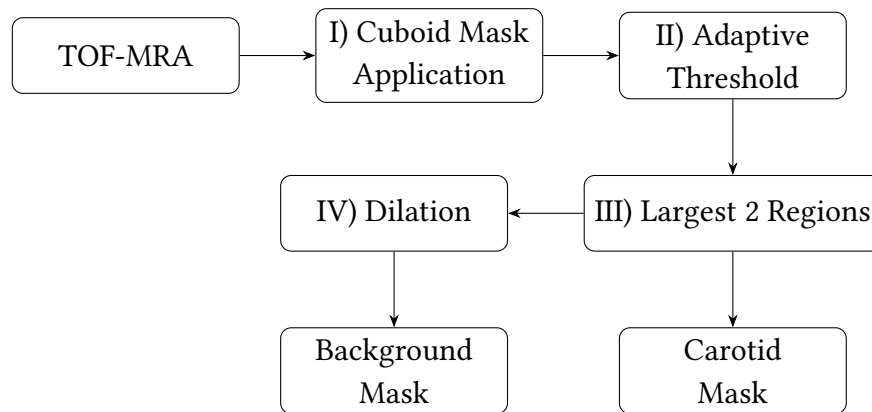
For both studies, the T1-weighted image was registered to the average PET image and to the TOF-MRA using the NiftyReg program, an affine registration method [TODO]. The two resulting affine transformation matrices were then composed to register the TOF-MRA directly to the PET space. Even though patients in the [ $^{18}\text{F}$ ]FDG study were unconscious and PET and MRI were acquired simultaneously in the same session, we performed this step

out of caution to eliminate the possibility of misregistration. The T1-weighted image served as an intermediary because it shares anatomical features with both modalities, whereas directly registering TOF-MRA to PET is impractical due to PET's lower spatial resolution and the limited axial field-of-view of TOF-MRA. Finally, the brain in the T1-weighted image was segmented into regions of interest (ROIs) using the Hammersmith brain atlas [6] to obtain regional masks, and these masks were applied to the dynamic PET to extract regional TACs.

## 2.3 Carotid Segmentation

Figure 2.1 summarizes the carotid segmentation pipeline. Because vessels appear hyperintense in TOF-MRA, a high-intensity threshold can extract arterial structures. However, lesions—which were common in the comatose cohort—and venous structures can also appear hyperintense, and may be inadvertently selected by thresholding. To exclude these, a cuboid volume of interest (VOI) was defined in a reference space to cover the anatomical region where the ICAs are most likely to appear (Step I). The threshold was set to the 1st percentile of all non-zero voxels in the image and applied only within the VOI (Step II). A 3D connected-components filter was then applied, and the two largest connected components were retained as the left and right internal carotid arteries (Step III).

The reference image was the standard MNI152 atlas, padded by 50% in the inferior direction (negative  $z$  in voxel space) because the original atlas excludes the subcranial region relevant for the ICAs. The TOF-MRA was registered to this reference using an affine transformation, and the resulting matrix was used to map the VOI into TOF-MRA space before applying the thresholding and connectivity steps.



**Figure 2.1:** Carotid and background mask segmentation pipeline

## 2.4 Partial Volume Correction

### 2.4.1 Geometric Transfer Matrix

As we discussed before in ??, direct extraction of the radioactivity in the arteries is not practical due to PVC. Geometric Transfer Matrix aims to account for this loss of signal by considering the observed TACs are linear combination of the true value and other effecting regions [7].

Here we define two regions, the carotid and the surrounding tissues (background). A mask for extracting the activities of the latter was obtained by dilating the carotid mask by 5 mm and subtracting the voxels corresponding to the carotid mask (Figure 2.1, Step IV).

$$\underbrace{\begin{bmatrix} T_c \\ T_{bg} \end{bmatrix}}_{\text{Observed}} = \underbrace{\begin{bmatrix} \omega_{c \rightarrow c} & \omega_{bg \rightarrow c} \\ \omega_{c \rightarrow bg} & \omega_{bg \rightarrow bg} \end{bmatrix}}_{\text{GTM}} \cdot \underbrace{\begin{bmatrix} T_{IF} \\ T_{tissue} \end{bmatrix}}_{\text{Unknown}}, \quad (2.1)$$

where  $\omega_{n \rightarrow m}$  are the spill-in and spill-over coefficient of region  $n$  onto region  $m$ , which is obtained by convolving the binary mask of region  $n$  with the system's point spread function and integrating the resulting intensity over region  $m$ , normalized by the total signal in region  $m$ . where

$$\omega_{n \rightarrow m} = \frac{\int_{\Omega_m} (h * \chi_n)(r) dr}{\int_{\Omega_m} (h * \chi_m)(r) dr}, \quad (2.2)$$

with  $\chi_n$  and  $\chi_m$  denoting the binary masks of regions  $n$  and  $m$ , respectively,  $h$  the system's point spread function, and  $\Omega_m$  the spatial domain of region  $m$ .

$T_c$  and  $T_{bg}$  are respectively the observed carotid and background TACs and  $T_{IF}$  and  $T_{tissue}$  are the real unknown TACs of the carotid (the input function) and the background tissue.

By inverting the GTM, this system of equations can be solved to recover the arterial and background TACs. However, GTM and other classical PVC methods often fail to fully recover the lost signal because they simplify the problem and ignore time-varying noise characteristics. As a result, they may propagate or even amplify noise. Time-varying noise arises from several factors, including the small diameter of the ICAs, very short early frames when the IF changes rapidly, and the exponential decay of the tracer, all of which reduce gamma counts detection by the PET camera and increase variance.

## 2.4.2 Bayesian Geometric Transfer Matrix

### Modelling

For each subject,  $T_{IF}$  is modeled as a linear combination of a population mean and principal components obtained by applying principal component analysis (PCA) to AIFs from the cohort. Specifically, for each subject, a subset randomly selected subjects—excluding the subject under study—is used to build a PCA basis. After zero-centering the training AIFs, PCA yields principal axes  $\phi_i(t)$  with explained variances  $\lambda_i$ . We define scaled components  $v_i(t) = \sqrt{\lambda_i} \phi_i(t)$ , and write

$$T_{IF}(t) = \mu_{IF}(t) + \sum_{i=1}^p \theta_i v_i(t), \quad (2.3)$$

where  $\mu_{IF}(t)$  is the population mean and the coefficients satisfy  $\theta_i \sim \mathcal{N}(0, 1)$  by construction. Here,  $p$  denotes the number of retained components.

Spectral analysis (SA) models tissue activity as the convolution of the input function with an impulse response written as a sum of exponentials [TODO]. This makes it more flexible and generalizable as it does not assume a particular compartment model and is convenient for describing background signal. Therefore, the background TAC is

$$T_{bg}(t) = \sum_{i=1}^s \alpha_i (T_{IF} \otimes e^{-\beta_i t})(t), \quad (2.4)$$

where  $\otimes$  denotes convolution,  $\alpha_i$  is the amplitude, and  $\beta_i$  is the decay rate of the  $i$ th spectral component.

### Noise

Accurate noise modeling is challenging because multiple sources contribute to the variance. Some effects (e.g. scatter, randoms, and motion) are partially corrected during reconstruction post-processing. The main contributor is low-count statistics, an inherent limitation of PET due to detector performance and overall camera sensitivity, which is particularly problematic for IDIF because the ICAs are small, early frames are very short during the bolus passage, and radioactive decay reduces detected events, increasing variance across frames.

While raw PET counts are Poisson-distributed, post-reconstruction noise is typically approximated as Gaussian [TODO]. To account for time-varying noise, we use per-frame weights to normalize variance across frames:

$$\omega_i = \frac{\Delta t_i}{c_i} \exp\left(-\frac{t_i \ln 2}{T_{1/2}}\right), \quad (2.5)$$

where  $\Delta t_i$  is the frame duration,  $c_i$  is the net true counts for frame  $i$ , and  $T_{1/2}$  is the tracer half-life. Because  $c_i$  was unavailable, we substitute  $C_T(t_i)$ , the total reconstructed activity concentration at the frame mid-time.

The time-varying noise level is summarized by a weighted average variance:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N \omega_i \sigma_i^2. \quad (2.6)$$

### Estimation

Let  $\mathcal{D}$  denote the observed PET TAC, and let  $\Theta = (\theta_1, \dots, \theta_p, \alpha_1, \beta_1, \dots, \alpha_s, \beta_s)$  be the unknown model parameters, with an unknown Gaussian noise variance  $\sigma^2$ . We aim to estimate the joint posterior  $p(\Theta, \sigma^2 \mid \mathcal{D})$ .

By Bayes' rule,

$$p(\Theta, \sigma^2 \mid \mathcal{D}) \propto p(\mathcal{D} \mid \Theta, \sigma^2) \pi(\Theta) \pi(\sigma^2), \quad (2.7)$$

where  $p(\mathcal{D} \mid \Theta, \sigma^2)$  is the likelihood and  $\pi(\cdot)$  are the priors. We can therefore sample the posterior once the likelihood and priors are obtained.

The observed data are linked to the latent TACs through GTM ???. For convenience, define

$$\mathcal{T}(t) = \mathcal{G}(t; \Theta) = \mathcal{G} \left( \begin{bmatrix} T_{IF}(t; \theta_1, \dots, \theta_p) \\ T_{tissue}(t; \alpha_1, \beta_1, \dots) \end{bmatrix} \right), \quad (2.8)$$

where  $\mathcal{G}$  is the GTM operator and  $\mathcal{T}$  is the predicted activity (after PVE). Assuming Gaussian noise with frame weights  $\omega_i$ , the likelihood is

$$p(\mathcal{D} \mid \Theta, \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{\omega_i (\mathcal{D}(t_i) - \mathcal{T}(t_i))^2}{2\sigma^2} \right). \quad (2.9)$$

We refer to this framework as the Bayesian Geometric Transfer Matrix (Bayesian GTM or BGTM).

The prior factorizes as

$$\pi(\Theta) = \prod_{i=1}^p \pi(\theta_i) \prod_{j=1}^s \pi(\alpha_j) \pi(\beta_j). \quad (2.10)$$

As noted above,  $\theta_i \sim \mathcal{N}(0, 1)$ . For the spectral parameters, we adopt broad uniform priors,

$$\alpha_i, \beta_i \sim \mathcal{U}(u_{\min}, u_{\max}). \quad (2.11)$$

For the noise variance, we use an inverse-gamma prior,  $\pi(\sigma^2) \sim \Gamma^{-1}(a_0, b_0)$ , conjugate to the Gaussian likelihood. We center this prior around an empirical weighted mean squared deviation between the observed carotid TAC and the mean AIF:

$$\mathbb{E}[X] = \frac{1}{N} \sum_{i=1}^N \omega_i (T_c(t_i) - \mu_{IF}(t_i))^2. \quad (2.12)$$

Given a chosen coefficient of variation (CV), we set

$$a_0 = 2 + \frac{1}{\text{CV}^2}, \quad b_0 = (a_0 - 1) \mathbb{E}[X]. \quad (2.13)$$



## Sampling

Direct evaluation of the posterior ?? is intractable, so we employ Monte Carlo sampling. We use a Markov chain Monte Carlo scheme with a Metropolis-within-Gibbs strategy.

Gibbs updates draw each parameter from its univariate conditional posterior  $p(\Theta_i \mid \mathcal{D}, \Theta_{-i}, \sigma^2)$  and the noise variance from  $p(\sigma^2 \mid \mathcal{D}, \Theta)$ . Metropolis–Hastings proposals explore  $\Theta$ : at iteration  $m$ , propose  $\Theta'_i = \Theta_i^{(m)} + \varepsilon_i b$ , where  $\varepsilon_i$  is the step size and  $b$  is a Brownian perturbation. Accept with probability

$$\min\left(1, \frac{p(\Theta'_i \mid \mathcal{D}, \Theta_{-i}, \sigma^2)}{p(\Theta_i^{(m)} \mid \mathcal{D}, \Theta_{-i}, \sigma^2)}\right). \quad (2.14)$$

Because the inverse-gamma prior is conjugate, the conditional posterior for  $\sigma^2$  is also inverse-gamma,

$$p(\sigma^2 \mid \mathcal{D}, \Theta) \sim \Gamma^{-1}(a, b), \quad (2.15)$$

with

$$a = a_0 + \frac{N}{2}, \quad b = b_0 + \frac{1}{2} \sum_{i=1}^N \omega_i (\mathcal{D}(t_i) - \mathcal{T}(t_i))^2. \quad (2.16)$$

Thus,  $\sigma^2$  can be sampled directly without Metropolis–Hastings.

To improve mixing, each parameter  $k$  has its own proposal step size  $\epsilon_k$ . During burn-in, step sizes are adapted every  $L = 50$  Metropolis–Hastings updates by an arbitrary acceptance rate of 0.5: if the empirical rate exceeds 0.5, increase  $\epsilon_k \leftarrow 1.1 \epsilon_k$ ; otherwise decrease  $\epsilon_k \leftarrow 0.9 \epsilon_k$ . After burn-in, step sizes are fixed and subsequent samples are retained.

## Inference

After sufficient sampling, we use a modified maximum a posteriori (MAP) estimate by averaging the parameters of the top 0.1% of samples ranked by posterior probability and reporting this average as the solution.

## 2.5 Plasma Fraction Correction

By construction, IDIFs extracted from PET represent whole-blood activity. For  $[^{18}\text{F}]\text{FDG}$ , plasma metabolites are generally negligible over the scan, so the whole-blood AIF was used without plasma or metabolite correction [TODO]. For  $[^{11}\text{C}]\text{Yohimbine}$ , kinetic modeling requires the plasma parent activity, so whole-blood IDIFs were converted to plasma and corrected for metabolites using

$$C_{P,\text{parent}}(t) = \frac{C_{WB,\text{IDIF}}(t)}{R_{bp}} \times \text{PPF}(t), \quad (2.17)$$

where  $R_{bp} = 0.661$  is the whole-blood-to-plasma ratio and  $PPF(t)$  is the plasma parent fraction modeled as a 3 exponential curve: [TODO]

$$PPF(t) = 0.30 e^{-t/0.88} + 0.09 e^{-t/11.49} + 0.61 e^{-t/65.77}. \quad (2.18)$$

Because  $V_T$  estimation uses total plasma parent activity, no additional free-fraction correction was applied.

## 2.6 Evaluation

### 2.6.1 IF Curves

The performance of the proposed IDIF estimation was first evaluated by computing the mean absolute error (MAE) between the cumulative area under the curve (cAUC) of the estimated IDIF and the *ground true* AIF. cAUC was considered to be a more suitable metric since it provides an integrated measure of tracer exposure over time and is less sensitive to local fluctuations or noise in the curve compared to the directly comparing the TACs.

$$cAUC(t) = \int_0^t IF(\tau) d\tau, \quad (2.19)$$

where  $IF$  is the input function.

### 2.6.2 Quantification

However, because the cAUC error does not fully capture the impact of IDIF deviations on kinetic parameters, absolute quantification was also performed to evaluate the performance of the estimated IDIF against the gold standard AIF. Quantification was performed using graphical methods implemented in TPCCLIB [8]. The Patlak plot was used for  $[^{18}\text{F}]\text{FDG}$  [3], and the Logan plot for  $[^{11}\text{C}]\text{Yohimbine}$  [2].

The brain atlas was applied to the PET image and regional TACs were obtained by averaging voxels over time. For  $[^{18}\text{F}]\text{FDG}$ , regional  $\text{MR}_{\text{glu}}$  was computed. The mean absolute percentage error (MAPE) of  $\text{MR}_{\text{glu}}$  in each ROI was then calculated and averaged across the dataset:

$$\text{Average MAPE}(\text{MR}_{\text{glu}}) = \frac{100}{N} \sum_{i=1}^N \left( \frac{1}{N_{\text{ROI}}} \sum_{j=1}^{N_{\text{ROI}}} \left| \frac{\text{MR}_{\text{glu},ij}^{\text{IDIF}} - \text{MR}_{\text{glu},ij}^{\text{AIF}}}{\text{MR}_{\text{glu},ij}^{\text{AIF}}} \right| \right), \quad (2.20)$$

where  $N$  is the number of subjects.

Similarly, for  $[^{11}\text{C}]\text{Yohimbine}$ , the volume of distribution ( $V_T$ ) was calculated and the dataset error was computed as

$$\text{MAPE}(V_T) = \frac{100}{N} \sum_{i=1}^N \left( \frac{1}{N_{\text{ROI}}} \sum_{j=1}^{N_{\text{ROI}}} \left| \frac{V_{T,ij}^{\text{IDIF}} - V_{T,ij}^{\text{AIF}}}{V_{T,ij}^{\text{AIF}}} \right| \right). \quad (2.21)$$

Additionally, linear least-squares regression was performed between the regional  $MR_{glu}$  and  $V_T$  obtained using AIF and IDIF for each subject. The coefficient of determination ( $R^2$ ) and regression slope ( $S$ ) were computed per subject. The mean absolute percentage errors of these metrics across the dataset are

$$MAPE(R^2) = \frac{100}{N} \sum_{i=1}^N |R_i^2 - 1| \quad (2.22)$$

and

$$MAPE(S) = \frac{100}{N} \sum_{i=1}^N |S_i - 1|, \quad (2.23)$$

where  $R_i^2$  and  $S_i$  denote the coefficient of determination and slope for subject  $i$ , respectively.

## 2.7 Simulation

PET simulations are integral to validating statistical and analytical PET methods because they provide ground truth that may be unavailable or affected by human or instrumental error in experimental data. In the previous section, we treated the AIF as the ground truth and evaluated methods against it. However, because AIF measurement and analysis involve multiple manual steps and devices operated by different personnel, errors can occur.

Therefore, as a complementary validation, we conducted realistic simulations and evaluated the method on these data as well. PET-SORTEO is a Monte Carlo PET simulation platform [TODO] whose accuracy has been validated repeatedly [TODO, TODO] and can closely approximate real acquisitions. This fidelity strongly depends on accurate descriptions of both the physical anatomy and tracer kinetics.

As input, PET-SORTEO requires the description and characteristics of the PET machine, a 3D emission phantom segmenting the tissues into emitting regions, a 3D attenuation phantom describing tissue attenuation coefficients, and TACs for each emitting region. For this study, we simulated the  $[^{18}F]$ FDG dataset using the Biograph mMR model to mirror the experimental datasets, and we derived numerical phantoms and TACs from the real MR and PET data.

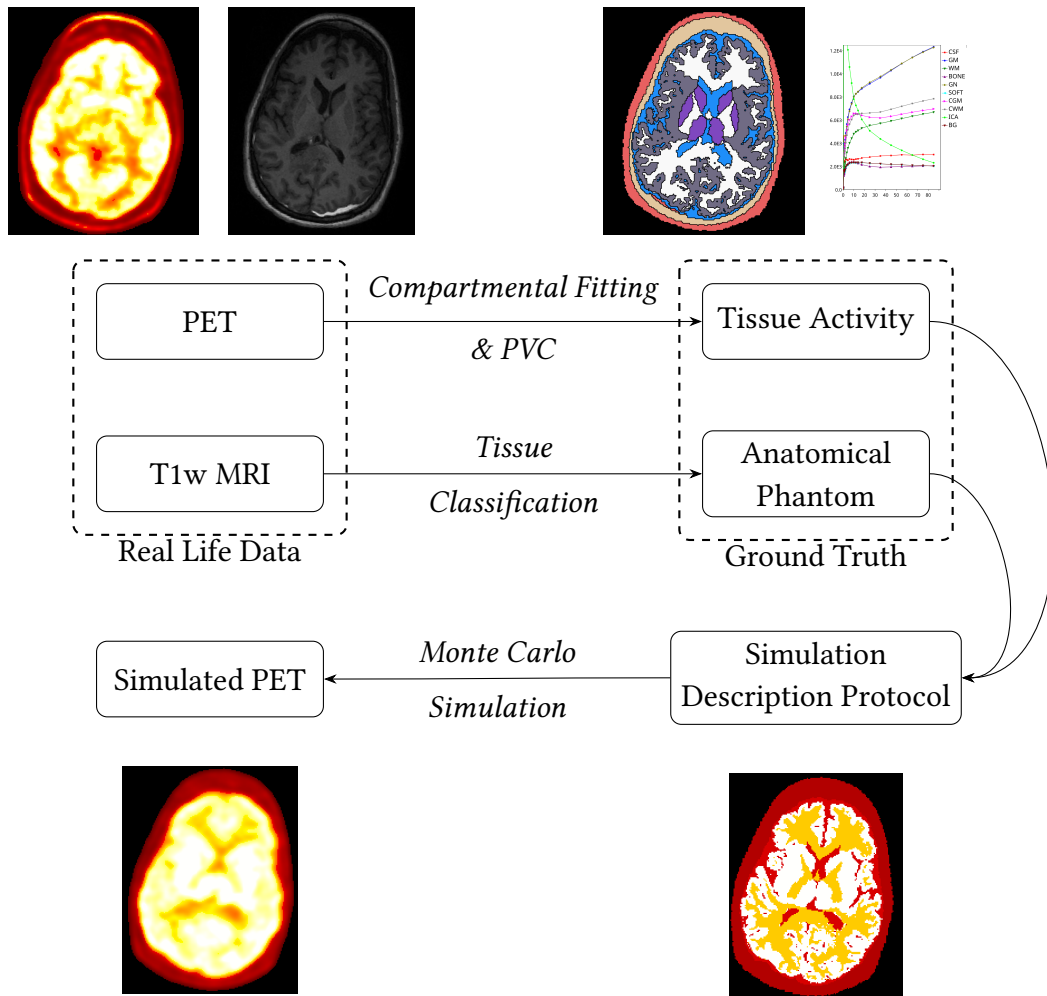
For each subject, first, the brain atlas subregions were merged into larger regions: white matter (WM), gray matter (GM), gray nuclei (GN), cerebellar gray matter (CGM), and cerebellar white matter (CWM). An algorithm originally developed for pseudo-CT generation from T1-weighted MRI for bone attenuation correction was used to create masks for the extracerebral regions[TODO]. Values above 300 were labeled as bone (BONE) and values between  $-300$  and  $300$  after excluding the brain as extracerebral soft tissues (SOFT) which includes the scalp, neck muscles, glands, eyes, etc. The leftover space between the skull and the brain was labeled as cerebrospinal fluid (CSF). Finally, the arterial mask (ICA) was obtained from the TOF-MRA segmentation (Section 2.3). These nine regions were combined to form the emission phantom.

The attenuation phantom was generated from the pseudo-CT by segmenting SOFT and BONE and assigning their respective attenuation coefficients.

The emission phantom was then registered to PET space to extract subject-specific TACs. These TACs were corrected for PVE using GTM and subsequently fitted with a 2TCM to denoise and obtain "ideal" TACs. The AIF used for fitting was the subject's measured AIF, itself fitted to an exponential model to remove noise [feng].

Several practical considerations were applied. Our SOFT definition is broad; while adequate for simulation, it violates homogeneity assumptions in compartment modeling, so the unfitted TAC was retained for SOFT. The pseudo-CT algorithm was permissive for BONE, leading to overestimated BONE TACs due to spill-in; empirically, BONE activity was scaled by 0.5.

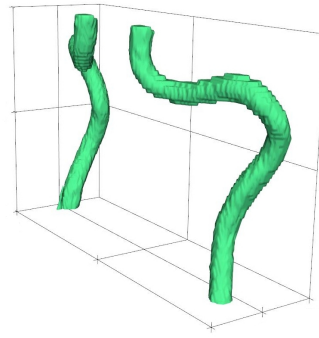
Extracting the correct activity of the CSF with this approach is not feasible since it's a very thin region which is heavily affected by PVE. Thus, CSF TAC was considered 0.2 of the fitted WM TAC.



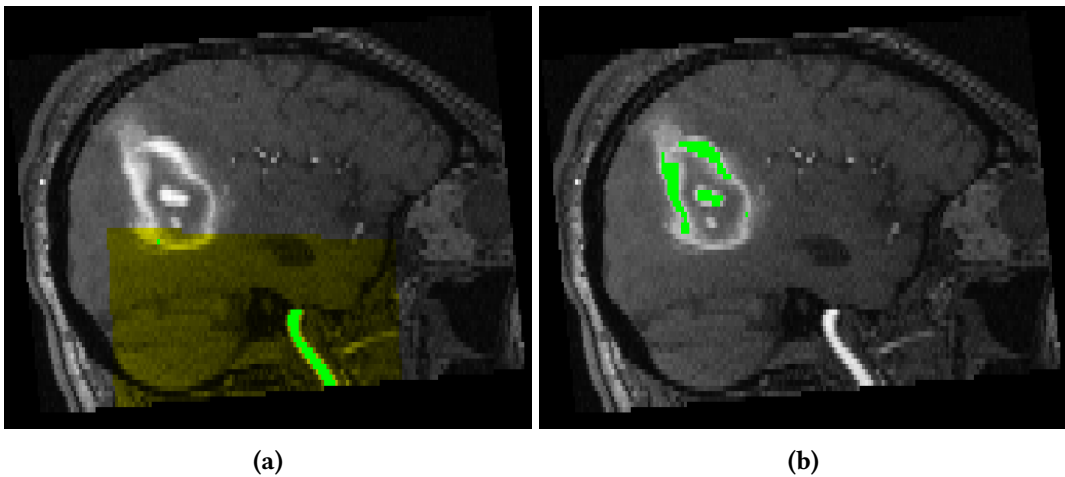
# Results

## 3.1 Carotid Segmentation from MRA-TOF

As illustrated in Figure 3.2, the cuboid mask plays a crucial role in the carotid segmentation. Because no ground truth segmentation is available, visual inspection was used to evaluate the results. In Figure ??, we see the 3D model of segmented ICA of a subject.



**Figure 3.1:** 3D visualization of segmented internal carotid arteries



**Figure 3.2:** Comparison of carotid segmentation (green) with (a) and without (b) a cuboid mask (yellow). In the absence of the cuboid mask, the segmentation algorithm fails to capture the carotid and instead incorrectly identifies the brain lesion

### 3.2 Hyperparameter Tuning

The number of principle components  $p$  was chosen as 3 by analysing the explained variance ratio plot in Figure ???. We see with only 3 components we get more than 90% explained variability. To prove the availability of the method, we seek to find the lowest possible number of population needed for the PCA in our  $[^{18}\text{F}]$ FDG dataset. To do this, the experiment was repeated 10 times each time with a different random seed for choosing population for  $N = \{5, 10, 15, 20, 30, 40, 50, 60\}$  and the median quantification errors were compared. As apparent in Figure ??, we see a plateau in the performance after  $N = 20$  and it was chosen as the ideal population sample count. For the  $[^{11}\text{C}]$ Yohimbine dataset since it only has 7 subjects, the maximum amount  $N = 6$  was chosen.

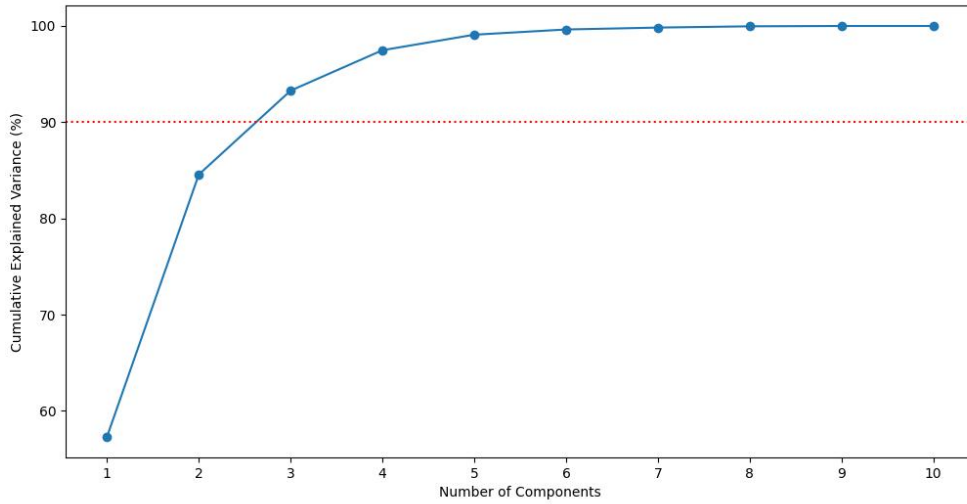


Figure 3.3

In classical model fitting using SA, an appropriate spectral range is chosen based on the radioisotope half-life (e.g.  $10^{-4}\text{s}^{-1}$  to  $1\text{s}^{-1}$  for  $^{18}\text{F}$ ) and is then divided into  $s = 100$  or  $s = 1000$  fixed frequencies ( $\beta$ ) then their amplitudes ( $\alpha$ ) are fitted to the TAC [TODO]. The peaks in the resulting spectrum are considered as the [TODO] frequencies of the TAC. However, estimating 100 or 1000 different amplitude parameters using MCMC would not be computationally feasible. Instead we must choose an appropriate number of spectral components and let the model to tune the frequencies and the amplitudes together.

To derive the number of basis functions, we assumed the activity in the surrounding tissue can either be modelled as 1TCM or 2TCM which their Impulse Response Function (IRF) have respectively one and two basis functions. Experiments with two basis functions showed that the model rarely results in two distinctive frequencies and would generally converge to similar values for both frequencies or one frequency would converge to infinity meaning no significant impact. Thus for both datasets one basis function was chosen.

As discussed in Section ??, the prior distribution used for the PCA weighting coefficients was  $\theta_i \sim \mathcal{N}(0, 1)$ . The prior for the SA parameters were considered a very uninformative

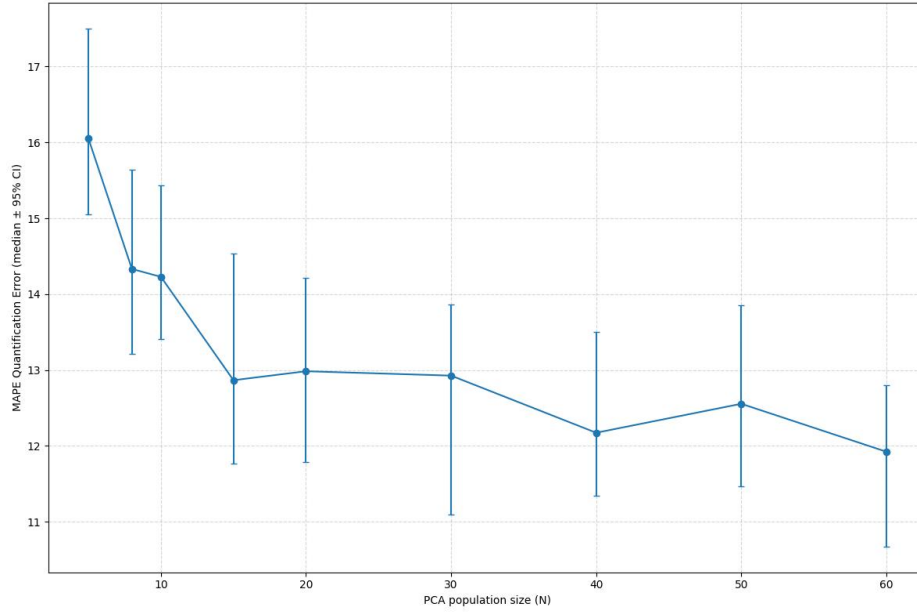


Figure 3.4

uniform distribution  $\mathcal{U}(10^{-5}, 10^{-2})$  as we did not have concrete prior knowledge on these parameters.

### 3.3 Simulation

From the 59 subjects in the  $^{18}\text{F}$ FDG dataset, 24 subjects were excluded due to having very large brain tumors or opened skulls due to open brain surgery which caused the tissue classification algorithm to fail. The rest of the 35 subjects were used to create the numerical phantoms and to generate TACs for the simulation protocol. PET-SORTEO utilized the protocols to simulate realistic PET kinetic and exported PETs in sinogram format. The sinograms were reconstructed with the same reconstruction program and settings as the experimental dataset to ensure uniformity with the experimental dataset.

Although the point of the simulation was not to replicate the same scan exactly but only use it as an inspiration for how tracer kinetic behavior, the results of the simulations were compared to the real scans as a quality control step. In Figure ?? we see an example of the 2nd (top) and last frame (bottom) sagittal slice of the real PET(inspiration source), the simulation input, and the simulation output from left to right. As expected we see less details in the neck area (SOFT region) since we did not consider the complete arterial and venous structure as well as glands. But more importantly, we can clearly see the activity in the internal carotids in the 2nd frame which roughly coincides with the peak of the AIF. In the last frame we can also see the steady-state activity of the brain to be very realistic and have realistic contrast in different regions.

In Figure ??, we see an example of the comparison of the Partial Volume Corrected TAC of the simulation output and the simulation input. As expected we notice slight discrepancy

due to lost signal as noise in the larger regions such as GM, CGM, and WGM since they are less affected by PVE and a very noticeable bias in the ICA which is affected the most by PVE.

As a post processing step, the brain atlases were also applied to the reconstructed simulated PETs to extract regional TTACs for the quantification step.

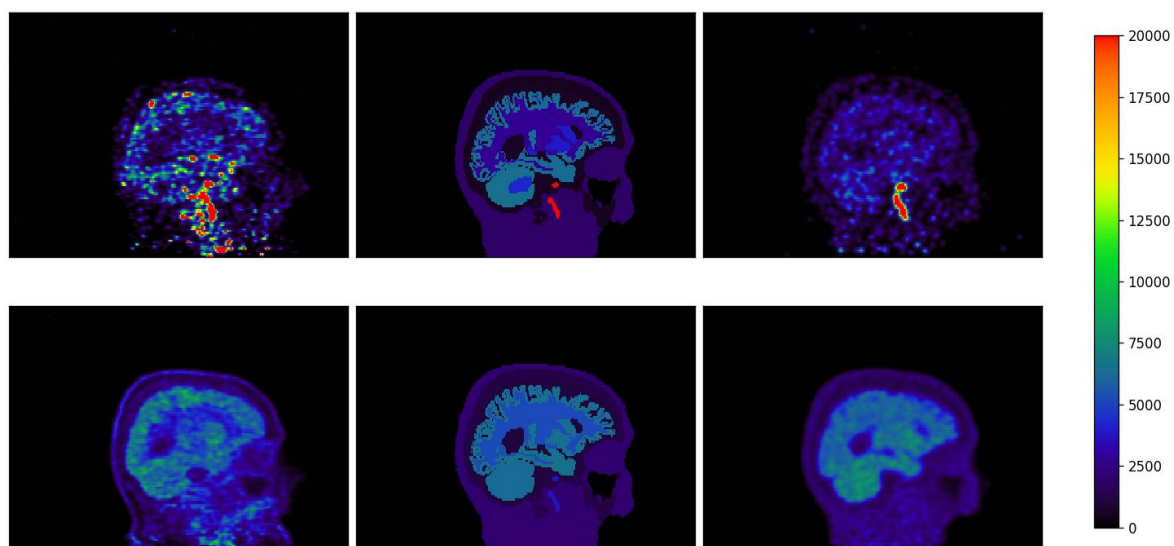


Figure 3.5

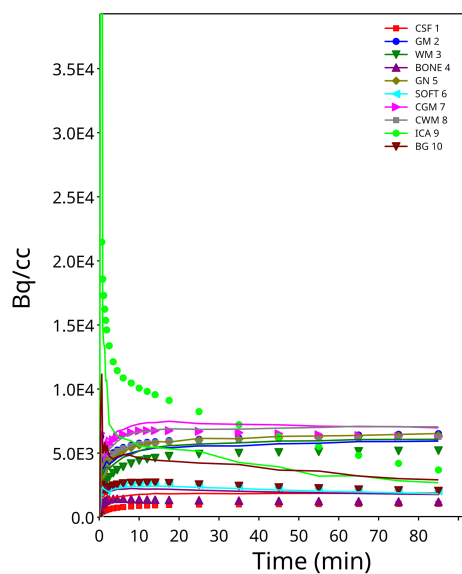


Figure 3.6



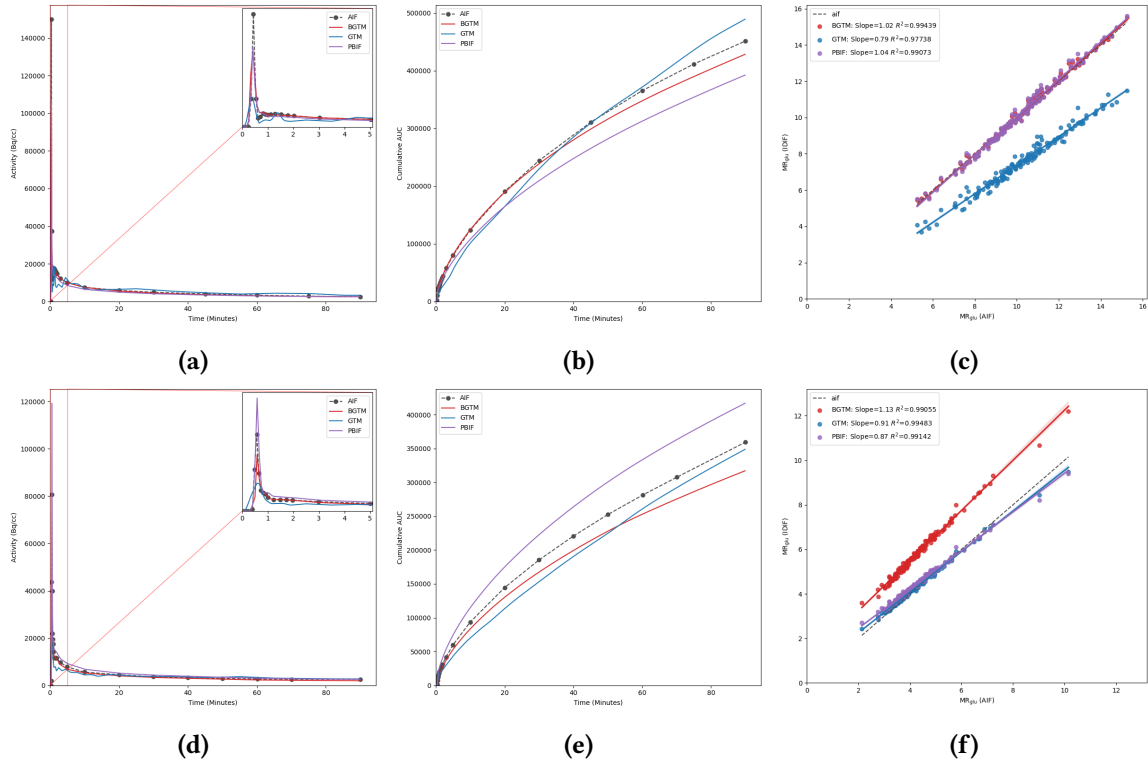
### 3.4 IDIF

Since BGTM relies on the GTM method and also depends on a population AIF. Thus the performance should be compared compared to the IDIF derived from GTM method and also the Population Based Input Function (PBIF) which is the mean AIF used in the PCA ( $\mu_{IF}$ ). The Markov Chain was run for 230,000 iteration of which the first 30,000 were in the burn-in stage.

#### 3.4.1 $[^{18}\text{F}]$ FDG Dataset

The reported statistic are from the experiment with the median results from the fine-tuning stage.

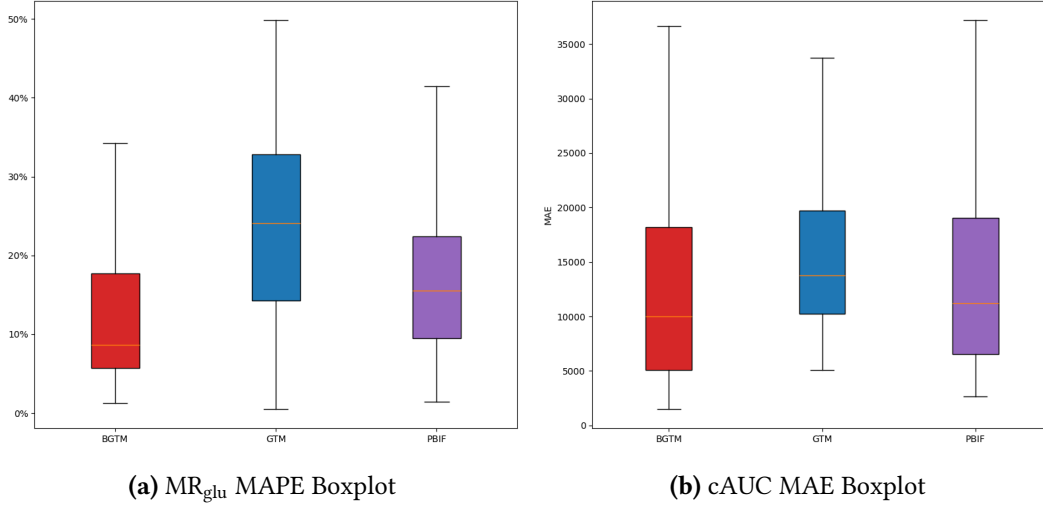
Figure 3.7 compares the two methods for one of the best- and worst-performing subjects. In the well-performing subject, BGTM significantly outperforms GTM and PBIF ( $\text{MR}_{\text{glu}}$  MAPE of 1.2% vs. 26% and 1.6% respectively); however, in the poorly performing case, BGTM falls short of GTM and PBIF ( $\text{MR}_{\text{glu}}$  MAPE of 36% vs. 2.2% and 4.75% respectively).



**Figure 3.7:** Comparison of the IFs (a,d), cumulative AUC curves (b,e), and  $\text{MR}_{\text{glu}}$  regression lines (c,f) for one of the best-(top row) and worst-performing(bottom row) subjects in the  $[^{18}\text{F}]$ FDG dataset

The average mean absolute error of the cAUC curves across the dataset was 13,024 for BGTM and 15,709 and 14,630 for GTM and PBIF respectively (Figure 3.8b). In quantification BGTM resulted in better performance compared to GTM and PBIF methods. Specifically, the BGTM, GTM, and PBIF methods achieved an average  $\text{MR}_{\text{glu}}$  mean absolute percentage

error (MAPE) of 13%, 24%, 17%, respectively (Figure 3.8a), and an average  $MR_{glu}$  MAE of 1.04, 1.98, and 1.4. In addition, the MAPE for the coefficient of determination ( $R^2$ ) and the regression slope ( $S$ ) were 1.75%, 11.2% for BGTM, compared to 4.1% and 17% for GTM and 1.5% and 15.75% for PBIF, respectively.



**Figure 3.8:** Boxplot of curve and quantification errors for the  $[^{18}F]$ FDG dataset

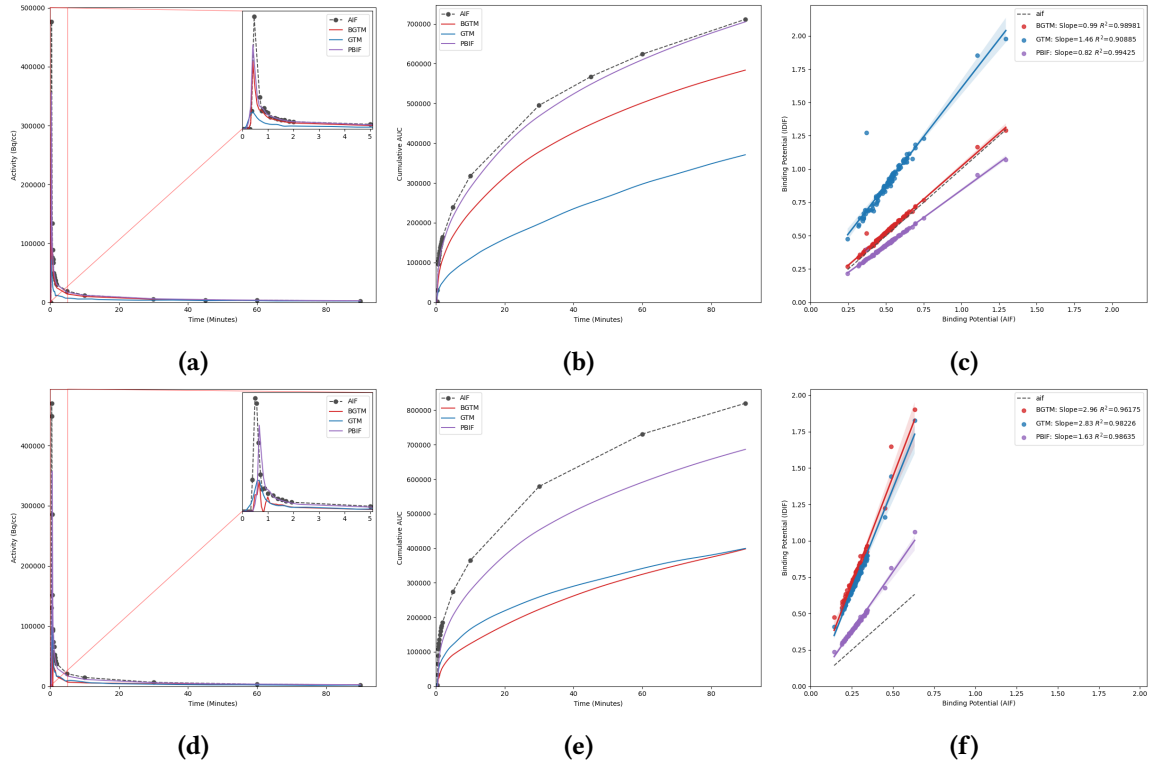
Paired t-tests were used to compare BGTM with GTM and PBIF on quantification metrics (Table 3.1). Relative to GTM, BGTM showed significantly lower  $MR_{glu}$  MAE ( $t = -5.934, p = 1.754 \times 10^{-7}$ ),  $MR_{glu}$  MAPE ( $t = -5.459, p = 1.042 \times 10^{-6}$ ),  $R^2$  absolute percentage error (APE) ( $t = -2.762, p = 7.677 \times 10^{-3}$ ), and slope APE ( $t = -4.030, p = 1.644 \times 10^{-4}$ ). Relative to PBIF, BGTM also yielded significantly lower  $MR_{glu}$  MAE ( $t = -2.974, p = 4.275 \times 10^{-3}$ ),  $MR_{glu}$  MAPE ( $t = -2.415, p = 1.890 \times 10^{-2}$ ), and slope APE ( $t = -3.036, p = 3.583 \times 10^{-3}$ ), while the difference in  $R^2$  APE was not significant ( $t = 1.680, p = 9.832 \times 10^{-2}$ ).

### 3.4.2 $[^{11}C]$ Yohimbine Dataset

The success of the  $[^{18}F]$ FDG dataset was not able to be replicated for the  $[^{11}C]$ Yohimbine dataset. In Figure 3.9 the best- and worst-performing subjects are compared. However, the 5% MAPE  $V_T$  in the good performing subject cannot reliably be attributed to the well performance of the algorithm rather it's mostly a lucky exception.

Overall the average cAUC MAE of the dataset was 55,499 for BGTM and 96,362 and 37,613 for GTM and PBIF, respectively (Figure 3.10b). In quantification using Logan plot method, the average  $V_T$  MAPE was 75.9%, 166.5%, and 29.5% for BGTM, GTM, and PBIF, respectively (Figure 3.10a).

As apparent by the Paired t-tests in Table 3.2, BGTM was able to perform better than GTM but performs significantly worse than the PBIF method.

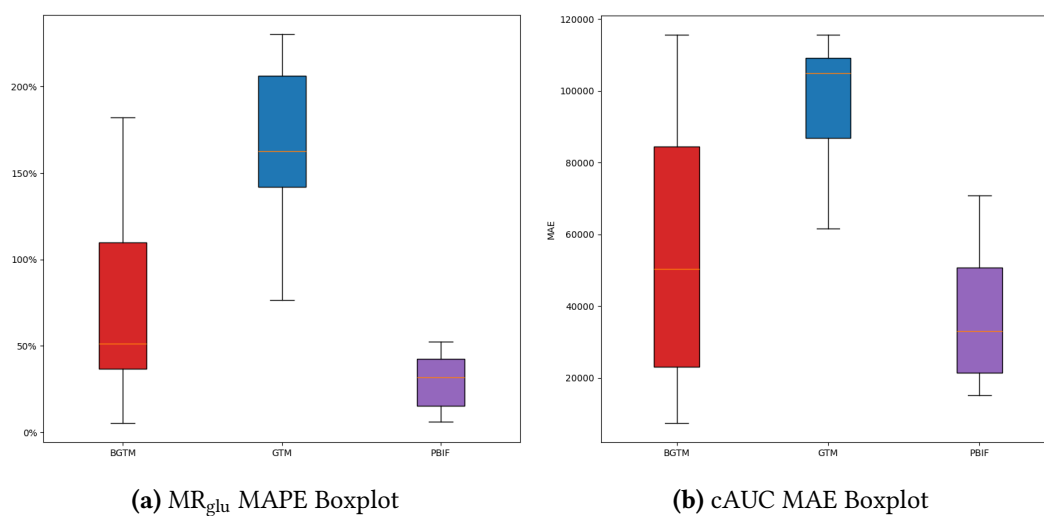


**Figure 3.9:** Comparison of the IFs (a,d), cumulative AUC curves (b,e), and  $V_T$  regression lines (c,f) for one of the best-(top row) and worst-performing(bottom row) subjects in the  $[^{11}\text{C}]\text{Yohimbine}$  dataset

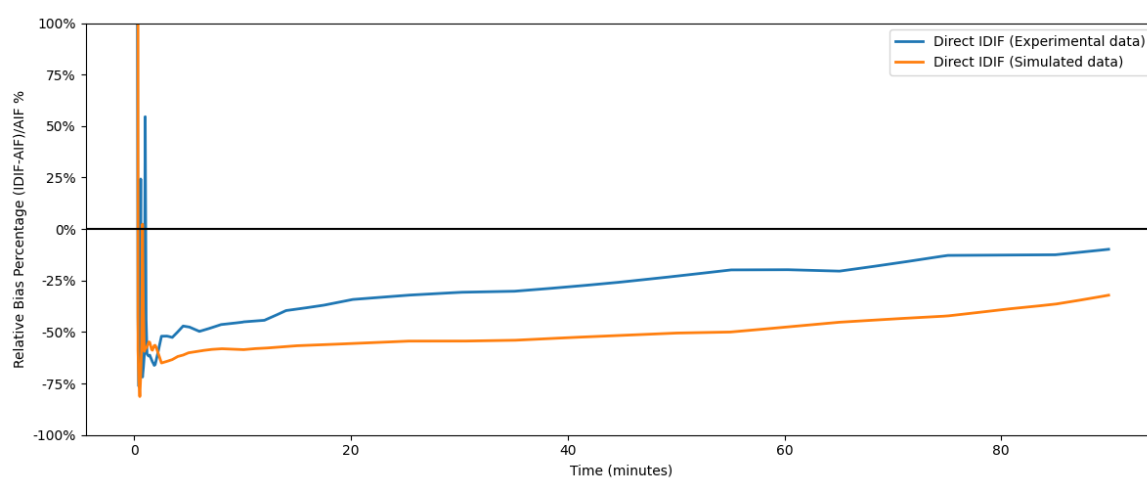
### 3.4.3 Simulated Dataset

As a complementary quality control step of the simulations, the performance of Direct IDIF (meaning with no PVC) of the experimental dataset was compared to the simulated dataset. The average cAUC MAE of the simulated dataset was significantly higher than the experimental dataset (85,138 vs 34,929). And the AUC APE of 46% for simulated compared to 22% for experimental. By comparing the temporal error of the Direct IDIF curves of both dataset (example subject in Figure 3.11), it was shown that the Direct IDIF of the simulated data underestimates the activity level compared to the experimental data.

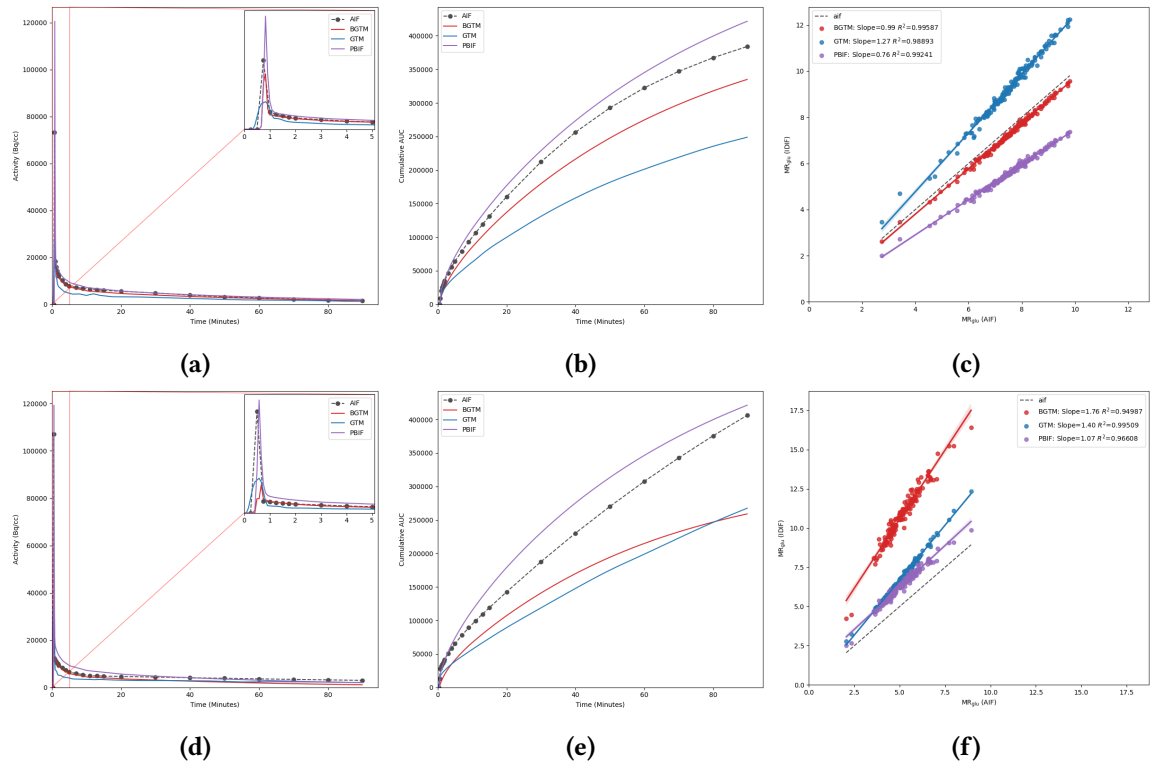
This underestimation and larger bias in the Direct IDIF was translated in to large bias in the other IDIF methods as well. Figure 3.12 compares one of the best and worst-performing subjects. BGTM was only able to show slight improvement in cAUC MAE compared to GTM with 53,997 vs 57,991. However, it performed significantly worse than PBIF with an average bias of 23,072(Figure 3.13b. In absolute quantification, BGTM underperformed compared to both methods with an average  $\text{MR}_{\text{glu}}$  MAPE of 48% compared to 35.6% and 18.5% for GTM and PBIF respectively. In Table 3.3, the performance in all metrics are detailed and no significant improvement was noticed in the paired t-tests.



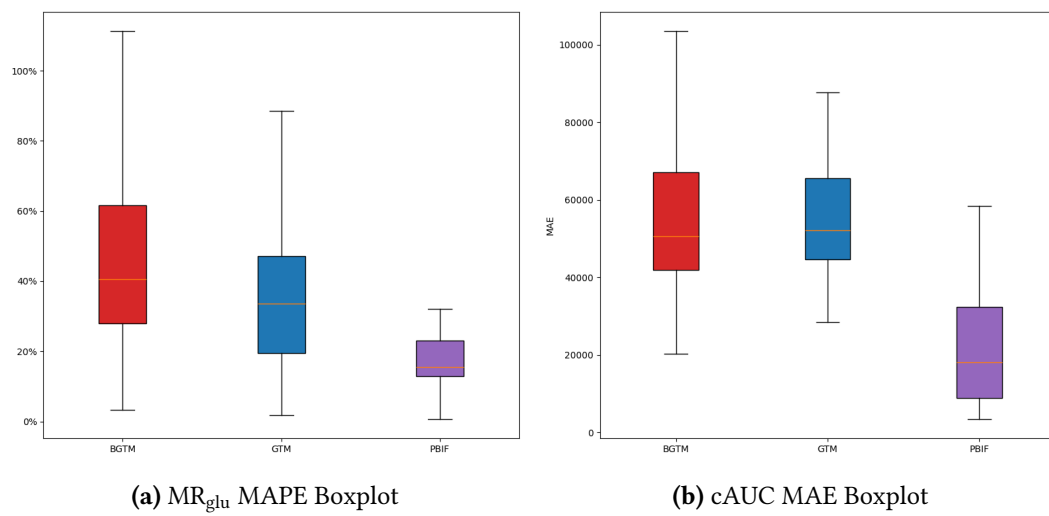
**Figure 3.10:** Boxplot of curve and quantification errors of the  $[^{11}\text{C}]$ Yohimbine dataset



**Figure 3.11:** Comparison of Relative Temporal Bias Percentage of the experimental and simulated data Direct IDIF of a subject



**Figure 3.12:** Comparison of the IFs (a,d), cumulative AUC curves (b,e), and  $MR_{glu}$  regression lines (c,f) for one of the best-(top row) and worst-performing(bottom row) subjects in the simulated dataset



**Figure 3.13:** Boxplot of curve and quantification errors of the simulated dataset

Metric	BGTM		GTM		PBIF		BGTM vs GTM		BGTM vs PBIF	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$t$	$p$	$t$	$p$
IF cAUC MAE	13,024	10,410	15,709	7,884	14,630	10,268	-2.388	$2.023 \times 10^{-2}$ *	-1.278	$2.065 \times 10^{-1}$
IF cAUC RMSE	20,293	16,078	23,380	11,979	23,334	16,803	-1.907	$6.151 \times 10^{-2}$ †	-1.439	$1.555 \times 10^{-1}$
IF AUC APE (%)	11.93	10.76	11.57	11.50	14.34	11.08	0.297	$7.674 \times 10^{-1}$	-1.862	$6.767 \times 10^{-2}$ †
MR <sub>glu</sub> MAPE (%)	13.19	10.57	24.36	13.68	17.17	12.41	-5.459	$1.042 \times 10^{-6}$ ***	-2.415	$1.890 \times 10^{-2}$ *
MR <sub>glu</sub> MAE	1.04	0.92	1.98	1.33	1.40	1.10	-5.934	$1.754 \times 10^{-7}$ ***	-2.974	$4.275 \times 10^{-3}$ **
$R^2$ APE (%)	1.76	4.07	4.13	8.90	1.50	3.53	-2.762	$7.677 \times 10^{-3}$ **	1.680	$9.832 \times 10^{-2}$
Slope APE (%)	11.28	9.37	17.05	10.58	15.76	11.41	-4.030	$1.644 \times 10^{-4}$ ***	-3.036	$3.583 \times 10^{-3}$ **

**Table 3.1:** Comprehensive curve and quantification metrics for the [ $^{18}\text{F}$ ]FDG dataset. Values are means ( $\mu$ ) and standard deviations ( $\sigma$ ) across subjects. Paired two-sided  $t$ -tests compare GTM or PBIF against BGTM for each metric. Significance codes: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , †  $p < 0.10$  (trend).

Metric	BGTM			GTM			PBIF			BGTM vs GTM			BGTM vs PBIF		
	$\mu$	$\sigma$		$\mu$	$\sigma$		$\mu$	$\sigma$		$t$	$p$		$t$	$p$	
IF cAUC MAE	55,499	40,271		96,362	20,343		37,613	20,456		-3.404	$1.443 \times 10^{-2*}$		0.959	$3.748 \times 10^{-1}$	
IF cAUC RMSE	83,282	60,130		147,032	28,318		52,313	27,845		-3.789	$9.083 \times 10^{-3**}$		1.196	$2.768 \times 10^{-1}$	
IF AUC APE (%)	26.14	19.02		52.75	3.53		17.74	14.78		-3.644	$1.078 \times 10^{-2*}$		0.872	$4.168 \times 10^{-1}$	
$V_T$ MAPE (%)	75.91	62.09		166.48	52.91		29.47	17.73		-3.628	$1.100 \times 10^{-2*}$		2.588	$4.133 \times 10^{-2}$	
$V_T$ MAE	0.324	0.211		0.812	0.360		0.132	0.072		-3.477	$1.318 \times 10^{-2*}$		3.042	$2.275 \times 10^{-2}$	
$R^2$ APE (%)	2.04	1.83		2.27	3.09		2.06	2.40		-0.148	$8.870 \times 10^{-1}$		-0.040	$9.692 \times 10^{-1}$	
Slope APE (%)	81.18	68.33		161.80	73.59		35.06	24.22		-3.167	$1.939 \times 10^{-2*}$		2.409	$5.263 \times 10^{-2}$	

**Table 3.2:** Comprehensive curve and quantification metrics for the  $[^{11}\text{C}]\text{Yohimbine}$  dataset. Values are means ( $\mu$ ) and standard deviations ( $\sigma$ ) across subjects. Paired two-sided  $t$ -tests compare GTM or PBIF against BGTM for each metric. Significance codes: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ,  $^\dagger p < 0.10$  (trend).

Metric	BGTM		GTM		PBIF		BGTM vs GTM		BGTM vs PBIF	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$t$	$p$	$t$	$p$
IF cAUC MAE	53,997	21,653	57,991	20,621	23,072	19,136	-1.353	$1.849 \times 10^{-1}$	9.013	$1.556 \times 10^{-10}$
IF cAUC RMSE	71,835	28,603	79,860	29,010	30,451	25,946	-2.105	$4.279 \times 10^{-2*}$	8.805	$2.732 \times 10^{-10}$
IF AUC APE (%)	33.35	9.20	36.13	9.46	12.94	10.59	-1.612	$1.162 \times 10^{-1}$	8.317	$1.043 \times 10^{-9}$
MR <sub>glu</sub> MAPE (%)	48.14	27.09	35.60	22.39	18.56	11.58	2.678	$1.133 \times 10^{-2}$	6.578	$1.545 \times 10^{-7}$
MR <sub>glu</sub> MAE	4.62	2.13	3.49	2.21	1.82	1.08	2.463	$1.901 \times 10^{-2}$	7.598	$7.919 \times 10^{-9}$
$R^2$ APE (%)	0.35	0.88	0.65	0.88	0.32	0.59	-1.636	$1.111 \times 10^{-1}$	0.609	$5.466 \times 10^{-1}$
Slope APE (%)	45.58	23.70	39.97	23.74	16.86	11.37	1.297	$2.035 \times 10^{-1}$	6.812	$7.750 \times 10^{-8}$

**Table 3.3:** Comprehensive curve and quantification metrics for the simulated dataset. Values are means ( $\mu$ ) and standard deviations ( $\sigma$ ) across subjects. Paired two-sided  $t$ -tests compare GTM or PBIF against BGTM for each metric. Significance codes: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ,  $^\dagger p < 0.10$  (trend).



# Discussion

The Bayesian Geometric Transfer Matrix (BGTm) method demonstrated significant improvements in IDIF accuracy compared to classical GTM approaches through enhanced handling of partial volume effects. This advancement was achieved by integrating TOF-MRA-guided carotid segmentation with population-based AIF prior knowledge within a Bayesian framework.

The refined segmentation algorithm, particularly through implementation of the cuboid mask, proved crucial for processing challenging data from comatose patients who frequently present brain lesions (Figure 3.2). However, occasional over-conservatism in vessel selection suggests the region-growing and thresholding parameters can be optimized to accurately capture the complete carotid structure. The lack of reference or ground truth carotid mask prevented the evaluation of the segmentation. This also disabled us from accurately evaluating the proposed IDIF algorithm, as the calculated errors are accumulated with the segmentation error as well. With labeled data, deep learning approaches such as U-Net can also be explored to improve segmentation accuracy [9].

The cAUC error reduction (Figure 3.8b) confirms Bayesian priors effectively constrain AIF estimates to physiological ranges. The strong correlation between cAUC errors and quantification errors (Figure ??) validates cAUC as a robust intermediate metric for IDIF assessment, enabling rapid validation without requiring full kinetic modeling.

Quantitative analysis revealed strong agreement between BGTm and AIF-derived  $MR_{glu}$  values, with the Bayesian approach achieving a 57% reduction in absolute error compared to conventional GTM methods (Table ??). Despite this improvement, the observed MAPE results exhibited substantial dispersion (Coefficient of Variation = 71%). This inconsistency is particularly evident in outlier cases where BGTm underperformed (Figure 3.7). Although, further investigation is required to fully characterize the underlying causes, one plausible factor could be TOF-MRA/PET misregistration resulting in inaccurate TAC extraction.

Our analysis was limited to a comatose patient cohort from a single imaging center, which may limit generalizability to populations with normal cerebral blood flow. Future validation should incorporate multi-center studies involving both healthy subjects and different radiotracers to establish broader applicability.

Additionally, the current PCA implementation of randomly selecting 10 subjects from the population lacks practical viability for standardized implementation. Future iterations should only be limited to a predefined population.

# Conclusion

This study demonstrates that combining TOF-MRA-guided carotid segmentation with partial volume correction in a bayesian framework improves image derived input function estimation. While the method showed significant improvement in IDIF estimation accuracy, variability in accuracy underscores the challenges posed in artery segmentation as well as partial volume correction. Future validation must address limitations in the current PCA framework by establishing predefined reference cohorts and expanding to multi-center studies with diverse patient populations. By refining anatomical alignment protocols and ensuring consistent prior knowledge integration, this approach can enhance reliability for clinical translation, ultimately supporting safer, patient-friendly quantitative PET imaging.

# References

- [1] John W Keyes. “SUV: standard uptake or silly useless value?” In: *Journal of Nuclear Medicine* 36.10 (1995), pp. 1836–1839.
- [2] Jean Logan et al. “Graphical analysis of reversible radioligand binding from time—activity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects”. In: *Journal of Cerebral Blood Flow & Metabolism* 10.5 (1990), pp. 740–747.
- [3] Clifford S Patlak, Ronald G Blasberg, and Joseph D Fenstermacher. “Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data”. In: *Journal of Cerebral Blood Flow & Metabolism* 3.1 (1983), pp. 1–7.
- [4] Peter Young et al. “Image-derived input functions from dynamic 15O–water PET scans using penalised reconstruction”. In: *EJNMMI physics* 10.1 (2023), p. 15.
- [5] Zacharie Irace et al. “Bayesian partial volume correction for image derived input function”. In: *Journal of Cerebral Blood Flow and Metabolism*. Vol. 41. 1\_ SUPPL. SAGE PUBLICATIONS INC. 2021, pp. 229–229.
- [6] Alexander Hammers et al. “Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe”. en. In: *Hum. Brain Mapp.* 19.4 (Aug. 2003), pp. 224–247.
- [7] Olivier G Rousset, Yilong Ma, and Alan C Evans. “Correction for partial volume effects in PET: principle and validation”. In: *Journal of nuclear medicine* 39.5 (1998), pp. 904–911.
- [8] Vesa Oikonen et al. *TPCCLIB*. Version 0.6.20. Retrieved on 2nd February 2025. Turku PET Centre, University of Turku, 2018. URL: <https://gitlab.utu.fi/vesoik/tpcclib>.
- [9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18. Springer. 2015, pp. 234–241.