博 士 学 位 論 文

氏名（本籍）	Paulus Kapundja Shigwedha
学位の種類	博　士（工学）
学位記番号	甲第１７０号
学位授与年月日	令和 3 年 4 月 26 日
学位授与の要件	学位規則第４条第　1　項
学位論文題目	Methods to reduce the bias and enhance the contrast of receptor density in positron emission tomography imaging

論文審査委員	主査	福岡　豊
	副査	於保　英作
	〃	斎藤　秀俊
	〃	木村　裕一
	〃	
	〃	
	〃	

工学院大学大学院

Dedications:

I dedicate this thesis to my late father, Tomas Shikongeni SHIGWEDHA, and to my two younger brothers, Leevi Panduleni SHIGWEDHA and Esra Gandja-Ayihe SHIGWEDHA.

Acknowledgements

Firstly, I would like to show my gratitude towards the Japanese government, which, through its Ministry of Education, Culture, Sports, Science and Technology (MEXT), offered me a scholarship to come further my studies in this wonderful nation of the rising sun, Japan.

I am also immensely grateful to my supervisor [Yutaka FUKUOKA, Professor, Faculty of Engineering, Department of Electrical and Electronic Engineering, Kogakuin University of Technology and Engineering (KUTE), Tokyo, Japan] for his vast guidance and motivation during my studies. He has been continuously supportive, and that made it possible for me to achieve the goal of my research. I appreciate the input by our external supervisor [Yuichi KIMURA, Professor, Department of Computational Systems Biology, Faculty of Biology-Oriented Science and Technology, Kindai University, Wakayama, Japan], he provided me with his well established MATLAB® library for the LGA execution, and his valuable assistance throughout the course of this research is fully acknowledged. In addition, I appreciate the members of the Thesis Committee, co-supervisor [Eisaku OHO, Professor, Faculty of Engineering, Department of Electrical and Electronic Engineering, KUTE, Tokyo, Japan] and [Hidetoshi SAITO, Professor, Faculty of Engineering, Department of Electrical and Electronic Engineering, KUTE, Tokyo, Japan], for their valuable time as well comments that greatly improved the content of this thesis. I am also thankful to my fellow labmates, in the Laboratory for Biomedical Informatics (LBMI), KUTE, academic years, 2015–2020, for their help and discussions in the laboratory. Further, I would also like to appreciate the support provided by KUTE's Student Support Division, and the entire University community throughout my studies.

My family has always been there for me during my studies. I would like to thank my mother and my late father, and my entire family and friends back home for their unlimited love, guidance and support in all forms. I would also like to thank the Namibian personnel in Japan for their productive moments we have had during my studies. I am as well thankful to my friends in Japan, especially Mr. Hiroki HARA for the time he has spent helping me in various ways, especially with social situations during the course of my stay in Japan.

I reserved special thanks for my best friend Hilda Theofilus. She has been encouraging and always ready to listen and attend to me whenever I needed her.

Contents

Chapter 1

Introduction

Positron emission tomography (PET) is a physiological imaging technique in nuclear medicine. PET is based on the fundamentals of nuclear dacay, specifically beta-plus (β^+) decay. Beta-plus decay is a nuclear disintegration process that occurs in a nucleus with more number of protons than neutrons. In a β^+ decay, a proton is transformed into a neutron, and then a positron and a neutrino are simultaneously released [1]. PET is used to evaluate physiological processes of biological systems, which is applicable in diagnosing, staging and treatment of various diseases, such as various heart and brain diseases and cancers. In a PET study, a β^+ radioactive substance, referred to as a radiotracer, is administered *in vivo*. When the administered radiotracer undergoes β^+ decay *in vivo*, emitted positrons annihilates with the electrons in tissues. The positron-electron annihilation gives off two γ-rays, which are approximately in opposite directions. The γ-rays are then detected to construct the concentration of the radiotracer in tissues over time; these data are referred to as tissue time-activity curves (tTACs). Various methods such as the Logan graphical analysis (LGA) can then be applied to tTACs for quantitative analysis of the radiotracer kinetics.

LGA can be thought of as a linearization technique for PET data. LGA transforms PET tTACs into linear relationships [2, 3]. Physiologically interpretable quantities, such as distribution volume (DV), distribution volume ratio (DVR) and the non-displaceable binding potential (BP_{ND}) can be obtained from the slope of the LGA linear relationship. LGA is computationally efficient, in that it is easy to implement and fast to compute, owing to which it has been well-accepted amongst researchers.

Problem statement and motivation

Traditionally, the LGA slope is estimated by the common ordinary least-squares (OLS) regression. However, in the presence of noise in PET data, it has been shown that the OLS-based estimates of the LGA slope are underestimated [4, 5, 6, 7]. In addition, this negative bias has been observed to increase with the magnitude of both noise and the LGA slope [4, 5, 6, 7]. The underestimation is due to that both the LGA variables are contaminated with a correlated noise, but OLS only accounts for the noise in the response variable. Specifically, the noise in the LGA variables is attributed to the noisy term, $C(t)$ (see Eqs. (2.21, 2.40)), in the denominators of both of the two LGA variables. Accordingly, the noise due to $C(t)$ propagates correlatively in the LGA variables, presenting a correlated errors-in-variables (EIV) problem. This explicitly makes OLS unsuitable for estimating the LGA slope.

Various methods have been proposed to reduce the influence of the noise in PET data on the estimated slope. These methods range from tTAC smoothing methods, such as the generalized linear least-squares (GLLS) [8] technique, to methods based on a rearrangement of the original LGA equation, such as the multilinear reference tissue models (MRTM 1 & 2) [9, 10]. Other methods including orthogonal distance regression (ODR) [11] simply employ a different line estimation technique. These methods, however, have been found to either only slightly reduce the bias or reduce the bias whilst causing variations in the estimates [5, 6]. This necessitates the need for further studies. This study aims to establish a method to reduce the bias in the estimates of the LGA slope, and improve the contrast of the resulting parametric images. In [11], it has been demonstrated that the regression method used to estimate the Logan slope influences the resulting bias. On this basis, this study employs an alternative linear regression method referred to as least-squares cubic (LSC) to estimate the LGA slope. LSC accounts for noise in both variables, by minimizing the squared residuals in both the predictor and response variables [12, 13, 14]. In addition, LSC also incorporates the correlation of the noise in the variables. To further improve the LSC estimates, LSC is for the first time combined with tTACs denoising techniques, principal component analysis (PCA) and correlated component analysis (CorrCA).

PCA is a dimensionality reduction technique used mostly for feature extraction and noise filtering. PCA transforms variables into new sets of variables that are linear functions of the original

variables [15]. The new sets of variables are uncorrelated and are defined by sets of orthogonal basis vectors and principal components (PCs) that optimally describe the variance in the data. The PCs and the corresponding variance are obtained by solving for the eigenvectors and eigenvalues of the covariance matrix of the data. The eigenvectors are the orthogonal basis vectors, which define the PCs, and the eigenvalues are the corresponding variances.

PCA is a common tool in PET parametric imaging; it has been previously used to denoise tTACs in [16], and it showed minimal effects on the variance. Therefore, this effect is employed in this study. PCA has also been used for clustering in [17]. Further, PCA has also been used for noise filtering in PET studies with irreversible radiotracers [18]. Here, LSC and PCA are used together for the first time, and the effect of this fusion on the image quality was assessed not only in the usual manner of visual inspection of the images but also numerically in terms of the contrast between grey matter and white matter. In other words, this study utilizes the combined effects of LSC's minimal bias and PCA's precision.

Like PCA, CorrCA is also a feature extraction method. CorrCA operates by identifying components that are maximally correlated between repetitions in multivariate data [19]. Specifically, CorrCA maximizes the ratio of between-repetition to within-repetition covariance [19]. In the context of our application to PET data, our repetitions are along the slices dimension. CorrCA will therefore maximize the ratio of between-slices to within-slices covariance. This ratio is referred to as inter-subject correlation (ISC). Specifically in our context, it translates to inter-slice correlation.

Maximizing the between-slices to within-slices covariances maximizes the mean-over-variance across slices, which has been asserted to define a signal-to-noise ratio measure [19]. CorrCA capitalize on the ability to simultaneously operate "within individual slices" and "across all slices", without arranging the slices in one plane. That means operating directly on the 3D-array, unlike PCA which requires the slices to be arranged side-by-side into a 2D-array. This could be advantageous over PCA, which means CorrCA can possibly provide a better noise filter method in comparison to PCA. This study therefore also seek to introduce CorrCA to PET parametric imaging and assess the performance of CorrCA in comparison to PCA. Hereafter, the fusions of LSC with PCA and CorrCA are respectively denoted as LSC–PCA and

LSC–CorrCA. Similarly, using OLS with PCA and CorrCA are herein referred to as OLS–PCA and OLS–CorrCA, respectively.

This study was conducted on both simulation and real data. Two sets of simulation data were used, one mimicking ^{11}C-carfentanil (CFN), and the other mimicking ^{11}C Pittsburgh compound B (^{11}C-PiB). The real data are for ^{11}C-PiB PET brain images. Both CFN and ^{11}C-PiB have specific binding sites in the brain. CFN binds to μ-opioid receptors, and ^{11}C-PiB binds to beta amyloid (Aβ) plaques.

Thesis outline

The next chapter (Chapter 2) goes into more detail about PET. In the subsections, the compartmental model, derivation of the LGA from the compartment model, as well as the interpretation of the parameters estimated from the LGA are discussed. OLS, ODR and MRTM2 are also discussed. Chapter 3 gives the derivation and analysis of LSC. The analysis of LSC is also given specifically in terms of the LGA variables. The results of the study of LSC-based LGA in comparison to OLS-based LGA and MRTM2 are presented in this chapter (Chapter 3). These are the results reported in the published paper: [20]. In Chapter 4, the mathematical details of both PCA and CorrCA are presented. The results of the study of LSC–CorrCA, LSC–PCA, LSC and OLS are given in this chapter (Chapter 4). Chapter 5 summarizes this thesis and discusses future directions.

Chapter 2

Positron emission tomography, compartmental model and the Logan graphical analysis

2.1 Positron emission tomography

PET is referred to as a functional imaging method. This is to differentiate it from other imaging techniques such as magnetic resonance imaging (MRI) which are defined as structural imaging. A functional imaging helps image the functioning status of cells *in vivo*, whereas structural imaging is used to examine the structural shapes of tissues and organs. PET is used to visualize physiological functioning of tissue cells by monitoring the interaction of the radiotracer with the system's biochemical processes and activities such as glucose metabolism, oxygen consumption and blood flow. PET is advantageous over other imaging techniques in that it can detect diseases from their early onsets. In a similar way PET can also detect immediate response of a patient to treatments.

A radiotracer is made to have an affinity to a particular region of interest (ROI). Therefore, eventually after administration, the radiotracer will accumulate in the ROI. As the radionuclide undergoes β^+ decay in the ROI, the time-course of radioactivity in tissues is recorded to be used to study the physiological activities of interest in the ROI. Figure 2.1 shows a simple scheme of PET imaging. Radiotracers are made of two components; a radionuclide and a biological molecule. This forms a biological radioactive compound. The biological molecule describes the

affinity aspects of the radiotracer.

Equation (2.1) [1] represents a β^+ decay process. The parent nucleus ${}^{A}_{Z}X_N$ has more number of protons than neutrons, i.e., $Z > (A - Z = N)$. A proton in the proton-rich parent nucleus is transformed into a neutron, accompanied by a simultaneous emission of a positron (e^+) and a neutrino (v). The resulting daughter nucleus, ${}^{A}_{Z-1}Y_{N+1}$, has "one-less number of protons" and "one-more number of neutrons" than the parent nucleus. The mass number (A) of the daughter nuclide remains the same as that of the parent nuclide, i.e., the daughter nuclide is an isobar of the parent nuclide —different chemical elements with same number of nucleons (A) and different number of protons. The daughter nuclide can be a stable isotope or a radioactive isotope which is closer (than the parent nuclide) to the stability region on the chart of nuclides.

$$ {}^{A}_{Z}X_N \rightarrow {}^{A}_{Z-1}Y_{N+1} + e^+ + v. \tag{2.1} $$

The β^+ decay equation for ^{11}C is represented in Eq. (2.2). ^{11}C is the radionuclide label for ^{11}C-PiB radiotracer, which is utilized for clinical ^{11}C-PiB PET images analyzed in this study.

$$ {}^{11}_{6}\mathrm{C}_5 \rightarrow {}^{11}_{5}\mathrm{B}_6 + e^+ + v + 0.96\,\mathrm{MeV}. \tag{2.2} $$

Equation (2.2) means that the unstable ^{11}C radionuclide undergoes β^+ decay to form the stable isotope ^{11}B, and a total of 0.96 MeV is released —which is mostly carried away as kinetic energy of the released particles. ^{11}C has a half-life of 20.364 minutes.

Beta-plus decay is in contrast to another form of beta decay, referred to as beta-minus (β^-) decay, in which a neutron is transformed into a proton, accompanied by a simultaneous emission of an electron (e^-) and an anti-neutrino ($\bar{v}$). Beta-minus decay can be represented by Eq. (2.3) [1] below:

$$ {}^{A}_{Z}X_N \rightarrow {}^{A}_{Z+1}Y_{N-1} + e^- + \bar{v}. \tag{2.3} $$

With reference to the type of beta particles released in β^+ and β^- decay, they can be referred to as **positron emission** and **electron emission**, respectively. The two beta decay processes can be further explained interms of *"up quarks"* and *"down quarks"* —which are the fundamental particles that forms the nucleons.

PET radiotracers are commonly made of low elements short–lived isotopes such as ^{11}C, ^{18}F, ^{15}O and ^{13}N. Due to its effectiveness in cancer detection and staging, [^{18}F]-fluorodeoxyglucose

(FDG) is the widely used radiotracer [1, 21]. Depending on factors such as the ROI, the radiotracer being used and the demographics of the subject, a radiotracer can be administered by either injection, swallowing or inhalation. The main purpose of a radiotracer is to trace the physiological process of interest in the ROI. Therefore, it is administered in such an amount that it does not perturb the physiological processes it is being used trace [22, 21].

Being a glucose analog, [^{18}F]-FDG, is transported into cells by the same mechanism as glucose. Intracellular, glucose is phosphorydated into glycolysis. Unlike glucose, [^{18}F]-FDG cannot enter glycolysis. This is because [^{18}F]-FDG does not have the 2-hydroxyl group (-OH) required for glycolysis [23, 24]. [^{18}F]-FDG therefore gets trapped in cells in the form of [^{18}F]-FDG-6-Phosphate [23]. [^{18}F]-FDG-6-Phosphate exits the cells via β^+ radioactive decay. Imaging of the distribution of the [^{18}F]-FDG by PET can therefore be used to trace the distribution of glucose uptake and phosphorylation.

The clinical data used in this study are of ^{11}C-PiB radiotracer. ^{11}C-PiB binds to Aβ plaques in the brain [25, 26], which are associated with Alzheimer's disease (AD) and and mild cognitive impairment (MCI). Quantification of ^{11}C-PiB binding therefore allows for visualization and quantification of *in vivo* Aβ deposition.

2.1.1 Positron emission tomography detection system

PET data acquisition is based on measuring the total energy lost by annihilation photons as they interact with the detector equipment. Annihilation photon detectors form a ring surrounding the subject to undergo PET imaging. The detection ring depicted in Fig. 2.1 is made of scintillation materials mounted onto photomultipliers tubes (PMTs) [27]. Photon detection is then based on their interaction with the scintillator material. The scintillator material is a specially-designed crystal which emits visible light upon interaction with annihilation photons [1]. This process is referred to as scintillation. In short, in the best case scenario, when there is no photoelectric or scatter events, annihilation photons reach and directly excite the scintillator valence electrons. The excited electrons get into the conduction band, and return (de-excite) to the ground state by releasing some energy via photon emission [1]. Scintillation materials are modified by adding impurities to pure crystals such that the scintillation photons are in the visible electromagnetic spectrum, i.e., luminescence [1]. The PMTs then detect the scintillation

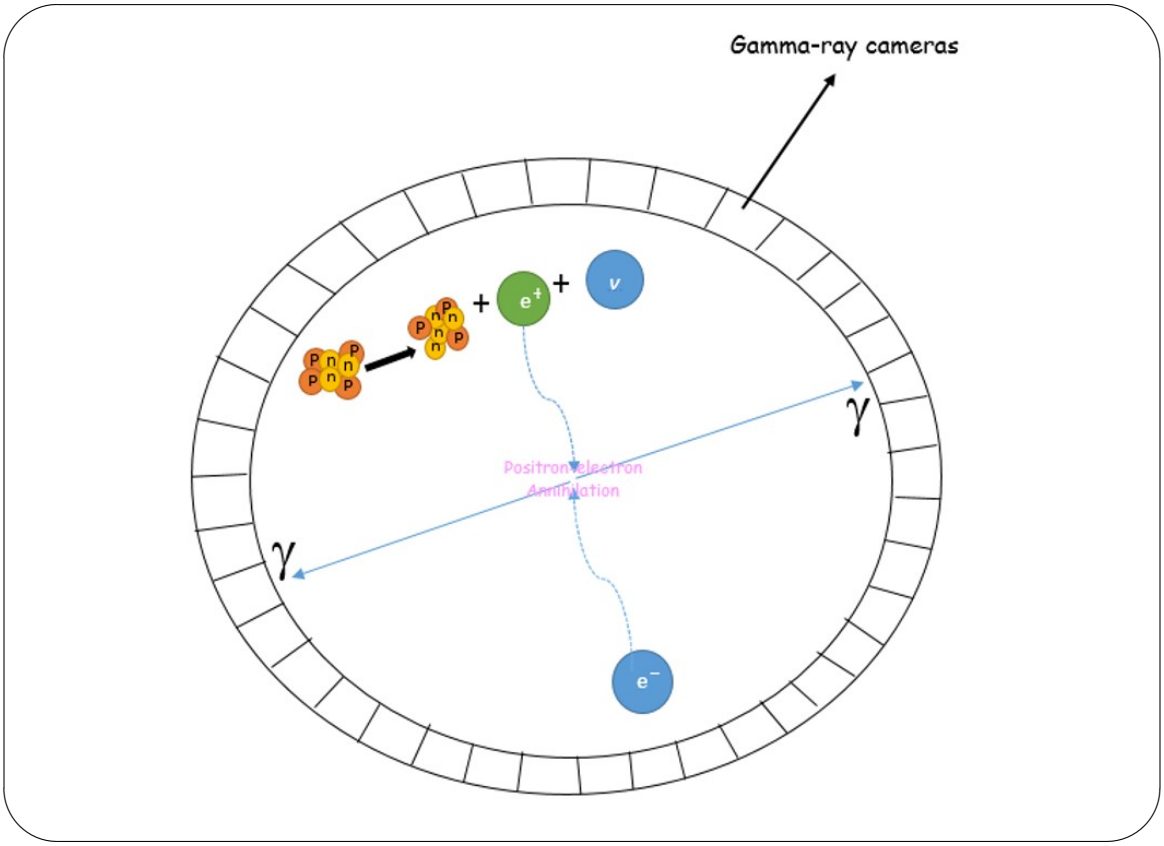


Figure 2.1: *A simple PET schematic diagram.*

photons by ionization mechanism of photoelectric effect. Two photons detected within a preset coincidence time window (usually a few nanoseconds (1 – 12)) are recorded and used to determine a line of response (the line connecting the centers of the two detectors.) [1, 27, 28]. The number of scintillation photons produced in the scintillator medium, and the subsequent electric charge produced in the PMT is proportional to the energy deposited by the annihilation photon in the scintillator medium. These data can then be used to evaluate radioactivity concentration in tissues.

The four main properties of scintillator materials are, stopping power for annihilation photon (511 keV), signal decay time, light output, and the intrinsic energy resolution [1]. The stopping power is defined by the inverse of the mean distance a photon can travel before depositing its energy in the detector material [1, 28]. This distance is termed attenuation length [1], and it is dependent on the density and effective atomic number of the scintillator material. The high the number of photons depositing their energy in the scintillator material the better, i.e., a scintillator material with a short attenuation length is preferable. The decay time of a scintillator

describes the rate at which the valence electrons excite and de-excite following annihilation photon absorption, i.e., the rate of photoluminescence. During this period the detector is referred to as being dead since it cannot process all other coincidence events [28], and are therefore lost —this time period is referred to as dead time. Decay time is usually in the range of hundreds of nanoseconds (e.g., 230 nanoseconds for Sodium Iodide doped with Thallium (NaI(Tl)) scintillator crystal [28]). The manner in which the decay time is associated with the efficiency of the scintillator material is that, short decay time will allow discrete cascades of electrons to be processed at high count rates, and minimize the number of random coincidences [1]. Light output refers to the number of scintillation photons produced by an incident annihilation photon. High light output allows for a good spatial resolution and good energy resolution [1, 28]. High energy resolution allows for improved discrimination of scatter and true events [1, 28]. The energy resolution of a scintillator is also dependent on the intrinsic energy resolution of the scintillator material, which is dependent on the inhomogeneities of the crystal and its associated properties, such as light-output uniformity [1].

There are many types of scintillator materials used in PET detection systems. One such is the NaI(Tl) crystal mentioned earlier. The NaI(Tl) scintillator has a high light output, providing a good energy and spatial resolution [1, 28]. However, NaI(Tl) scintillator crystal has a slow decay time, which means increased detector dead time, as well as increased random coincidences [1]. Because of its low density, NaI(Tl) also has a low stopping power than the other scintillator materials [1, 28]. One other scintillator material is the Gadolinium Oxyorthosilicate doped with Cerium (GSO). GSO has a lower stopping power and energy output, but it is mostly emplyed for its improved energy resolution and uniform light output in comparison to other scintillator materials [1]. Another scintillator material is Barium Flouride (BaF_2). BaF_2 is noted for its very short decay time (600 picoseconds), an excellent feature that makes up for the effects of its low stopping power [1].

PMTs are comprised of an enclosed vacuum incorporated with a photo-cathode at the entrance window [1]. The vacuum tube of the PMT consists of dynodes, arranged at consecutively increasing potentials. An incident scintillation photon interacts with the atoms of the photocathode layer by photoelectric effect [1]. The resulting photo-electron is accelerated to the first dynode by an applied electric field [1]. This accelerated photo-electron will cause further mul-

tiple electron emissions at the dynode. This trail of electron acceleration and further multiple secondary emissions continues at consecutive dynodes, leading to an amplified electric signal. The final component of the dynode series is an anode through which the collected electric charge leaves the vacuum tube for signal processing.

In PET data acquisition, three main kinds of possible coincidence events, —namely, true, random, and scatter coincidences— can be recorded. True coincidences are the events of interest in a PET study. These events occur when two photons coming from one annihilation event are detected by two oppositely located detectors, without encountering any major interactions with the tissues in the field of view. Random coincidences are accidental detection of two unrelated photons as coincidence photons, detected within the preset coincidence detection window. Two photons from a single annihilation event that are detected after either one or both have undergone a major scatter event results in what is termed scatter coincidences. Random and scatter events are some of the major contribution to sources of errors in PET data, in addition to factors such as positron range. Positron range refers to the distance traveled by a positron after emission from the nucleus before annihilation [1]. The aim of PET measurement is to determine the position of the nucleus in which beta-plus decay took place, therefore if the annihilation event took place at some distance from the nucleus then that will add to errors in PET data.

2.1.2 Compartmental model and the Logan graphical analysis

To better understand the LGA, one needs to understand it in terms of its mathematical construction. LGA is derived from a well known concept in kinetic modeling referred to as compartmental model. Compartmental model forms the foundation of PET kinetic analysis. Speciffically, compartmental model is used to describe the kinetics of a radiotracer *in vivo* via a set of differential equations. In compartmental model, it is assumed that there are discrete physiological regions in which the radiotracer is concentrated [21, 29]. These physiological regions are referred to as compartments. Applying differential Calculus, one obtains a set of differential equations describing the rates of change in the concentration of the radiotracer in the compartments with time. The LGA is derived from the differential equations of the compartmental model, giving a linear relationship of which the slope is a physiologically interpretable quantity.

Figure 2.2 shows schematic figures for compartmental models; Two compartments (one tissue)

model in 2.2(a); Three compartments (two tissues) model in 2.2(b); and Four compartments (three tissues) model in 2.2(c). The compartments C_P, C_F, C_N, and C_S represent the concentration of the radiotracer in arterial plasma, bound-free radiotracer, non-specifically bound radiotracer, and specifically bound radiotracer, respectively. The naming of the compartments are self-explaining. The arterial plasma compartment refers to the radiotracer in the arterial plasma, which has not yet entered the tissues. Specifically bound compartment refers to the radiotracer specifically bound to target receptors in tissues. Non-specifically bound refers to the radiotracer that is bound to receptors other than the target receptors.

The transportation of the radiotracer between neighboring compartments is described by the exchange rate constants, K_1 $[ml\,g^{-1}\,min^{-1}]$ and $k_{2\text{–}4}$ $[min^{-1}]$ [30, 21]. K_1 denotes the delivery rate of the radiotracer from the arterial plasma to tissues; k_2 is the clearance rate from the tissues back to the plasma; k_3 and k_4 are the respective association and dissociation rates of the radiotracer to and from the specific binding sites; and k_5 and k_6 are the respective association and dissociation rates to and from the non-specific binding sites. The pairs of the exchange rate constants and the difference of the radiotracer concentration of the two involved compartments are assumed to be linearly related [21]. This relationship makes it possible to form the differential equations.

The concentrations of the radiotracer in the tissue compartments at time t $[min]$ for the one-tissue compartment model in Fig. 2.2(a) are described by the differential equation below,

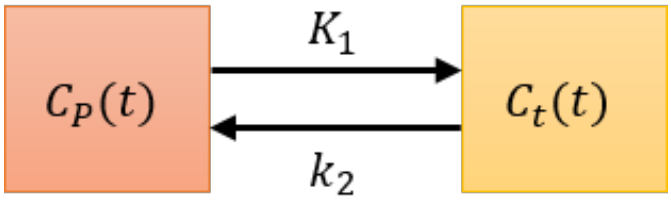
$$\frac{dC_t(t)}{dt} = K_1 C_P(t) - k_2 C_t(t). \tag{2.4}$$

Similarly, the concentrations of the radiotracer in the tissue compartments for a two- (Fig. 2.2(b)) and three- (Fig. 2.2(c)) tissues models are respectively described by,

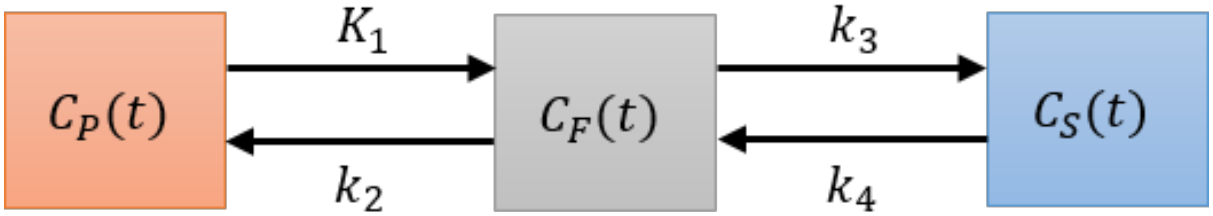
$$\begin{aligned} \frac{dC_F(t)}{dt} &= K_1 C_P(t) + k_4 C_S(t) - (k_2 + k_3) C_F(t) \\ \frac{dC_S(t)}{dt} &= k_3 C_F(t) - k_4 C_S(t), \end{aligned} \tag{2.5}$$

and,

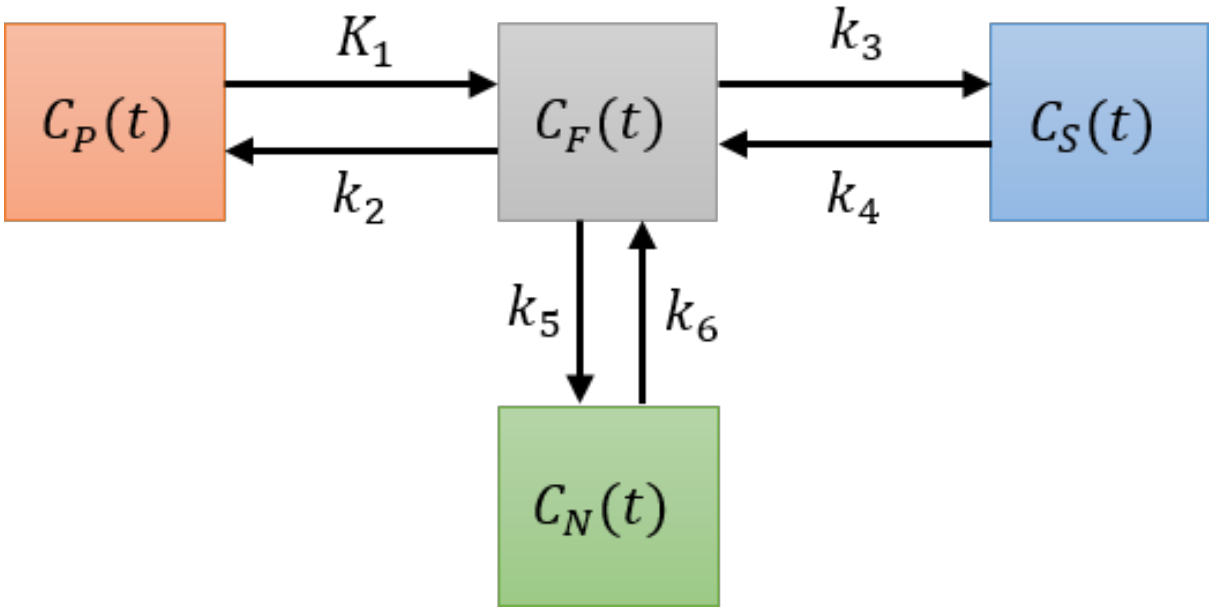
$$\begin{aligned} \frac{dC_F(t)}{dt} &= K_1 C_P(t) + k_6 C_N(t) + k_4 C_S(t) - (k_2 + k_3 + k_5) C_F(t) \\ \frac{dC_N(t)}{dt} &= k_5 C_F(t) - k_6 C_N(t) \\ \frac{dC_S(t)}{dt} &= k_3 C_F(t) - k_4 C_S(t). \end{aligned} \tag{2.6}$$



(a) *Two compartments (one-tissue) model.*



(b) *Three compartments (two-tissues) model.*



(c) *Four compartments (three-tissues) model.*

Figure 2.2: *Schematic figures of the (a) one-, (b) two- and (c) three-tissues compartmental models.*

$C_P(t)$, $C_F(t)$, $C_S(t)$ and $C_N(t)$ denote the concentrations of radioactivity in the respective compartments at time t.

PET data are not recorded separately for the compartments. The recorded PET data represents a sum of the radioactivity in all tissue compartments. Denoting the observed PET data by $C_{PET}(t)$, for Fig. 2.2(c) we can write,

$$C_{PET}(t) = C_F(t) + C_S(t) + C_N(t). \tag{2.7}$$

Non-linear least-squares method can be used to fit measured tTACs to the compartmental model to estimate the kinetic parameters (exchange rate constants) [21], using arterial plasma time-activity curve, $C_P(t)$, as input function. Estimated kinetic parameters are then used to evaluate physiological processes of the target tissues [21]. The non-linear approach is dependent on the number of components. Graphical analysis methods such as the LGA provides a great advantage because they are independent of the assumed model (i.e. number of compartments) [6, 21, 2]. This makes it possible to be able to obtain parameters such as DV, DVR and BP_{ND} irrespective of the assumed compartmental model.

The procedure below demonstrate the linearization of the compartmental model equations to obtain the LGA. The LGA forms a simple linear relationship of two variables, and it is the slope of this relationship, which is a function of the kinetic parameters, that defines the physiological functioning of the target tissues.

The differential equations in (2.4), (2.5) and (2.6) can be represented in a general form as,

$$\frac{d\boldsymbol{A}(t)}{dt} = \boldsymbol{K}\boldsymbol{A}(t) + C_P(t)\,\boldsymbol{Q}. \tag{2.8}$$

$\boldsymbol{A}$ represents a column vector of the radioactivity concentrations in tissue compartments, $\boldsymbol{K}$ represents a matrix made of the kinetic parameters, k_{2-6}, and $\boldsymbol{Q}$ denotes K_1. Writing these

expressions, $\boldsymbol{A}$, $\boldsymbol{K}$ and, $\boldsymbol{Q}$, out; for the one-tissue model we get,

$$\begin{aligned} \boldsymbol{A} &= C_t(t) \\ \boldsymbol{K} &= -k_2 \\ \boldsymbol{Q} &= K_1. \end{aligned} \tag{2.9}$$

The two-tissues model takes the form,

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} C_F(t) & C_S(t) \end{bmatrix}^T \\ \boldsymbol{K} &= \begin{bmatrix} -(k_2+k_3) & k_4 \\ k_3 & -k_4 \end{bmatrix} \\ \boldsymbol{Q} &= \begin{bmatrix} K_1 & 0 \end{bmatrix}^T, \end{aligned} \tag{2.10}$$

and similarly, the three-tissues model has the form,

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} C_F(t) & C_S(t) & C_N(t) \end{bmatrix}^T \\ \boldsymbol{K} &= \begin{bmatrix} -(k_2+k_3+k_5) & k_4 & k_6 \\ k_3 & -k_4 & 0 \\ k_5 & 0 & -k_6 \end{bmatrix} \\ \boldsymbol{Q} &= \begin{bmatrix} K_1 & 0 & 0 \end{bmatrix}^T. \end{aligned} \tag{2.11}$$

The $^{.T}$ represents matrix transpose.

The total radioactivity concentration in tissues, $C_t(t)$, is given by,

$$C_t(t) = C_{PET}(t) = \begin{cases} C_t(t), & \text{for a one-tissue model} \\ C_F(t) + C_S(t), & \text{for a two-tissues model} \\ C_F(t) + C_S(t) + C_N(t) & \text{for a three-tissues model} \end{cases}, \tag{2.12}$$

such that the total radioactivity concentration in the whole ROI is,

$$C_T(t) = C_t(t) + V_P C_P(t), \tag{2.13}$$

where V_P denotes the effective plasma in tissues.

Let us define a unit vector $\boldsymbol{U}$ as,

$$\boldsymbol{U} = \begin{cases} 1 & \text{for a one-tissue model} \\ \begin{bmatrix} 1 \\ 1 \end{bmatrix} & \text{for a two-tissues model} \\ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} & \text{for a three-tissues model} \end{cases},$$

so that we can write $C_t(t)$ as

$$C_t(t) = \boldsymbol{U}^T \boldsymbol{A}(t). \tag{2.14}$$

Equation (2.8) can be rearranged into,

$$d\boldsymbol{A}(t) = \boldsymbol{K}\,\boldsymbol{A}(t)\,dt + C_P(t)\,\boldsymbol{Q}\,dt. \tag{2.15}$$

Re-arranging Eq. (2.15) and integrating gives,

$$\begin{aligned} \boldsymbol{A}(t)dt &= -\boldsymbol{K}^{-1}\boldsymbol{Q}C_P(t)dt + \boldsymbol{K}^{-1}d\boldsymbol{A}(t) \\ \int_0^t \boldsymbol{A}(u)du &= -\boldsymbol{K}^{-1}\boldsymbol{Q}\int_0^t C_P(u)du + \boldsymbol{K}^{-1}\boldsymbol{A}. \end{aligned} \tag{2.16}$$

Integration of Eq. (2.14) is,

$$\int_0^t C_t(u)\,du = \boldsymbol{U}^T \int_0^t \boldsymbol{A}(u)\,du. \tag{2.17}$$

Now, substituting Eq. (2.16) into Eq. (2.17) gives,

$$\int_0^t C_t(u)\,du = -\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q}\int_0^t C_P(u)du + \boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}. \tag{2.18}$$

Integration of Eq. (2.13) is,

$$\int_0^t C_T(u)\,du = \int_0^t C_t(u)\,du + V_P\int_0^t C_P(u)\,du. \tag{2.19}$$

Substituting Eq. (2.18) into Eq. (2.19) gives,

$$\begin{aligned}\int_0^t C_T(u)\,du &= -\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q}\int_0^t C_P(u)du + \boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A} + V_P\int_0^t C_P(u)\,du \\ &= (-\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q} + V_P)\int_0^t C_P(u)du + \boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}.\end{aligned} \tag{2.20}$$

Dividing Eq. (2.20) by $C_T(t)$ results in,

$$\begin{aligned}\frac{\int_0^t C_T(u)\,du}{C_T(t)} &= (-\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q} + V_P)\frac{\int_0^t C_P(u)du}{C_T(t)} + \frac{\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}}{C_T(t)} \\ &= \alpha\frac{\int_0^t C_P(u)du}{C_T(t)} + \beta.\end{aligned} \tag{2.21}$$

Equation (2.21) is the LGA. The LGA is such that after some time t^*, the term,

$$\frac{\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}}{C_T(t)} = \frac{\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}}{C_t(t) + V_P C_P(t)}, \tag{2.22}$$

becomes constant in time [2, 31]. In turn, Eq. (2.21) becomes a linear relationship between the terms,

$$\frac{\int_0^t C_P(u)du}{C_T(t)}, \tag{2.23}$$

and,

$$\frac{\int_0^t C_T(u)\,du}{C_T(t)}. \tag{2.24}$$

The slope of the LGA linear relationship is $\alpha = -\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q} + V_P$, and the y-intercept is $\beta = \dfrac{\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}}{C_T(t)}$. A linear regression method can then be applied to estimate these regression parameters, slope and y-intercept.

Equation (2.21) can therefore be expressed simply as,

$$Y = \alpha X + \beta, \tag{2.25}$$

which is a simple linear relationship between X and Y, with a slope of α, and y-intercept β, where,

$$\begin{aligned}X &= \frac{\int_0^t C_P(u)du}{C_T(t)} \\ Y &= \frac{\int_0^t C_T(u)\,du}{C_T(t)},\end{aligned} \tag{2.26}$$

and,

$$\begin{aligned} \alpha &= -\boldsymbol{U}^T \boldsymbol{K}^{-1} \boldsymbol{Q} + V_P \\ \beta &= \frac{\boldsymbol{U}^T \boldsymbol{K}^{-1} \boldsymbol{A}}{C_T(t)}. \end{aligned} \tag{2.27}$$

The slope of the LGA described in Eqs. (2.21–2.27) is termed DV. DV defines the capacity of tissue receptors to bind the radiotracer [5]. DV is also interpreted as the equilibrium ratio of the radioactivity concentration in ROI, $C_T(t)$ $[kBq \cdot cm^{-3}]$, to that in plasma, $C_P(t)$ $[kBq \cdot mL^{-1}]$. Equilibrium state is achieved when there is no more irreversible uptake of the radiotracer between the compartments. At this point, all radiotracer flow is due to reversible uptakes, and all tissue compartments and the arterial plasma compartment are at equilibrium with each other. In equilibrium state, the radiotracer concentration in the compartments becomes constant, meaning that the flow of the radiotracer into and out of a given compartment is the same. In other words, equilibrium state is achieved after time t^* when the y-intercept term of the LGA has become constant.

The slope and y-intercept term are functions of the kinetic parameters, and it is demonstrated for the different models as follows:

One-tissue model: α (α_1) and β (β_1) are,

$$\begin{aligned} \alpha_1 &= -\boldsymbol{U}^T \boldsymbol{K}^{-1} \boldsymbol{Q} + V_P \\ &= (-1)(-k_2)^{-1} K_1 + V_P \\ &= \frac{K_1}{k_2} + V_P, \end{aligned} \tag{2.28}$$

and,

$$\begin{aligned} \beta_1 = \frac{\boldsymbol{U}^T \boldsymbol{K}^{-1} \boldsymbol{A}}{C_T(t)} &= \frac{(1)(-k_2)^{-1} C_t(t)}{C_T(t)} \\ &= \frac{C_t(t)}{-k_2 C_T(t)} + V_P \\ &= \frac{C_t(t)}{-k_2 (C_t(t) + V_P C_P(t))} \\ &= \frac{1}{-k_2 \left(1 + V_P \frac{C_P(t)}{C_t(t)}\right)}. \end{aligned} \tag{2.29}$$

In this case, equilibrium of the LGA is reached when the term $C_P(t)/C_t(t)$ in Eq. (2.29) is constant against time. V_P is assumed to be negligible such that the slope and y-intercept term in the expressions in Eqs. (2.28) and (2.29), respectively, are functions of the kinetic parameters only. Thus, Eq. (2.28) becomes,

$$\alpha_1 = \frac{K_1}{k_2}, \tag{2.30}$$

which denotes the DV obtained from a one-tissue compartmental model. Accordingly, Eq. (2.29) becomes,

$$\beta_1 = -\frac{1}{k_2}. \tag{2.31}$$

Two-tissues model: α (α_2) and β (β_2) are,

$$\begin{aligned}
\alpha_2 &= -\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q} + V_P \\
&= -\frac{1}{k_2k_4}\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} -k_4 & -k_4 \\ -k_3 & -(k_2+k_3) \end{bmatrix}\begin{bmatrix} K_1 \\ 0 \end{bmatrix} + V_P \\
&= \frac{K_1}{k_2}\left(1+\frac{k_3}{k_4}\right) + V_P,
\end{aligned} \tag{2.32}$$

and,

$$\begin{aligned}
\beta_2 &= \frac{\boldsymbol{U}^T\boldsymbol{U}^{-1}\boldsymbol{A}}{C_T(t)} \\
&= \frac{-\frac{1}{k_2k_4}\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} -k_4 & -k_4 \\ -k_3 & -(k_2+k_3) \end{bmatrix}\begin{bmatrix} C_F(t) \\ C_S(t) \end{bmatrix}}{C_T(t)} \\
&= \frac{-1}{k_2k_4}\left((k_3+k_4)\frac{C_t(t)}{C_T(t)} + k_2\frac{C_S(t)}{C_T(t)}\right).
\end{aligned} \tag{2.33}$$

Three-tissues model: α (α_3) and β (β_3) are,

$$\begin{aligned}\alpha_3 &= -\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q} + V_P \\ &= \frac{1}{k_2k_4k_6}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} -k_4k_6 & k_4k_6 & -k_4k_6 \\ k_3k_6 & k_2k_6+k_3k_6 & k_3k_6 \\ -k_4k_5 & k_4k_5 & -k_2k_4-k_4k_5 \end{bmatrix}\begin{bmatrix} K_1 \\ 0 \\ 0 \end{bmatrix} + V_P \\ &= -\frac{K_1}{k_2} + \frac{K_1}{k_2k_4}\left(k_3 - \frac{k_4k_5}{k_6}\right) + V_P,\end{aligned} \tag{2.34}$$

and,

$$\begin{aligned}\beta_3 &= \frac{\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}}{C_T(t)} \\ &= \frac{-\frac{1}{k_2k_4k_6}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} -k_4k_6 & k_4k_6 & -k_4k_6 \\ k_3k_6 & k_2k_6+k_3k_6 & k_3k_6 \\ -k_4k_5 & k_4k_5 & -k_2k_4-k_4k_5 \end{bmatrix}\begin{bmatrix} C_F(t) \\ C_S(t) \\ C_N(t) \end{bmatrix}}{C_T(t)} \\ &= \frac{-k_3}{k_2k_4}\frac{C_t(t)}{C_T(t)} + \frac{1}{k_2}\left(1+\frac{k_5}{k_6}\right)\left(\frac{C_F(t)-C_S(t)+C_N(t)}{C_T(t)}\right) - \frac{1}{k_6}\frac{C_N(t)}{C_T(t)} - \frac{1}{k_4}\frac{C_S(t)}{C_T(t)}.\end{aligned} \tag{2.35}$$

2.2 Logan graphical analysis using a reference region

Distribution volume ratio and non-displaceable binding potential

Arterial blood sampling to obtain plasma time-activity curve (pTAC), $C_p(t)$, is invasive and uncomfortable for patients [21, 32]. It can also be expensive, which is also undesirable. An alternative approach is to use a time-activity curve (TAC) of a reference region [3] instead of pTAC to avoid blood sampling. A reference region has ignorable specific binding sites for the radiotracer and a density of nonspecific binding sites equal to that in the ROI [3]. For the ^{11}C-PiB PET data used in this study, the cerebellum gray matter region was used as the reference region [33]. This approach of using a reference region TAC in place of pTAC is referred to as

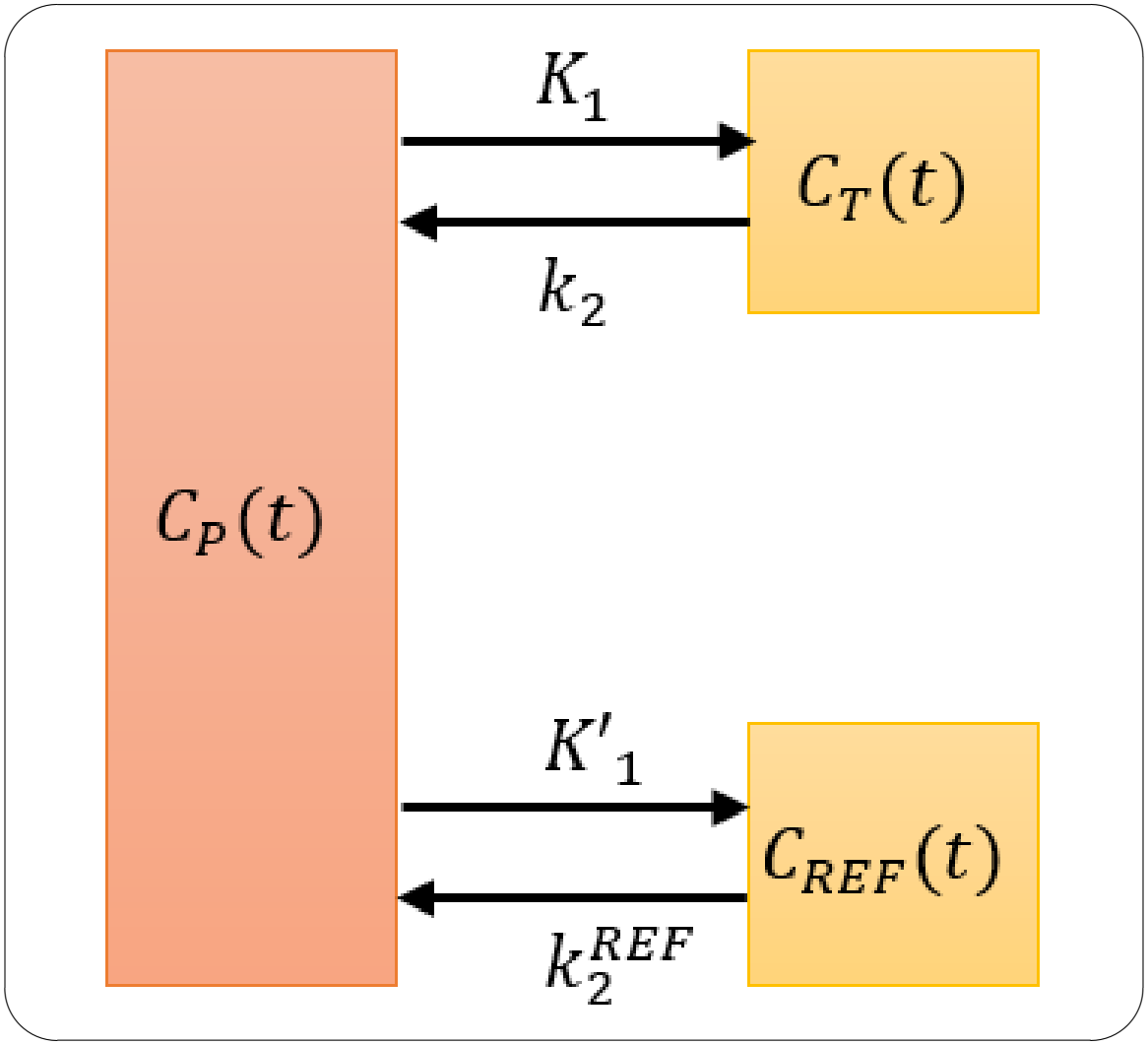


Figure 2.3: *One-tissue reference region compartmental model.*

reference LGA, and the output parameter is the distribution volume ratio (DVR). DVR is a measure of the ratio of DV in a receptor-rich region (ROI) to a non-receptor (reference) region, and is used for assessing receptor availability [3].

Replacing $\int_0^t C_T(u)\,du$ and $C_T(t)$ in Eq. (2.21) by $\int_0^t C_{REF}(u)\,du$ and $C_{REF}(t)$, respectively, and making $\int_0^t C_P(u)\,du$ the subject of the formula we get,

$$\int_0^t C_P(u)\,du = \frac{1}{DV^{REF}}\left(\int_0^t C_{REF}(u)\,du - \beta_{REF}C_{REF}(t)\right), \tag{2.36}$$

where the superscript ^{REF} is for reference region.

Substituting for $\int_0^t C_P(u)\,du$ in Eq. (2.21) as per Eq. (2.36), we get that,

$$\frac{\int_0^t C_T(u)\,du}{C_T(t)} = \frac{DV}{DV^{REF}}\left(\frac{\int_0^t C_{REF}(u)\,du - \beta_{REF}C_{REF}(t)}{C_T(t)}\right) + \beta. \tag{2.37}$$

For a two-compartment (one-tissue) reference region shown in Fig. 2.3, we write,

$$\beta_{REF} = -\frac{1}{k_2^{REF}}, \tag{2.38}$$

as per Eq. (2.31). Substituting for β_{REF} in Eq. (2.37) gives,

$$\frac{\int_0^t C_T(u)\,du}{C_T(t)} = DVR\left(\frac{\int_0^t C_{REF}(u)\,du + C_{REF}(t)/k_2^{REF}}{C_T(t)}\right) + \beta, \tag{2.39}$$

where $DVR = \dfrac{DV}{DV^{REF}}$.

Replacing k_2^{REF} in Eq. (2.39) with an average value, $\bar{k}_2^{REF}$, over all voxels in the reference region, we get that,

$$\frac{\int_0^t C_T(u)\,du}{C_T(t)} = DVR\left(\frac{\int_0^t C_{REF}(u)\,du + C_{REF}(t)/\bar{k}_2^{REF}}{C_T(t)}\right) + \beta, \tag{2.40}$$

Equation (2.40) is the reference LGA, which uses a reference region TAC instead of pTAC.

Figure 2.4 shows an example of a simulated reference time-activity curve (C_{REF}) and a ROI time-activity curve (C_T) in 2.4(a), and a corresponding Logan plot in 2.4(b). The tTACs in Fig. 2.4(a) are averages of 1024 simulated noisy tTACs, which the details will be given in Section 4.1.1. The data plot in the Logan plot in Fig. 2.4(b) achieves a linear relationship at a later time after the radiotracer administration. These data points, indicated with filled circles, are the ones used to estimate the Logan slope (DVR in Eq. (2.40)). Time t^* is then the time point before the first filled point, which correspond to 30 minutes in Fig. 2.4(a).

DVR can also be interpreted in terms of BP_{ND}, which is related to DVR by,

$$BP_{ND} = DVR - 1. \tag{2.41}$$

BP_{ND} compares the concentration of the administered radiotracer in receptor-rich to receptor-free regions [34]. BP_{ND} is proportional to the density of binding sites [3, 34], which denotes the concentration of the target receptors in ROI. Imaging BP_{ND} is therefore a direct quantification

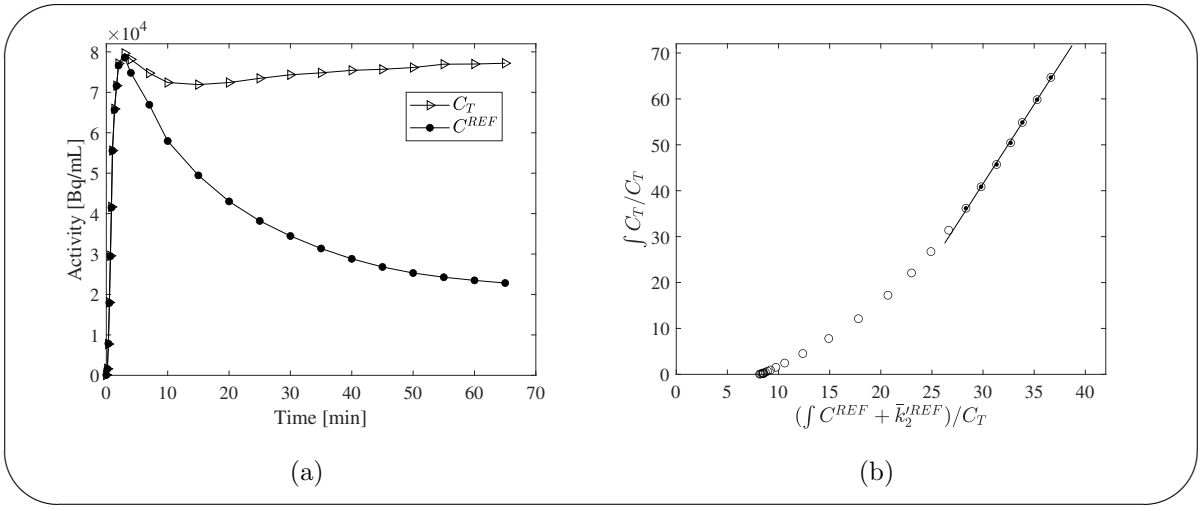


Figure 2.4: *An example of a reference time-activity curve (C_{REF}) and a ROI time-activity curve (C_T) in (a), and a corresponding Logan plot in (b).*

of the receptors of interest. BP_{ND} can also be comprehended in terms of its *in vitro* equivalence quantity, $B_{max}\, K'_D$, which assesses the capacity of the ROI for radiotracer-receptor binding [35]. This quantity is a product of the total number of receptors (B_{max}, i.e., free receptor concentration in the ROI) and their affinity (K'_D) for the radiotracer [35, 36].

Here, let us take a deeper look into previous methods for bias reduction: Despite the efficiency and ease of use presented by LGA, further investigations [4, 5, 6, 7] showed that the estimates of the LGA slope are negatively biased due to the inherent noise in PET data (tTACs), and that the bias increases with both the noise level and the LGA slope. Due to these findings, a number of studies have been carried out to try to reduce this bias. The noise in the LGA variables is specifically due to the noisy term $C_T(t)$ in the denominator of both the LGA variables, as it can be seen in Eqs. (2.26) and (2.40), presenting a correlated errors-in-variables problem. The bias is then due to that the OLS regression only accounts for the errors in the response variable, assuming the predictor variable is error-free. This has been analysed in [4].

One approach [11] for addressing the bias problem is that, since OLS does not account for errors in both variables, ODR, which minimizes the perpendicular distances from data points to the estimated regression line, does account for errors in both variables and thus would reduce the influence of the noise in the estimates of the LGA slope. This approach only removed a part

of the bias, and it came at the cost of precision [16]. The reason to not removing much of the bias could be due to that ODR is based on the assumption that the variances of the error in variables are equal, which is generally not the case in practical situations, and specifically not the case for LGA.

As it is the tradition, assuming that the integral terms in the numerators of Eq. (2.21) are noise-free, another approach [37] proposed a different method by re-arranging the original LGA by replacing the noisy term $C_T(t)$ in the denominators of Eq. (2.21) by $C_P(t)$ and continued as follows:

$$\begin{aligned}\frac{\int_0^t C_T(u)\,du}{C_P(t)} &= (-\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{Q} + V_P)\frac{\int_0^t C_P(u)du}{C_P(t)} + \frac{\boldsymbol{U}^T\boldsymbol{K}^{-1}\boldsymbol{A}}{C_P(t)} \\ &= \alpha\frac{\int_0^t C_P(u)du}{C_P(t)} + \beta. \end{aligned} \tag{2.42}$$

Multiplying through by $C_P(t)$ gives,

$$\int_0^t C_T(u)\,du = \alpha\int_0^t C_P(u)du + \beta C_P(t). \tag{2.43}$$

OLS is then used to estimate DV (slope α) from Eq. (2.43), thus supposedly accounting for all the noise since the right-hand-side (RHS) of Eq. (2.43) is supposed to be noise-free in this case. This is because the unintegrated noisy term, $C_T(t)$, no longer appears on the RHS of Eq. (2.43); it does not appear on the left hand side either. This method did not really bring up about a great improvement, it was shown to be consistent with the classic LGA [37]. This could be due to that, though these integrals, $\int_0^t C_T(u)\,du$ and $\int_0^t C_P(u)\,du$ are assumed to be noise-free, they are actually not; they are only less noisy compared to $C_T(t)$, but not noise-free. It was also pointed out in another study [5] that with this formulation, the true equilibrium state condition must apply before linearity is achieved, and this generally requires a longer time compared to the classic LGA formulation, with $C_T(t)$ in the denominator.

Another approach, multilinear analysis 1 (MA1) [9, 10], is also a mathematical rearrangement of the classic LGA equation into a format that would otherwise minimize the error in the LGA variables. In MA1, Eq. (2.21) is rearranged to bring the noisy term, $C_T(t)$, on one side as,

$$C_T(t) = -\frac{\alpha}{\beta}\int_0^t C_P(u)\,du + \frac{1}{\beta}\int_0^t C_T(u)\,du. \tag{2.44}$$

OLS can then be used to estimate the regression parameters of Eq. (2.44), and DV would be

found from the ratio of the regression parameters, $\frac{-\alpha}{\beta}$ and $\frac{1}{\beta}$, as,

$$DV = -\left(-\frac{\alpha}{\beta}\right)\left(\frac{\beta}{1}\right) = \alpha. \tag{2.45}$$

MA1 reduced the bias but did so at the expense of precison [9, 2]. An extension for MA1 is multilinear reference tissue model 2 (MRTM2). MRTM2 can be obtained by multiplying through Eq. (2.40) by $C_T(t)$ to obtain,

$$\int_0^t C_T(u)\,du = DVR\left(\int_0^t C_{REF}(u)\,du + \frac{C_{REF}(t)}{\bar{k}_2'^{REF}}\right) + \beta^* C_T(t), \tag{2.46}$$

and making $C_T(t)$ the subject of the formula gives,

$$C_T(t) = -\frac{DVR}{\beta^*}\left(\int_0^t C_{REF}(u)\,du + \frac{C_{REF}(t)}{\bar{k}_2'^{REF}}\right) + \frac{1}{\beta^*}\int_0^t C_T(u)\,du. \tag{2.47}$$

OLS can then be used to estimate the regression parameters, $-\frac{DVR}{\beta^*}$ and $\frac{1}{\beta^*}$, and DVR can be found as,

$$DVR = -\left(-\frac{DVR}{\beta^*}\right)\left(\frac{\beta^*}{1}\right) = DVR. \tag{2.48}$$

The GLLS approach [8] is a temporal smoothing method that attempts to denoise tTACs. In this approach, each noisy tTAC is separated into two, which are then separately fit to two-compartments (one-tissue) model using GLLS [38]. This method performs well at low noise level, but large bias and poor precision has been observed at high noise level [39, 8].

Similar to GLLS, the method of PCA denoising [16] is also a temporal smoothing method in which a PCA-based lower dimension linear model is used to denoise tTACs. PCA would reduce the noise in tTACs by transforming the tTACs into a feature space with the same dimensionality as the original tTACs such that the tTACs could be represented by a reduced number of dominant features only, while retaining most of the important inherent characteristics of the original tTACs. In this case, all the noise component in the less important PCs would be lost, hence reducing the noise in the reduced dimension data set comprised of the most important PCs [16]. So the source of the correlated errors in the LGA variables is thus minimised. This method performs considerably well, especially in terms of variance reduction, but there still remains rooms of improvement in terms bias reduction.

It is therefore that; OLS is highly biased with moderate precision, ODR removes only a portion of the bias which comes with poor precision. The GLLS, MA1 and MRTM2 methods are significantly effective in removing the bias, but this happens at the expense of precision. PCA quite maintains precision and removes the bias significantly, but leaving a space for further researches over a wide range of radiotracers, as well as in terms of determining the appropriate number of PCs. Not withstanding this tremendous effort to reduce the effect of the noise in PET data on the estimates of the LGA slope, and though some of the methods significantly reduce the bias, the bias persists and further studies are necessary.

This study takes a two-step approach to address the bias problem. The first step is to reduce the noise in the LGA estimates from undenoised tTACs. In the second step, the approach in the first step is employed on denoised tTACs. Specifically, the first step is based on the demonstrations in [11, 4]. In these studies it can be seen that the regression method used to estimate the Logan slope influences the resulting bias. On this basis, the "first step" in this study approached the LGA bias problem from the EIV perspective by employing an alternative linear regression method referred to as LSC to estimate the LGA slope. Unlike OLS or ODR, LSC accounts for the errors in both the predictor and response variables by minimizing the weighted squared residuals in both the predictor and response variables. Even more, LSC incorporates the correlation of errors in these variables. Thus, LSC is expected to be appropriate for correlated errors in the LGA variables. LSC is mathematically superior to both OLS and ODR, since it is a general solution to the linear least-squares problem, and also because all other commonly known regression methods, including OLS and ODR, can be derived from LSC as special cases, under certain assumptions about the statistical properties of the data being studied. Furthermore, LSC has been shown to improve regression parameter estimates in the fields of geophysics [12] and biogeosciences [40]. LSC was assessed against MRTM2, which is the most accepted method, and the conventional OLS-based LGA. In the second step, LSC is combined with tTACs denoising techniques, PCA and CorrCA.

Ordinary least-squares

Since LGA is a linear equation. Let us first refresh on OLS and ODR regression methods. OLS regression is a statistical method used to estimate the relationship between one or more predictor variables and a response variable, by estimating the unknown parameters in a lin-

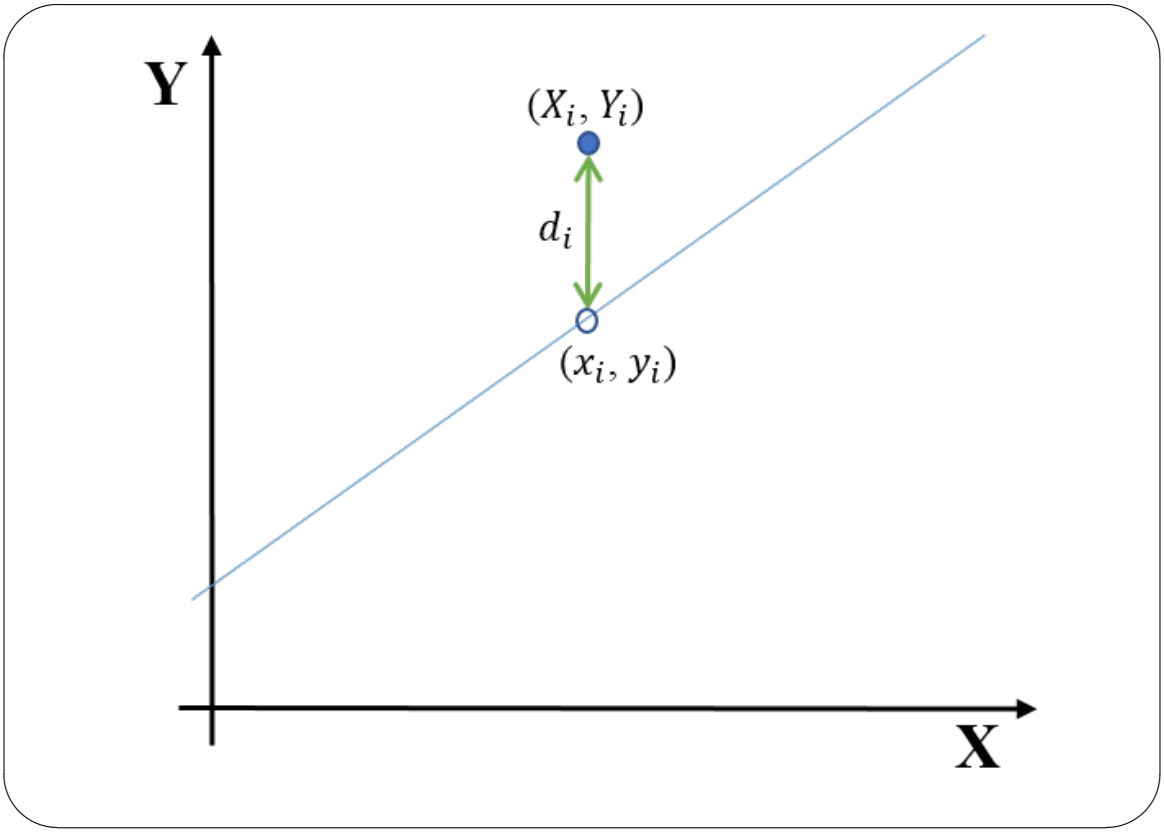


Figure 2.5: *A schematic figure for OLS.*

ear regression model. OLS estimates this relationship by minimizing the total squared errors [41, 42, 43], i.e., the sum of the squared differences between the observed and the predicted values of the response variable. Geometrically, this is the sum of the squared distances parallel to the axis of the response variable, between each scattered data point and the corresponding point on the estimated regression line. This interpretation can be seen in Fig. 2.5. Since the LGA is a bivariate system, i.e., one predictor variable and one response variable, henceforward OLS is presented based on a bivariate system.

Let us define $X = \{X_i \,|\, i = 1, 2, ..., n\}$ and $Y = \{Y_i \,|\, i = 1, 2, ..., n\}$ as the predictor and response variables, respectively. With reference to Fig. 2.5, the OLS line of best fit, $y = a + bx$, minimizes the quantity,

$$\begin{aligned} D &= \sum_{i=1}^{n} d_i^2 \\ &= \sum_{i=1}^{n} (Y_i - y_i)^2, \end{aligned} \tag{2.49}$$

subject to,

$$y_i = a + bX_i \qquad \text{for} \qquad i = 1, 2, \cdots, n, \tag{2.50}$$

where $\{y_i \,|\, i = 1, 2, ..., n\}$ are the estimated data points, lying on the estimated straight line. Substituting Eq. (2.50) into Eq. (2.49) gives,

$$D = \sum_{i=1}^{n} (Y_i - bX_i - a)^2 . \tag{2.51}$$

We therefore wish to solve for b and a by minimizing Eq. (2.51).

To minimize Eq. (2.51), consider that,

$$D = \sum_{i=1}^{n} (Y_i - bX_i - a)^2 = D(b, a). \tag{2.52}$$

We then differentiate Eq. (2.52) with respect to b and a to get,

$$\frac{\partial d^2}{\partial b} = -2 \sum_{i=1}^{n} (Y_i - bX_i - a) X_i, \tag{2.53}$$

and

$$\frac{\partial d^2}{\partial a} = -2 \sum_{i=1}^{n} (Y_i - bX_i - a) . \tag{2.54}$$

Setting Eq. (2.54) to zero we get,

$$\begin{aligned} &\sum_{i=1}^{n} (Y_i - bX_i - a) = 0 \\ &\implies a = \bar{Y} - b\bar{X}, \end{aligned} \tag{2.55}$$

where $\bar{X} = \sum_{i=1}^{n} X_i$ and $\bar{Y} = \sum_{i=1}^{n} Y_i$.

Setting Eq. (2.53) to zero we get, $\sum_{i=1}^{n} (Y_i - bX_i - a) = 0$, and substituting for a as per Eq. (2.55) gives,

$$\begin{aligned}
&\sum_{i=1}^{n}\left(Y_i - bX_i - \bar{Y} + b\bar{X}\right) = 0 \\
&\implies b\sum_{i=1}^{n}\left(X_i - \bar{X}\right) = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right) \\
&\implies b = \frac{\sum\left(Y_i - \bar{Y}\right)}{\sum\left(X_i - \bar{X}\right)} \\
&\therefore \quad b = \frac{\sum\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)}{\sum\left(X_i - \bar{X}\right)^2}.
\end{aligned} \tag{2.56}$$

Thus, Eqs. (2.56) and (2.55) respectively give the OLS estimates of slope b and y-intercept a of the straight line in Eq. (2.50).

We can see in Eq. (2.49) that OLS only minimizes the squared residuals of the response variable. This is based on the assumption that only the response variable is contaminated with measurement errors. However in practice, this is not always the case, and specifically for LGA as described earlier, both variables are contaminated and the contamination in both variables need to be accounted for.

In order to address the issues encountered with OLS, a variety of regression methods referred to as EIV have been developed. Before discussing the LSC method which is employed in this study, let us have a look at one of the EIV methods, ODR, which has been previously applied to LGA [11].

Orthogonal distance regression

ODR (also referred to as PCA linear regression in some literature, eg., [6]) is an EIV regression method that was developed to account for errors in both the predictor and response variables by minimising the perpendicular distances from the estimated regression line to the observed data points [44, 45], unlike OLS which minimizes the vertical distances. Fig. 2.6 shows a schematic diagram for ODR.

In ODR, both X_i and Y_i measurement are considered to be noisy. Therefore, ODR estimates

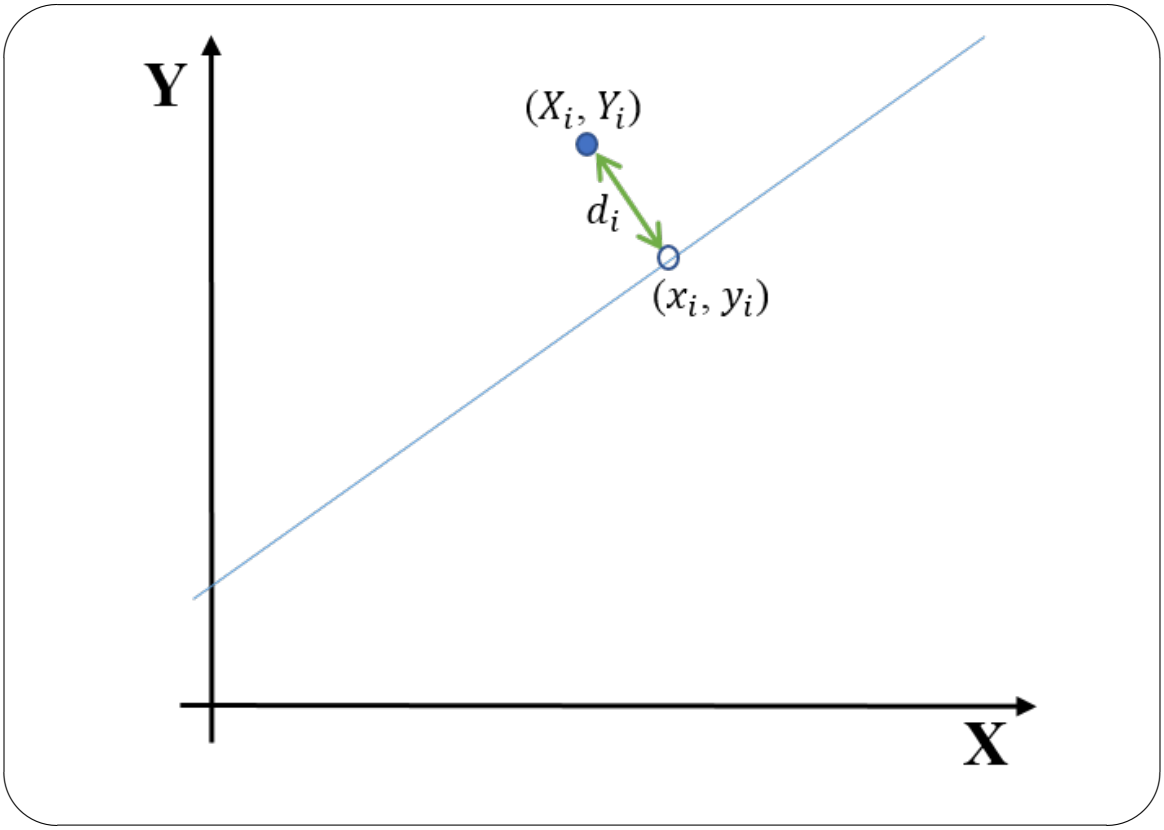


Figure 2.6: *A schematic figure for ODR.*

the best fit line, $y = a + bx$, by minimizing the residuals in both the X_i and Y_i measurements as represented in the equation,

$$D = \sum_{i=1}^{N} (X_i - x_i)^2 + k(Y_i - y_i)^2, \tag{2.57}$$

subject to,

$$y_i = bx_i + a, \tag{2.58}$$

where $\{(x_i, y_i) \,|\, i = 1, 2, ..., n\}$ are the estimated points sitting on the estimated straight line, and k is the ratio of variances of the errors in the measured data given by $\sigma_{xi}^2/\sigma_{yi}^2$. Note here that Eq. (2.58) differs from Eq. (2.50), in that Eq. (2.58) has x_i instead of X_i which is in Eq. (2.50).

Substituting Eq. (2.58) in Eq. (2.57) gives,

$$D = \sum_{i=1}^{N} (X_i - x_i)^2 + k(Y_i - bx_i - a)^2. \tag{2.59}$$

The respective solutions of the slope and y-intercept from Eq. (2.59) are,

$$b = \frac{(kS_{yy} - S_{xx}) + \sqrt{(S_{xx} - kS_{yy})^2 + 4kS_{xy}^2}}{2kS_{xy}}, \tag{2.60}$$

and

$$a = \bar{Y} - \alpha' \bar{X}, \tag{2.61}$$

where $S_{xx} = \sum (X_i - \bar{X})^2$, $S_{yy} = \sum (Y_i - \bar{Y})^2$ and $S_{xy} = \sum (X_i - \bar{X})(Y_i - \bar{Y})$.

In Eq. (2.60), if we assume that $\sigma_{xi} = \sigma_{yi}$, then $k = 1$, and Eq. (2.60) becomes,

$$b = \frac{(S_{yy} - S_{xx}) + \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}}{2S_{xy}}. \tag{2.62}$$

Equations (2.62) and (2.61) are the ODR solutions of the slope and y-intercept, respectively. Thus, ODR is mainly based on this assumption that the variances of the errors in the variables are equal. Clearly this is not always the case in practical situations, that is why ODR regression does not make the best regression method. Next, I present LSC linear regression method according to [13, 14, 12, 46] and show how it can be used to estimate the LGA slope.

Chapter 3

Reducing the bias in the BP_{ND} estimates using least-squares cubic linear regression

3.1 Least-squares cubic linear regression

LSC is an EIV-based regression method. Building upon the work of Demming [47], LSC was developed by Derek York in (1966 [13]), in which he considered the errors in both the predictor and response variables but not the correlation of the errors in variables. He later (1969 [14]) extended it to take into account the correlation of measurement errors in variables. Further, in 2004, York and colleagues [12] followed up, and presented compact equations for the slope of LSC, and unified the the expressions of the standard errors of the slope and y-intercept with maximum likelihood expressions. LSC fully considers the errors in both the predictor and response variables [12]. This is achieved by minimizing the sum of weighted squared errors in both variables, and by including the correlation of these errors [12]. Under specific assumptions, LSC can be shown to reduce to either OLS or other common regression methods (see Appendix A), rendering LSC to be a general solution to the linear regression problem [13, 14].

To discuss the mathematical context of LSC, consider the observed predictor and response variables, $X = \{X_i \,|\, i = 1, 2, ..., n\}$ and $Y = \{Y_i \,|\, i = 1, 2, ..., n\}$, respectively. The LSC slope is

obtained by minimizing the equation below [12];

$$S = \sum_{i=1}^{n} \left\{ w(X_i)(x_i - X_i)^2 - 2r_i\alpha_i(x_i - X_i)(y_i - Y_i) + w(Y_i)(y_i - Y_i)^2 \right\} \frac{1}{1 - r_i^2}, \tag{3.1}$$

where $(x_i,\ y_i)$ are the estimated points of $(X_i,\ Y_i)$ sitting on the estimated straight line, and $\alpha_i = \sqrt{w(X_i)w(Y_i)}$. Equation (3.1) includes the residuals in both variables $[(x_i - X_i),\ (y_i - Y_i)]$, the weights $[w(X_i),\ w(X_i)]$, and the correlation of the errors in variables (r_i).

Deriving Eq. (3.1)

The derivation of Eq. (3.1) is demonstrated in [46] as follows:

For the respective i^{th} observed and estimated points $(X_i,\ Y_i)$ and $(x_i,\ y_i)$, let us denote the residuals column vector by,

$$\mathbf{v}_i^T = [u(X_i),\ v(Y_i)]\,, \tag{3.2}$$

where, superscript T denotes a transpose, and

$$\begin{aligned} u(X_i) &= x_i - X_i \\ v(Y_i) &= y_i - Y_i. \end{aligned} \tag{3.3}$$

We can represent the column vector of residual vectors as,

$$\begin{aligned} \mathbf{v}^T &= [v(X_1)\ \ v(Y_1)\ \mid\ v(X_2)\ \ v(Y_2)\ \mid\ \cdots\ \ \cdots\ \mid\ v(X_n)\ \ v(Y_n)] \\ &= \left[\mathbf{v}_1^T,\ \ \mathbf{v}_2^T, \cdots\ \ \mathbf{v}_n^T\right]. \end{aligned} \tag{3.4}$$

Let $\left(s^2(X_i),\ \ s^2(Y_i)\right)$ and $s(X_i\,Y_i)$, respectively, denote the variances and covariances associated with the observed pair $(X_i,\ \ Y_i)$. The block-diagonal variance matrix is defined

as,

$$\mathbf{Q} = \left[\begin{array}{cc:cc:ccc:cc} s^2(X_1) & s(X_1\,Y_1) & 0 & 0 & \cdots & \cdots & 0 & 0 \\ s(X_1\,Y_1) & s^2(Y_1) & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \hdashline 0 & 0 & s^2(X_2) & s(X_2\,Y_2) & \cdots & \cdots & 0 & 0 \\ 0 & 0 & s(X_2\,Y_2) & s^2(Y_2) & \cdots & \cdots & 0 & 0 \\ \hdashline \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \vdots & \vdots \\ \hdashline 0 & 0 & 0 & 0 & \cdots & \cdots & s^2(X_n) & s(X_n\,Y_n) \\ 0 & 0 & 0 & 0 & \cdots & \cdots & s(X_n\,Y_n) & s^2(Y_n) \end{array}\right]$$

$$= \begin{bmatrix} \mathbf{Q_1} & 0 & \cdots & 0 \\ 0 & \mathbf{Q_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q_n} \end{bmatrix}, \tag{3.5}$$

in which the 2-by-2 sub-matrices, $\mathbf{Q_1}, \mathbf{Q_2}, \cdots, \mathbf{Q_n}$, are symmetric. If we refer to $X_1, X_2, \cdots X_N$ as a finite population of N quantities, the population mean and variance are respectively given by,

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \quad \text{and} \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (X_i - \mu)^2 . \tag{3.6}$$

Given a sample of n quantities from an infinite population, the unbiased estimates of the mean and variance are respectively,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad s^2(X) = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 . \tag{3.7}$$

The standard deviation is given by the positive square-root of the variance, i.e.,

$$s(X) = \sum_{i=1}^{n} \sqrt{\frac{\left(X_i - \bar{X}\right)^2}{n-1}}. \tag{3.8}$$

For a pair of two samples (X_i, Y_i) both of size n, an unbiased estimate of covariance is,

$$s(XY) = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) . \tag{3.9}$$

The block-diagonal weight matrix of the measurements denoted by $\mathbf{W}$ is defined by,

$$\mathbf{W} = \mathbf{Q^{-1}} \begin{bmatrix} \mathbf{W_1} & 0 & \cdots & 0 \\ 0 & \mathbf{W_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{W_n} \end{bmatrix}, \tag{3.10}$$

in which the 2-by-2 sub-matrices, $\mathbf{W_1} = \mathbf{Q_1^{-1}}$, $\mathbf{W_2} = \mathbf{Q_2^{-1}}, \cdots$, $\mathbf{W_n} = \mathbf{Q_n^{-1}}$, are symmetric. This means that the weight of measurements are given by the inverse of the variances.

From Eq. (3.5) we can extract the i^{th} variance matrix $\mathbf{Q_i}$ as,

$$\mathbf{Q_i} = \begin{bmatrix} s^2(X_i) & s(X_i\,Y_i) \\ s(X_i\,Y_i) & s^2(Y_i) \end{bmatrix}, \tag{3.11}$$

and the inverse of the i^{th} variance matrix will give the i^{th} weight matrix $\mathbf{W_i}$ as,

$$\mathbf{W_i} = \mathbf{Q_i^{-1}} = \frac{1}{s^2(X_i)s^2(Y_i) - s^2(X_iY_i)} \begin{bmatrix} s^2(Y_i) & -s(X_i\,Y_i) \\ -s(X_i\,Y_i) & s^2(X_i) \end{bmatrix}. \tag{3.12}$$

The correlation (r_i) between a pair of measurements (X_i, Y_i) is given by,

$$r_i = \frac{s(X_iY_i)}{s(X_i)s(Y_i)}, \tag{3.13}$$

which gives,

$$\frac{1}{s^2(X_i)s^2(Y_i) - s^2(X_iY_i)} = \frac{1}{\left(1 - r_i^2\right) s^2(X_i)s^2(Y_i)}. \tag{3.14}$$

Substituting Eq. (3.14) into Eq. (3.12) gives,

$$\mathbf{W}_i = \frac{1}{1-r_i^2} \left[\begin{array}{c:c} \frac{1}{s^2(X_i)} & -\frac{r_i}{s(X_i)s(Y_i)} \\ \hdashline -\frac{r_i}{s(X_i)s(Y_i)} & \frac{1}{s^2(Y_i)} \end{array}\right]. \tag{3.15}$$

Expressing the weights as the inverses of the unbiased estimates of the variances we get,

$$w(X_i) = \frac{1}{s^2(X_i)} \quad \text{and} \quad w(Y_i) = \frac{1}{s^2(Y_i)}. \tag{3.16}$$

We can then write Eq. (3.15) as,

$$\mathbf{W}_i = \frac{1}{1-r_i^2} \left[\begin{array}{c:c} w(X_i) & -r_i(\alpha_i) \\ \hdashline -r_i\alpha_i & w(Y_i) \end{array}\right], \tag{3.17}$$

where,

$$\alpha_i = \sqrt{w(X_i)w(Y_i)} = \frac{1}{s(X_i)s(Y_i)}. \tag{3.18}$$

The least-squares function S in Eq. (3.1) is equal to the sum of weighted squares of residuals, and can be defined as a matrix product,

$$S = \mathbf{v}^T\mathbf{W}\mathbf{v}. \tag{3.19}$$

By Eqs. (3.4) and (3.10), we can write S as below,

$$S = \mathbf{v_1}^T\mathbf{W_1}\mathbf{v_1} + \mathbf{v_2}^T\mathbf{W_2}\mathbf{v_2} + \mathbf{v_3}^T\mathbf{W_3}\mathbf{v_3} + \cdots + \mathbf{v_n}^T\mathbf{W_n}\mathbf{v_n}, \tag{3.20}$$

and using Eq. (3.17) we can write,

$$\mathbf{v_i}^T\mathbf{W_i}\mathbf{v_i} = \frac{1}{1-r_i^2}\begin{bmatrix} v(X_i) & v(Y_i) \end{bmatrix} \left[\begin{array}{c|c} w(X_i) & -r_i(\alpha_i) \\ \hline -r_i\alpha_i & w(Y_i) \end{array}\right] \begin{bmatrix} v(X_i) \\ \\ v(Y_i) \end{bmatrix}$$

$$= \frac{1}{1-r_i^2}\left\{w(X_i)(v(X_i))^2 - 2r_i\alpha v(X_i)v(Y_i) + w(Y_i)(v(Y_i))^2\right\}$$

$$= \frac{1}{1-r_i^2}\left\{w(X_i)(x_i - X_i)^2 - 2r_i\alpha(x_i - X_i)(y_i - Y_i) + w(Y_i)(y_i - Y_i)^2\right\}. \tag{3.21}$$

Summing over $i = 1, 2, \cdots, n$ gives,

$$S = \sum_{i=1}^{n} \frac{1}{1-r_i^2}\left\{w(X_i)(x_i - X_i)^2 - 2r_i\alpha(x_i - X_i)(y_i - Y_i) + w(Y_i)(y_i - Y_i)^2\right\} \tag{3.22}$$

as in Eq. (3.1).

Solving for the least-squares cubic line of best fit

The line of best fit $y = a + bx$ minimizes the sum of the weighted squares of residuals [13, 14]. The adjusted/estimated points $(x_i,\ y_i)$ of the measurements $(X_i,\ Y_i)$ are required to lie on the estimated straight line, and therefore the regression parameters a, and b can be obtained by minimizing S (Eqs. (3.22) and (3.1)) subject to the constraints that,

$$y_i = a + bx_i \qquad \text{for} \qquad i = 1, 2, \cdots, n. \tag{3.23}$$

It is shown in [46] how this task can be achieved using the **Lagrange multipliers** optimization. The function to be minimized is the **Lagrangian** L written as,

$$L = S + 2\sum_i \lambda_i \left(a + bx_i - y_i\right), \tag{3.24}$$

where λ_i are unknown **Lagrange multipliers**. The **Lagrangian**, L, has five variables, λ_i, a, b, x_k, and y_k, and can be minimized by setting its first partial derivatives to zero as below,

$$\frac{\partial L}{\partial \lambda_i} = \frac{\partial L}{\partial a} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial y_i} = 0. \tag{3.25}$$

Working out the zero-equated partial derivatives in Eq. (3.25) gives,

$$\begin{aligned}
\frac{\partial L}{\partial \lambda_i} &= 2\sum_i (a + bx_i - y_i) = 0 \\
\frac{\partial L}{\partial a} &= 2\sum_i \lambda_i = 0 \\
\frac{\partial L}{\partial b} &= 2\sum_i \lambda_i x = 0 \\
\frac{\partial L}{\partial x_i} &= \sum_i \frac{1}{1-r_i^2}\left\{2w(X_i)(x_i - X_i) - 2r_i\alpha_i(y_i - Y_i)\right\} + 2\sum_i \lambda_i b = 0 \\
\frac{\partial L}{\partial y_i} &= \sum_i \frac{1}{1-r_i^2}\left\{2w(Y_i)(y_i - Y_i) - 2r_i\alpha_i(x_i - X_i)\right\} - 2\sum_i \lambda_i = 0.
\end{aligned} \tag{3.26}$$

In order, the expressions in Eq. (3.26) gives,

$$y_i = a + bx_i \tag{3.27}$$

$$\lambda_i = 0 \tag{3.28}$$

$$\lambda_i x = 0 \tag{3.29}$$

$$w(X_i)(x_i - X_i) - r_i\alpha_i(y_i - Y_i) + \lambda_i(1 - r_i^2)b = 0 \tag{3.30}$$

$$w(Y_i)(y_i - Y_i) - r_i\alpha_i(x_i - X_i) - \lambda_i(1 - r_i^2) = 0, \tag{3.31}$$

for each i.

Respectively solving for $x_i - X_i$ and $y_i - Y_i$ from Eqs. (3.30) and (3.31) gives,

$$x_i - X_i = \frac{1}{\alpha_i^2}\left(r_i\alpha_i\lambda_i - bw(Y_i)\lambda_i\right) \implies x_i = X_i + \frac{1}{\alpha_i^2}\left(r_i\alpha_i\lambda_i - bw(Y_i)\lambda_i\right) \tag{3.32}$$

$$y_i - Y_i = \frac{1}{\alpha_i^2}\left(w(X_i)\lambda_i - br_i\alpha_i\lambda_i\right) \implies y_i = Y_i + \frac{1}{\alpha_i^2}\left(w(X_i)\lambda_i - br_i\alpha_i\lambda_i\right). \tag{3.33}$$

Substituting Eqs. (3.32) and (3.33) into Eq. (3.27) and solving for the **Lagrange multipliers** gives,

$$\lambda_i = W_i(a + bX_i - Y_i), \tag{3.34}$$

where W_i is the weight function given by,

$$W_i = \frac{\alpha_k^2}{b^2 w(Y_i) + w(X_i) - 2br_i\alpha_i}. \tag{3.35}$$

Substituting Eq. (3.34) into Eqs. (3.28) and (3.29), respectively, gives,

$$\sum W_i(a + bX_i - Y_i) = 0 \tag{3.36}$$

and

$$\sum x_i W_i(a + bX_i - Y_i) = 0. \tag{3.37}$$

Expanding Eq. (3.36) and solving for a gives,

$$a = \frac{\sum W_i Y_i}{\sum W_i} - b\frac{\sum W_i X_i}{\sum W_i} \tag{3.38}$$

$$\therefore \quad a = \bar{Y} - b\bar{X}, \tag{3.39}$$

where $(\bar{X},\ \bar{Y})$ is the **Centroid** of the data:

$$\bar{X} = \frac{\sum W_i X_i}{\sum W_i}, \quad \bar{Y} = \frac{\sum W_i Y_i}{\sum W_i}. \tag{3.40}$$

Equations (3.38) and (3.39) mean that, according to the LSC approach, the best straight line goes through the centroid/center of gravity of the data set given by $(\bar{X},\ \bar{Y})$ in Eq. (3.40).

Let us define the **centroidal** coordinates $(U_i,\ V_i)$ as,

$$U_i = X_i - \bar{X}, \quad V_i = Y_i - \bar{Y}. \tag{3.41}$$

Then we can write Eq. (3.39) as

$$\begin{aligned} a + bX_i - Y_i &= a + b(U_i + \bar{X}) - (V_i + \bar{Y}) \\ &= a + bU_i + b\bar{X} - V_i - \bar{Y} \\ &= bU_i - V_i \end{aligned} \tag{3.42}$$

Respectively substituting Eq. (3.42) into Eqs. (3.34) and (3.37), we get,

$$\lambda_i = W_i(bU_i - V_i) = 0 \tag{3.43}$$

and

$$\sum W_i(bU_i - V_i) = 0. \tag{3.44}$$

Substituting Eqs. (3.32), (3.42), (3.43), and (3.44) into Eq. (3.37), gives,

$$\begin{aligned} &b^3 \sum \frac{W_i^2}{w(X_i)} U_i^2 - b^2 \sum W_i^2 \left(\frac{r_i}{\alpha_i} U_i + \frac{2}{w(X_i)} V_i \right) U_i \\ &- b \sum W_i \left(U_k^2 \frac{2 W_i r_i}{\alpha_i} U_i V_i - \frac{W_i}{w(X_i)} V_i^2 \right) + \sum W_i \left(U_i - \frac{W_i r_i}{\alpha_i} V_i \right) = 0. \end{aligned} \tag{3.45}$$

Equation (3.45) is referred to as the **Generalized least-squares cubic regression** [14, 12]. Being a general solution to the least-squares problem means that all other known regression methods, like ODR and OLS can be obtained from Eq. (3.45) as special cases, under specific assumptions. In addition, the solution obtained by York in [13] in which it is assumed that the errors in the variables are uncorrelated can also be obtained from Eq. (3.45) by setting r_i to zero. The "*cubic*" in "**Generalized least-squares cubic regression**" refers to that Eq. (3.45) is a cubic (or rather pseudo-cubic, as explained below) equation of b.

The pseudo-cubic equation:

Equation (3.45) is not really a cubic equation of b because the qantity W_i (Eq. (3.35)) and the **centroidal** coordinates $(U_i,\ V_i)$ (Eq. (3.41)) are functions of b. However, this pseudo-cubic equation can be reduced to a truc cubic equation by substituting in an approximate value for b in the weight function W_i. From there, it can then be solved directly as a cubic equation, giving three roots for b.

A concise solution for b is given in [12]. It was obtained by first reducing Eq. (3.45) to a quadratic equation as,

$$b^2 \sum W_i^2 \left(\frac{U_i V_i}{w(X_i)} - \frac{r U_i^2}{\alpha_i} \right) + b \sum W_i^2 \left(\frac{U_i^2}{w(Y_i)} - \frac{V_i^2}{w(X_i)} \right) - \sum W_i^2 \left(\frac{U_i V_i}{w(Y_i)} - \frac{r V_i^2}{\alpha_i} \right) = 0, \tag{3.46}$$

and subsequently to a linear equation as,

$$b \sum W_i^2 \left(\frac{U_i}{w(Y_i)} + \frac{b V_i}{w(X_i)} - \frac{b r_i U_i}{\alpha_i} \right) U_i - \sum W_i^2 \left(\frac{U_i}{w(Y_i)} + \frac{b V_i}{w(X_i)} - \frac{r_i V_i}{\alpha_i} \right) V_i = 0. \tag{3.47}$$

The slope coefficient, b, can then be obtained from Eq. (3.47) as,

$$b = \frac{\sum W_i^2 \left(\frac{U_i}{w(Y_i)} + \frac{b V_i}{w(X_i)} - \frac{r_i V_i}{\alpha_i} \right) V_i}{\sum W_i^2 \left(\frac{U_i}{w(Y_i)} + \frac{b V_i}{w(X_i)} - \frac{b r_i U_i}{\alpha_i} \right) U_i}. \tag{3.48}$$

Further, as shown in [46], Eq. (3.47) can be written as,

$$\begin{aligned} b\sum W_iU_iW_i\left(\frac{U_i}{w(Y_i)}+\frac{bV_i}{w(X_i)}-\frac{br_iU_i}{\alpha_i}\right)-b\sum W_iU_iW_i\left(\frac{r_iV_i}{\alpha_i}\right) \\ =\sum W_iV_iW_i\left(\frac{U_i}{w(Y_i)}+\frac{bV_i}{w(X_i)}-\frac{r_iV_i}{\alpha_i}\right)-b\sum W_iU_iW_i\left(\frac{r_iV_i}{\alpha_i}\right), \end{aligned} \tag{3.49}$$

which can be further simplified to give the expression,

$$\begin{aligned} b\sum W_iU_iW_i\left(\frac{U_i}{w(Y_i)}+\frac{bV_i}{w(X_i)}-\frac{r_i}{\alpha_i}(bU_i+V_i)\right)=\sum W_iV_iW_i \\ \left(\frac{U_i}{w(Y_i)}+\frac{bV_i}{w(X_i)}-\frac{r_iV_i}{\alpha_i}(bU_i+V_i)\right). \end{aligned} \tag{3.50}$$

The LSC slope parameter, b, can then be expressed in simplified terms as,

$$b=\frac{\sum W_i\beta_iV_i}{\sum W_i\beta_iU_i}. \tag{3.51}$$

where β_i is given by,

$$\beta_i=W_i\left(\frac{U_i}{w(Y_i)}+\frac{bV_i}{w(X_i)}-(bU_i+V_i)\frac{r_i}{\alpha_i}\right). \tag{3.52}$$

LSC solution for the slope, b, and y-intercept, a:

$$b=\frac{\sum W_iB_iV_i}{\sum W_iB_iU_i} \qquad \text{from Eq. (3.51)}$$

$$a=\bar{Y}-b\bar{X} \qquad \text{from Eq. (3.39)}$$

Respective variances of a and b:

$$\begin{aligned} \sigma_b^2&=\frac{1}{\sum W_iu_i^2} \\ \sigma_a^2&=\frac{1}{\sum W_i}+\bar{x}^2\sigma_b^2 \end{aligned} \tag{3.53}$$

Estimated/adjusted points $(x_i,\, y_i)$:

$$\begin{aligned} x_i&=\bar{X}+\beta_i \\ y_i&=\bar{Y}+b\beta_i \end{aligned} \tag{3.54}$$

Residuals (x_i, y_i):

$$v(X_i) = x_i - X_i = \beta_i - U_i$$
$$v(Y_i) = y_i - Y_i = b\beta_i - V_i \tag{3.55}$$

$\bar{x}$ in Eq. (3.53), and the corresponding $\bar{y}$ are respectively given by,

$$\bar{x} = \frac{\sum W_i x_i}{\sum W_i} \quad \text{and} \quad \bar{y} = \frac{\sum W_i y_i}{\sum W_i}, \tag{3.56}$$

where,

$$u_i = x_i - \bar{x} \quad \text{and} \quad v_i = y_i - \bar{y}. \tag{3.57}$$

Here let us again note that W_i itself (which is a function of b), and β_i, V_i and, U_i, which are all functions of W_i (and therefore functions of b) appear in Eq. (3.51). Therefore an appropriate approximate value for b shoul be substituted in W_i so that Eq. (3.51) can be used to estimate b. In this study, the approximate value which was substituted in W_i was estimated by OLS. For improved estimates of b, the estimates were iterated. A complete mathematical feature of LSC is provided in [46].

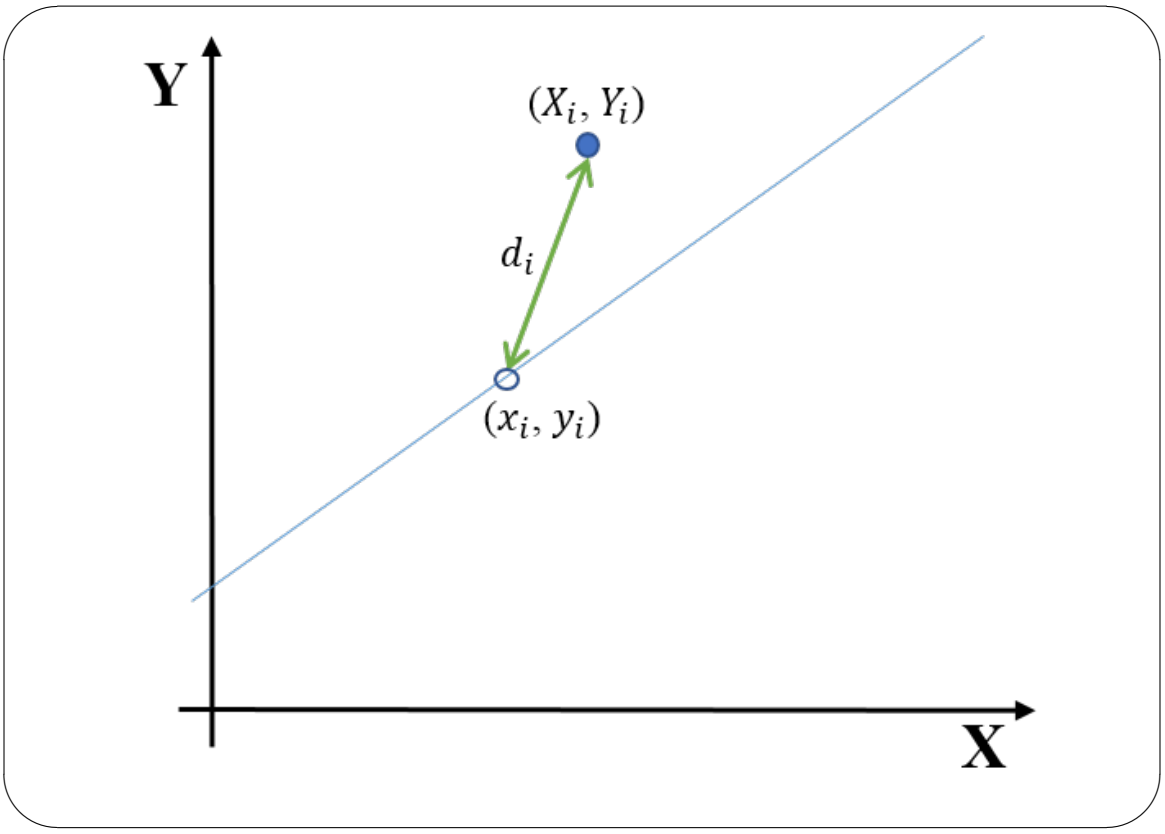


Figure 3.1: *An illustration of LSC regression: LSC for a general Y vs X. The observed point $(X_i,\ Y_i)$ is projected onto the estimated point $(x_i,\ y_i)$, making a residual d_i.*

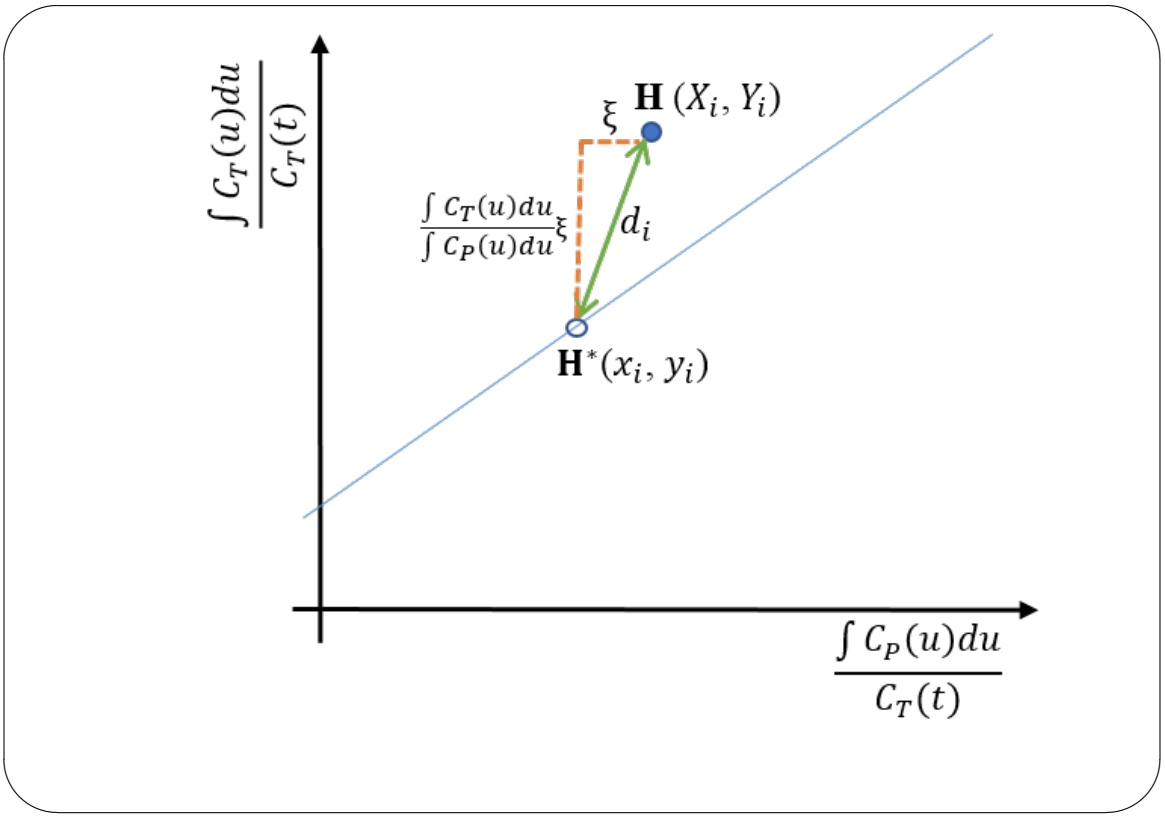


Figure 3.2: *An illustration of LSC regression: LSC visualized in terms of the LGA variables. The main point is that the direction along the minimization distance is not restricted to being either vertical (as is for OLS) or perpendicular (as is for ODR), but may vary depending on the value of* $\frac{\int C_T(u)\,du}{\int C_P(u)\,du}\xi/\xi = \frac{\int C_T(u)\,du}{\int C_P(u)\,du}$, *as implied in Eq. (3.58).*

Analysing least-squares cubic regression in terms of the Logan graphical analysis

Figure 3.1 shows a schematic diagram for a general LSC regression of Y vs X, and Fig. 3.2 expresses LSC specifically in terms of the LGA variables.

To explain LSC in specific terms of LGA, let us consider the LGA in Eq. (2.21). If the two integral terms, $\int C_P(u)\,du$ and $\int C_T(u)\,du$, are free from measurement errors, then all the errors in the LGA variables, $\frac{\int C_T(u)\,du}{C_T(t)}$, and $\frac{\int C_p(u)\,du}{C_T(t)}$, in Eq. (2.21) are entirely due to $C_T(t)$. Accordingly, the errors in these two LGA variables will be perfectly correlated, i.e., $r_i = 1$. The assumption that the integral terms are noise-free is based on the fact that integrated measurements has less noise than the discrete data. To simplify the explanation, we assume this is the case. Given the above, lets us say the value of $\frac{\int C_P(u)\,du}{C_T(t)}$ at point $\boldsymbol{H}$ in Fig. 3.2 has an error of ξ due to $C_T(t)$ error, then, according to York in [13], the value of $\frac{\int C_T(u)\,du}{C_T(t)}$ will have an error of $\frac{\int C_T(u)\,du}{\int C_P(u)\,du}\xi$ due to that same $C_T(t)$ error. Now, the observed point $\boldsymbol{H}$ is shifted from the true point $\boldsymbol{H}^*$ along a line with a slope given by,

$$\begin{aligned}\frac{\Delta Y}{\Delta X} &= \frac{\frac{\int_0^t C_T(u)\,du}{\int_0^t C_P(u)\,du}\xi}{\xi} \\ &= \frac{\int_0^t C_T(u)\,du}{\int_0^t C_P(u)\,du}.\end{aligned} \tag{3.58}$$

In a less compact format than Eq. (3.55), the LSC residuals can be expressed as,

$$\begin{aligned}x_i - X_i &= \frac{W_i\,(a + bX_i - Y)\,(r_i\alpha_i - b\,w(Y_i))}{w(X_i)w(Y_i)} \\ y_i - Y_i &= \frac{W_i\,(a + bX_i - Y)\,(w(X_i) - br_i\alpha_i)}{w(X_i)w(Y_i)}.\end{aligned} \tag{3.59}$$

Using Eq. (3.59), the slope of the straight line connecting points $\boldsymbol{H}$ and $\boldsymbol{H}^*$ in Fig. 3.2 is given by,

$$\begin{aligned}d_{slope} &= \frac{Y_i - y_i}{X_i - x_i} \\ &= \frac{\text{y-residual}}{\text{x-residual}} \\ &= \frac{w(X_i) - br_i\alpha_i}{r_i\alpha_i - b\,w(Y_i)}.\end{aligned} \tag{3.60}$$

Under the assumption that the errors are perfectly correlated, we have, $r = 1$, and Eq. (3.60) becomes,

$$\begin{aligned} d_{slope} &= \frac{w(X_i) - b\sqrt{w(X_i)w(Y_i)}}{\sqrt{w(X_i)w(Y_i)} - b\,w(Y_i)} \\ &= \sqrt{\frac{w(X_i)}{w(Y_i)}}. \end{aligned} \tag{3.61}$$

From Eq. (3.58) we have that,

$$\frac{\text{error in } X \text{ measurement}}{\text{error in } Y \text{ measurement}} = \frac{\text{error in } \frac{\int_0^t C_T(t)\,du}{C(t)}}{\text{error in } \frac{\int_0^t C_P(t)\,du}{C(t)}} = \frac{\Delta Y}{\Delta X} = \frac{\int_0^t C_T(t)\,du}{\int_0^t C_P(t)\,du}, \tag{3.62}$$

and by using the definition of the weights as the inverse of the squared errors, we get that,

$$\sqrt{\frac{w(X_i)}{w(Y_i)}} = \left.\frac{1/\left(\text{error in } \frac{\int_0^t C_P(t)\,du}{C(t)}\right)}{1/\left(\text{error in } \frac{\int_0^t C_T(t)\,du}{C(t)}\right)}\right]_i = \left.\frac{\text{error in } \frac{\int_0^t C_T(t)\,du}{C(t)}}{\text{error in } \frac{\int_0^t C_P(t)\,du}{C(t)}}\right]_i = \left.\frac{\int_0^t C_T(t)\,du}{\int_0^t C_P(t)\,du}\right]_i, \tag{3.63}$$

i.e.,

$$\sqrt{\frac{w(X_i)}{w(Y_i)}} = \left.\frac{\int_0^t C_T(t)\,du}{\int_0^t C_P(t)\,du}\right]_i. \tag{3.64}$$

Equation (3.61) serves to inform that the residuals are not necessarily, vertical, horizontal or perpendicular to the estimated straight line. They can be at any angle which is defined by their slope. We see in Eq. (3.61) that the slope of the residuals is dependent on the weight functions. This means that the weights will serve to determine the direction in which the observed measurements should be adjusted.

To put it into perspective, for the unweighted OLS, $w(Y_i) = 1$ and $w(X_i) = 0$, and therefore $d_{slope} = \inf$. This means that the residuals of the OLS regression are in the vertical direction, which is exactly the case for OLS, minimizing only the errors in the Y measurements.

Equation (3.64) shows how the weights —when assumed to be the inverse of measurement errors— relate to the integral terms in the LGA variables. Given that the errors in measured data are practically unknown, the aim is to define the weight functions such that they satisfy the relationship in Eq. (3.64).

The weights $[w(X_i), w(Y_i)]$ and the correlation of the errors r_i

The variances in the measurements are not known, and therefore the weights cannot be directly calculated by,

$$w(X_i) = \frac{1}{s^2(X_i)} \qquad \text{and} \qquad w(Y_i) = \frac{1}{s^2(Y_i)}. \tag{3.65}$$

This study adopted the method of "relative error regression" [48, 49]. In relative error regression, instead of minimizing the squared-sum of the absolute errors, $\sum(y_i - Y_i)$, the squared-sum of the relative errors, $\sum\left(y_i - Y_i)/Y_i^2\right)$, is minimized. In statistical prediction, relative errors are considered to be more informative than absolute errors [49]. Expressing the relative errors in this manner can help adjust measurements in a weighted linear regression with the weights given by, $1/Y_i^2$ [48, 49]. After adopting this technique for LSC, the X and Y weights can be respectively given as,

$$w(X_i) = \frac{1}{X_i^2} \qquad \text{and} \qquad w(Y_i) = \frac{1}{Y_i^2}. \tag{3.66}$$

These weight functions suffices for the LSC-based LGA since they satisfy Eq. (3.64) as below,

$$\begin{aligned} \sqrt{\frac{w(X_i)}{w(Y_i)}} &= \sqrt{\frac{1/X_i^2}{1/Y_i^2}} = \frac{Y_i}{X_i} \\ &= \frac{\int C_T(t_i)\, du/C_T(t_i)}{\int C_P(t_i)\, du/C_T(t_i)} \\ &= \frac{\int C_T(t_i)\, du}{\int C_P(t_i)\, du}. \end{aligned} \tag{3.67}$$

The correlation of the errors were estimated by the correlations between the variables themselves. These correlations were estimated by the correlations of determination. The correlation of determination is a number between 0 and 1, and it determines the proportion of variation in the response variable that can be explained by the predictor variable [50, 51]. To iterate, the slope obtained in the first run was substituted in W_i and β_i. The iteration process was repeated until the relative differences between consecutive estimates of the slope, $(b_j - b_{j+1})/b_{j+1}$, were less than 10^{-10}.

In summary, as per [12], here is how one would carry out the estimation of the regression parameters by LSC;

1. Begin by obtaining an approximate initial value for b. In this study, it was obtained by OLS.

2. Approximate the weights, $w(X_i)$ and $w(Y_i)$, for each point, and the correlation between the errors in measurements, r_i. In this study, the weights were obtained by Eq. (3.66), and r_i was estimated by the correlation between the measurements themselves.

3. Using the weights, the initial approximate for b and the correlation, r_i, evaluate W_i for each point.

4. Calculate $\bar{X}$ and $\bar{Y}$ by Eq. (3.40), calculate U_i and V_i by Eq. (3.41), and then calculate β_i by Eq. (3.52), for each point.

5. Using W_i, U_i, V_i, and β_i, calculate b by Eq. (3.51).

6. Now use the new b in step 5 as the initial estimate and repeat steps, 3, 4, and 5 until consecutive estimates of b converges to some desired condition.

7. With this final value of b, together with the final $\bar{X}$ and $\bar{Y}$, calculate a by Eq. (3.39).

8. For each (X_i, Y_i), calculate the adjusted/estimated points, x_i and y_i, by Eq. (3.54).

9. Use the estimated points, (x_i, y_i), and W_i to calculate $(\bar{x}, \bar{y})$ by Eq. (3.56), and then calculate (u_i, v_i) by Eq. (3.57).

10. Using W_i, $\bar{x}$ and u_i , calculate σ_b, and then σ_a by Eq. (3.53).

3.1.1 Simulation studies

Simulation data design

In order to assess the performance of LSC against that of MRTM2 and conventional LGA when estimating BP_{ND}, a set of PET data of a two-tissue compartmental model was simulated using a clinically measured pTAC. The simulated data mimicked a well-known radiotracer, ^{11}C-carfentanil (CFN), which is a reversible radiotracer that binds to μ-opioid in the brain. The kinetic parameters, K_1, $k_{2\text{–}4}$, and the non-displaceable distribution volume (V_{ND}) used for simulations were adopted from a range of values used in a previous study [52];

- 11 values of BP_{ND} were set in the range of [0.0, 3.0], covering a range of [0.0, 0.35] for k_3.

- $[K_1 \;\; k_4] = [0.1835\ \mathrm{mL\,cm^{-1}\,min^{-1}} \quad 0.115/\mathrm{min}]$.
- $[k_2 \;\; k_3] = [K_1/V_{ND} \quad BP_{ND} \cdot k_4]$.
- $V_{ND} \quad = 1.59\ \mathrm{mL\,cm^{-1}}$.

Simulated data were made for a 90 min, 33-frame scan (6×0.1667 min, 3×0.5 min, 5×1 min, 5×2.5 min, and 14×5 min). The reference region was formed with a zero noise level, zero BP_{ND}, and delivery and clearance rates equal to those of the target tissues, due to its physiological assumptions. Using these kinetic parameter values, 11 noise-free tTACs corresponding to the 11 BP_{ND} values were formed. Statistical noise was added to these noise-free tTACs, and 1024 noisy tTACs were simulated as a slice of 32-by-32 pixels for each noise-free tTAC. The BP_{ND} values were then re-estimated from the noisy tTACs by LSC, MRTM2, and conventional LGA, and the results were compared. The true value for the dissociation rate of the reference region (k_2^R), which was used to produce the simulation data, was used for all methods.

The noise was assumed to have a zero-mean Gaussian distribution with a variance proportional to the true tTAC; this is the same to the noise model used in previous studies [6, 8, 16, 11]. The noise scaling factor was set such that the actual magnitude of the noise level in the simulated tTACs varied within a range of 0 to 30%; this range covered the actual noise observed in voxel-based PET data. The magnitudes of the noise level were calculated by the percentage of the ratios of the "standard deviation of the noisy tTACs" to the "mean of the noise-free tTAC" for the portion of the time range used for the BP_{ND} estimations.

Results

Figure 3.3 shows the simulated noise-free (thick lines) and their corresponding noisy (thin lines) tTACs for the 11 BP_{ND} values. The 12th blue curve at the bottom is the curve of the reference region. Out of the the 1024 simulated noisy tTACs, only one is shown in Fig. 3.3 for each noise-free tTAC.

The results in Fig. 3.4 compares the percentage bias (difference from 100%) in the BP_{ND} values estimated by three methods, LSC, MRTM2 and OLS, from the noisy tTACs. The standard deviations (in percentages) are shown as error bars. In these results, LSC estimates are the least biased along the entire range of BP_{ND}, with up to a maximum of 12% and 28% bias

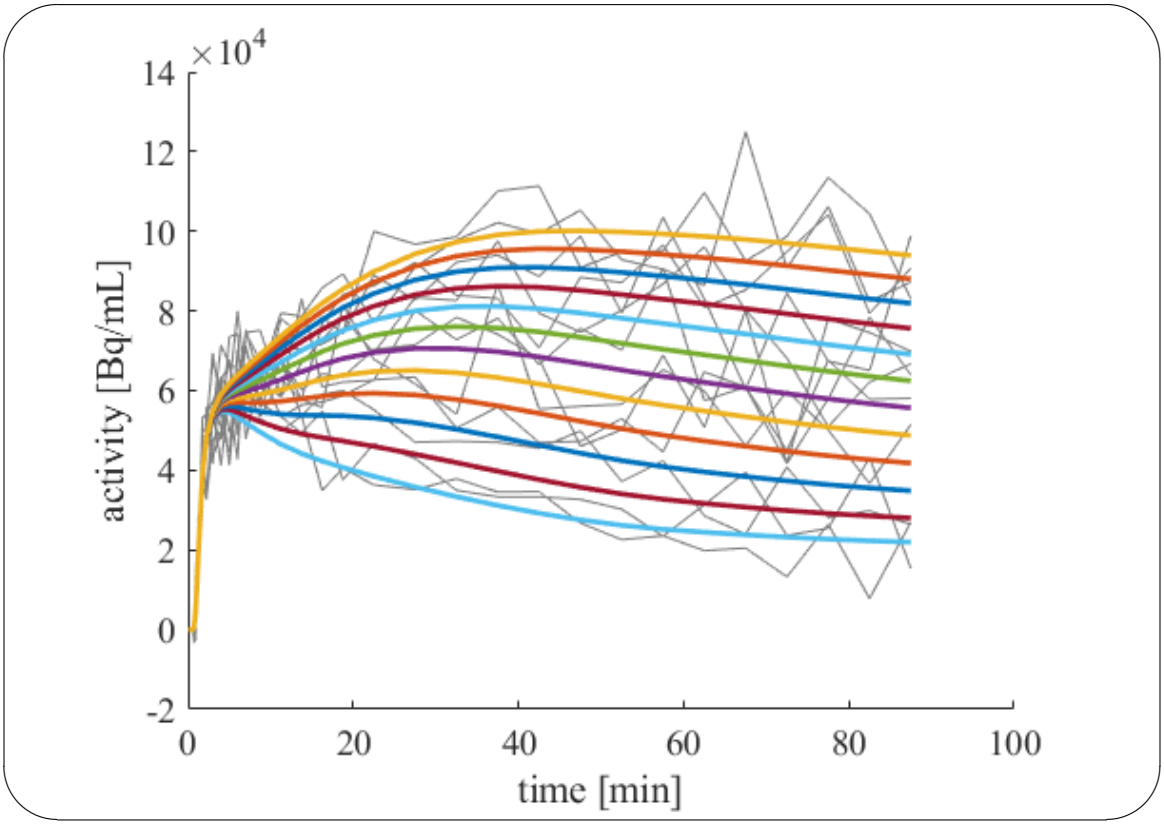


Figure 3.3: *Simulated tTACs. The 12 thick curves denote the noise-free tTACs. Eleven of these correspond to the 11 BP_{ND} values. The remaining one (lowermost) represents the reference region. The thin curves denote the noisy tTACs. Only one noisy tTAC is shown out of the 1024 simulated noisy tTACs for each of the noise-free tTACs.*

difference in comparison with MRTM2 and conventional OLS-based LGA, respectively. The error bars show that while the conventional LGA estimates have the smallest variability, LSC estimates have smaller variability than MRTM2 estimates. With the above, it can be seen that LSC outperforms MRTM2 in terms of both bias and variance, i.e., LSC estimates have both smaller bias and variance than MRTM2 estimates.

Figure 3.5 shows the distributions of the errors obtained from LSC- and OLS-based LGAs. The errors for OLS refers to the y-residuals. The errors for LSC are calculated as a sum of Eqs. (3.27–3.31), which makes the condition for the minimization term to equate to zero. The histograms in Fig. 3.5 are for residuals of only 8 random of the simulated tTACs. In this figure, it can be seen that LSC estimates make a better normal distribution —both in terms of the

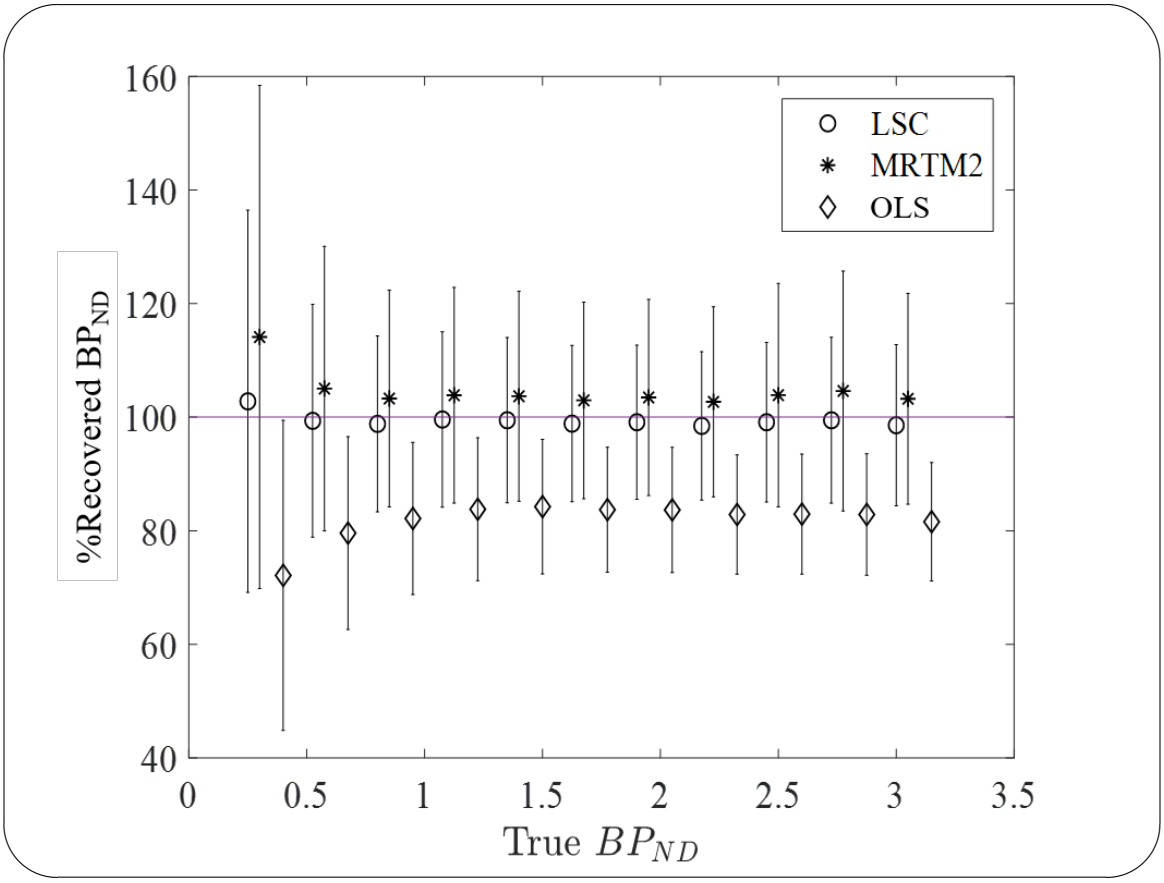


Figure 3.4: *Comparisons of the percentage bias and variability in the estimated BP_{ND} values. The percentage bias and variability in the BP_{ND} estimated from the noisy tTACs by LSC, MRTM2, and conventional LGA, are compared for the 11 values of true BP_{ND}.*

bell shape and in terms of narrowness— in comparison to those of OLS. Figure 3.6 shows the histograms of the errors of all 1024 simulated tTACs for $BP_{ND} = 3$ —a further confirmation of the results in Fig. 3.5.

3.1.2 Clinical ^{11}C-PiB PET data studies

Study design

Two ^{11}C-PiB PET data sets, of which one is Aβ-negative and the other is Aβ-positive, were used for this analysis. The two participants underwent a ^{11}C-PiB PET scan for 70 minutes with 25 frames (6 $\times$ 0.1667 min, 3 $\times$ 0.3333 min, 2 $\times$ 1 min, 2 $\times$ 3 min, and 12 $\times$ 5 min). The scan was carried out according to the standard ADNI ^{11}C-PiB PET procedure. The study protocol was approved by the Ethics Committee of Kindai University Hospital, and written informed

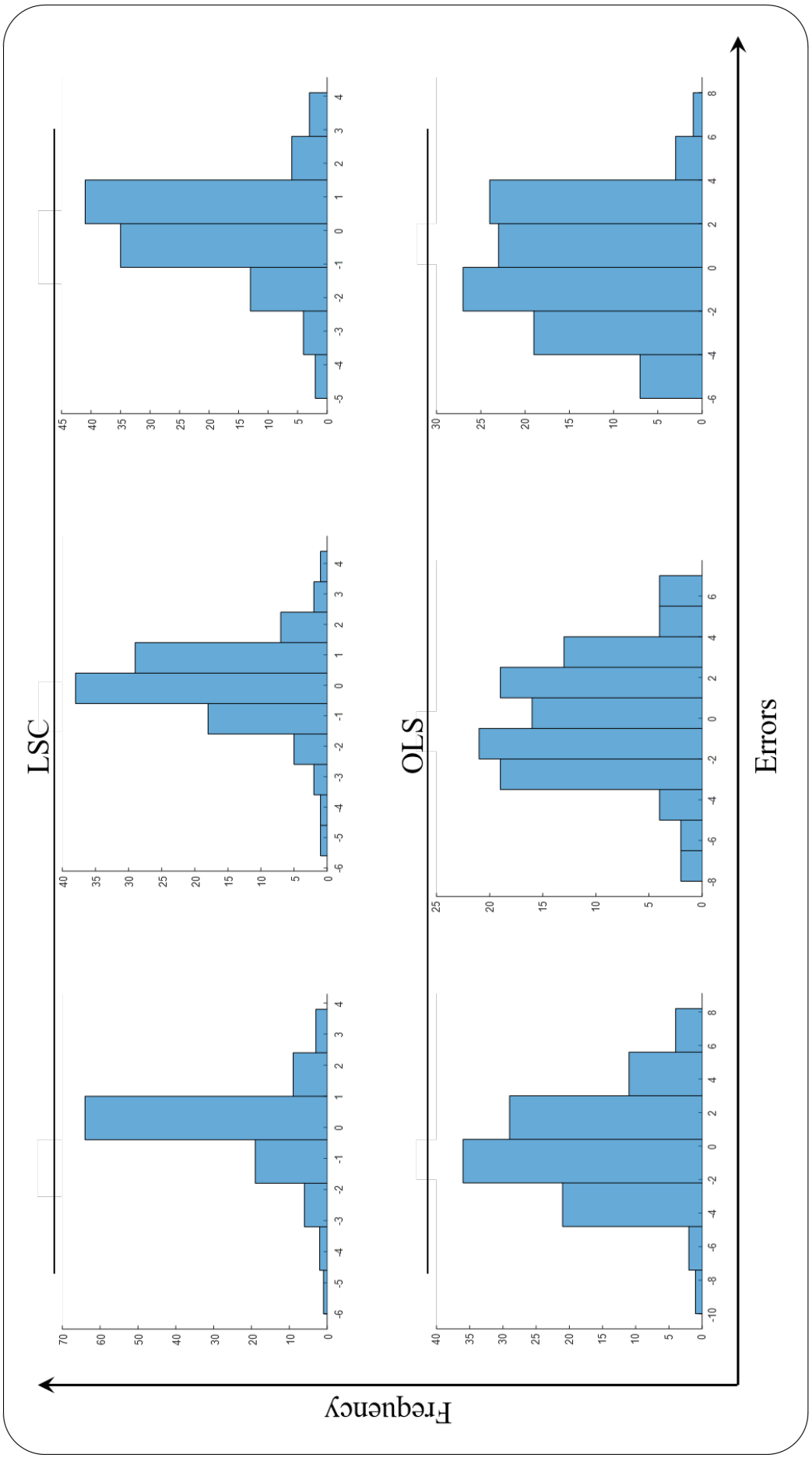


Figure 3.5: *Comparison of the distributions of LSC against OLS errors. The upper panel shows the distributions of the errors estimated by LSC, and the lower panel shows those of OLS.*

consent was obtained from the participants.

The data were collected in arrays of 128 voxels × 128 voxels × 47 slices, with voxel sizes of 2.1 × 2.1 × 3.4 mm^3. BP_{ND} parametric images – reflecting Aβ deposits in the brain regions – were then generated using three algorithms, LSC, MRTM2, and conventional OLS-based LGA, implemented in MATLAB (The MathWorks, Inc., Natick, MA, United States).

Time t^*(30 min.), which was used for both synthetic and actual data, was graphically determined from simulated data.

For the actual data, the values of k_2^R that were used for all methods were estimated by MA1. MA1 allows estimation of k_2^R with little bias [10]. Specifically for each slice, the average value, $\bar{k}_2^R$, calculated from the $\mid k_2^R \mid$ of all voxels in the reference region, was used.

Results

Figure 3.7 shows the BP_{ND} parametric images obtained by the three methods, LSC, MRTM2, and conventional OLS-based LGA. These images are generated from ^{11}C-PiB PET brain images, and therefore they reflect Aβ deposits in the brain regions. The left panel shows images of three slices of the Aβ-negative patient, obtained by each of the three methods, whereas the right panel shows the images of the Aβ-positive patient.

In Fig. 3.7, the images obtained by LSC have significantly higher total BP_{ND} estimates than those of conventional OLS-based LGA for both patients. The images obtained by MRTM2 have the highest total BP_{ND} estimates. For MRTM2 images, however, considerably higher levels of noise are observed. These results are consistent with those of CFN observed in the simulated data.

Table 3.1 shows the average computational time taken by each method for the corresponding sets of three images in Fig. 3.7. LSC took moderate computational time, one order of magnitude higher than conventional OLS-based LGA, for both patients' images. This was expected of LSC because it is an iterative procedure. MRTM2 took the longest computational time amongst the three methods, two orders of magnitude higher than conventional OLS-based LGA, for both

Table 3.1: *Means (± standard deviations) of computational times taken by each method for the respective slice images in Fig. 3.7.*

Human data computational time (seconds)		
	L	R
LSC	0.161 ± 0.033	0.156 ± 0.016
MRTM2	1.045 ± 0.031	1.014 ± 0.031
LGA	0.073 ± 0.024	0.083 ± 0.009

The two columns indicated by L and R denote the left and right panel images, respectively.

patients' images.

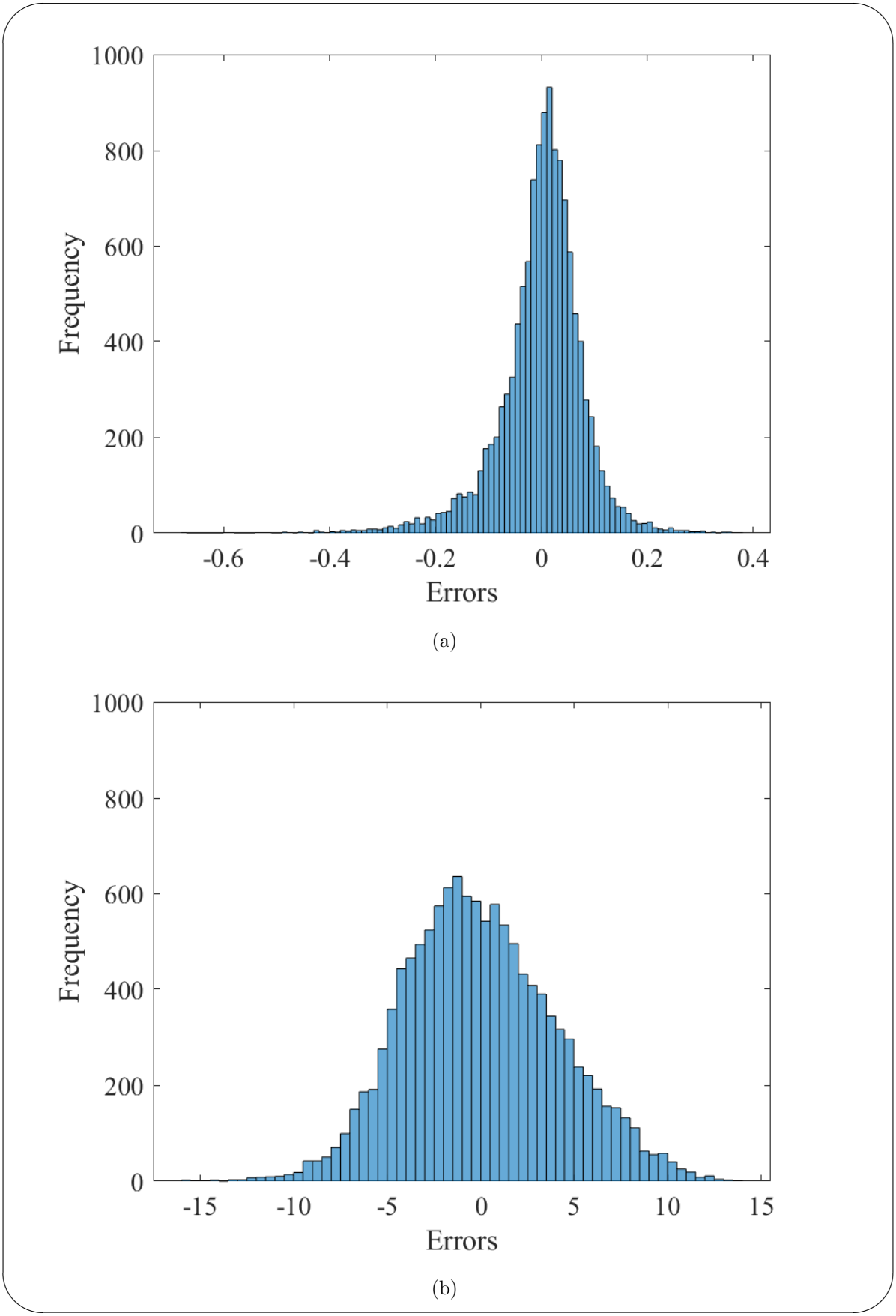


Figure 3.6: *Comparison of the distributions of LSC errors (a) against OLS errors (b) for all 1024 simulated noisy tTACs for* $BP_{ND} = 3$.

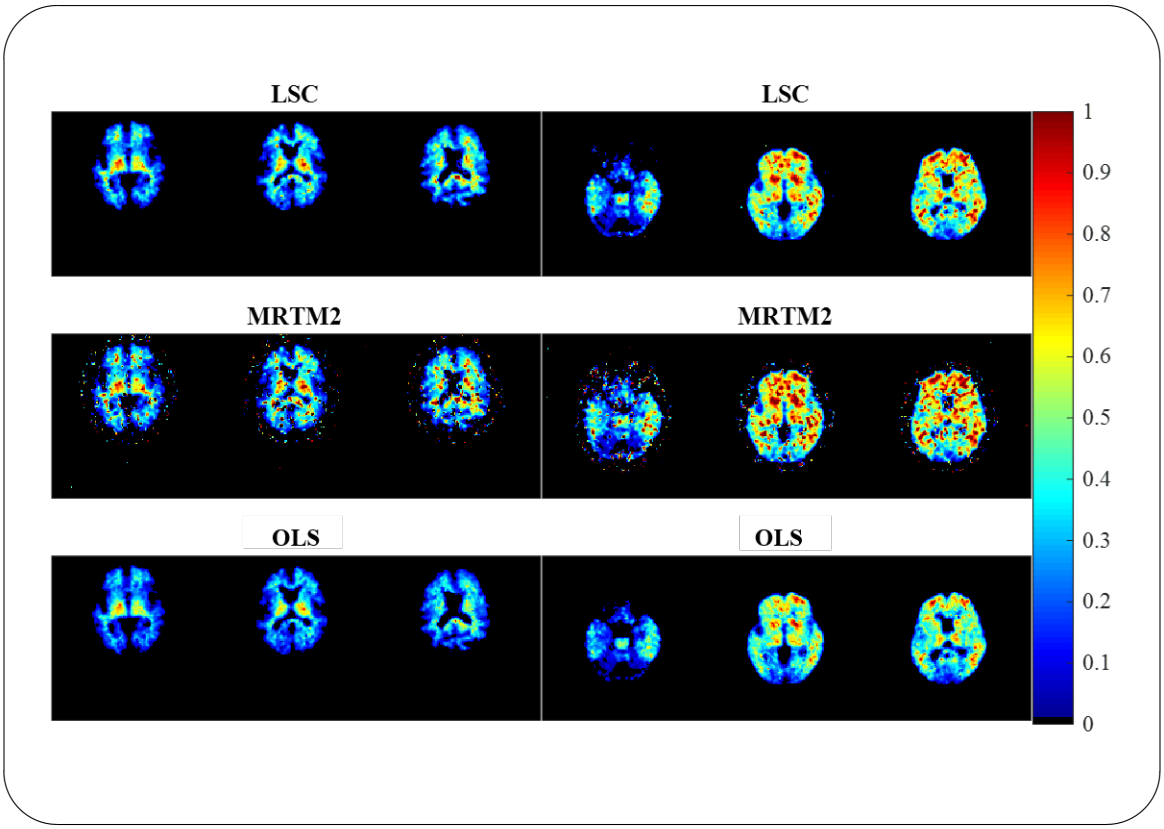


Figure 3.7: *Axial slice images of brain regions obtained by three methods, LSC, MRTM2, and conventional OLS-based LGA. These are BP_{ND} images of Aβ quantification of ^{11}C-PiB brain PET data of the two patients. The left panel shows the images of the Aβ-negative patient, and the right panel shows the images of the Aβ-positive patient.*

3.2 Discussion

The results in the above analysis shows that LSC estimates of the BP_{ND} emerged to be less biased than those of MRTM2 and conventional OLS-based LGA. In addition, it was also seen that LSC estimates have smaller variations in comparison with those for MRTM2, providing a better trade-off between bias and variability.

The underestimation which is observed in the simulated data for conventional OLS-based LGA estimates is also reflected in the real data, and accordingly, the reduction of the underestimation seen in the simulated data for the LSC method is also reflected in the real data. Furthermore, the underestimations observed for conventional OLS-based LGA method in both simulation and human data studies are in consistency with previous studies [33, 52], and this strengthens the reliability of the results reported here.

Both LSC and OLS are based on the assumption that the errors in the measurements are normally distributed. However, this assumption may not always be satisfied. A visual inspection of the distributions of the LSC error values and the OLS residuals —in Figs. 3.5 and 3.6— showed that the LSC error values have a better normal distribution compared with that of the OLS residuals. This can be attributed to the fact that LSC takes into account the varying magnitude of errors at each data point as demonstrated in [14].

LSC considers the errors in both variables without altering the original LGA equation, thus maintaining its simplicity and inducing less variability. In addition, LSC does not only consider the errors associated with term $C_T(t)$, but also the errors associated with each LGA variable as a whole. The effort to include more details in LSC has increased the number of terms in Eq. (3.1), compared with the fundamental OLS regression in Eq. (2.49), which only comprises the residuals of the response variable. This could be a factor in the slight increase in the variability of LSC-based estimates compared to conventional OLS-based LGA estimates.

MRTM2 rearranges the conventional LGA equation into a multilinear regression form so that the noted noisy term, $C_T(t)$, only appears in the response variable [9, 10]. This only considers the errors associated with $C_T(t)$. However, the rest of the terms in the LGA equation are not entirely noise-free; they are less noisy compared with $C_T(t)$. Not considering the noise in the

rest of the terms could lead to the bias in the MRTM2 estimates.

Based on the computational times of the three methods for obtaining slice images, the conventional LGA has the highest computational efficiency, followed by LSC, and then MRTM2. The computation time for LSC would depend on the number of iterations, which in turn depends on the iteration condition, with the computation time increasing with the number of iterations. All LSC images in Fig. 3.7 required at most three iterations. The iteration limiting condition in this study, $(\alpha'_i - \alpha'_{i+1})/\alpha'_{i+1} < 10^{-10}$, was arbitrarily chosen. However, it was found that by choosing a stricter condition, say 10^{-11}, the number of iterations did not increase.

Although the weighting formulas in Eq. (3.66) are inverse squares of the measurements, they will not cause the usual problem of overproportional weighting of smaller measurements. This is because they are not the direct weighting coefficients. The final weighting coefficients are appearing in Eq. (3.51) as $W_i B_i$. Because they appear in both the numerator and denominator, the overprortional weighting effect can be canceled out.

The proposed method, LSC-based LGA, significantly reduced the bias in the estimates of BP_{ND}, up to 12 and 28% difference compared with MRTM2 and conventional OLS-based LGA, respectively. Considering the conventional OLS-based LGA as the standard method, LSC reduced the bias at a slight expense of increased variation. However, LSC caused smaller variations and required shorter computation times than MRTM2. Thus, LSC outperformed MRTM2 in all aspects (bias, variability, and computational efficiency).

Chapter 4

Tissue time-activity curves denoising techniques for improving image contrast

4.1 Principal component analysis

PCA is a dimensionality reduction technique that is mostly used for feature extraction and noise filtering. The fundamental idea of PCA is to represent the data in a reduced dimension while preserving as much variance as possible, i.e., reducing a large data set of variables to a smaller set that retains most of the important information in the original data set. PCA transforms variables into new sets of variables that are linear functions of the original variables [15, 53, 54]. The new set of variables are uncorrelated and are defined by a set of orthogonal basis vectors and PCs that optimally describe the variance in the data. That is, the first PC describes the largest variance in the data. The second PC is orthogonal to the first PC and describes the largest of the remaining variance. It goes on until the last PC – orthogonal to all earlier PCs – that describes the least variance than all other PCs. The orthogonal basis vectors and the corresponding variance are found by solving for the eigenvalues and eigenvectors of the covariance matrix of the data. The eigenvectors are the orthogonal basis vectors, and the eigenvalues are the corresponding variances.

In dimension reduction, one keeps only the projections along the dimensions of the first few largest PCs. This gives a reduced dimensional projection that maintains most of the important

information of the original variables, allowing for a simpler interpretation.

In noise filtering application, one attempts to recover the original variables using the lower-dimensional projections. This results in denoised estimates of the original variables. The idea is that the larger PCs correspond to the dimensions with most of the signals of interest and less noise. On the other hand, smaller PCs correspond to the dimensions with much of the unwanted noise and less of the signal of interest. Thus, reconstructing the data using only the few largest PCs, preserves the signals of interest whilst throwing away the unwanted noise signals.

Let us denote the PET brain volume data $\mathbf{C}$ to be made up of R tTACs with p frames. Then we can write,

$$\begin{aligned}\mathbf{C} &= \begin{bmatrix} \mathrm{C}_1(t_1) & \mathrm{C}_1(t_2) & \dots & \mathrm{C}_1(t_p) \\ \mathrm{C}_2(t_1) & \mathrm{C}_2(t_2) & \dots & \mathrm{C}_2(t_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{C}_R(t_1) & \mathrm{C}_R(t_2) & \dots & \mathrm{C}_R(t_p) \end{bmatrix}_{\mathbf{R\times p}} \\ &= \begin{bmatrix} \mathrm{C}(t_1) & \mathrm{C}(t_2) & \dots & \mathrm{C}(t_p) \end{bmatrix}\end{aligned} \tag{4.1}$$

Applying PCA to the data matrix $\mathbf{C}$, it finds a set of $p \times p$ vectors, $\mathbf{V} \in \mathbb{R}^{p\times p}$. These vectors are simply the eigenvectors of the covariance matrix of the mean-centered data, $\mathbf{C_0}$. Using only the first k ($k \leq p$) eigenvectors, a new set of denoised original variables can be estimated by,

$$\hat{\mathbf{A}} = (\mathbf{C_0 W})\mathbf{W}' + \bar{\mathbf{C}}, \tag{4.2}$$

where, $\mathbf{W} = \mathbf{V}(:, 1:k) \in \mathbb{R}^{p\times k}$, is the set of the first k eigenvectors. $\bar{\mathbf{C}} \in \mathbb{R}^{1\times p}$ is the mean vector of $\mathbf{C}$. The eigenvectors define the directions of the PCs, $\mathbf{Z} \in \mathbb{R}^{R\times p}$, which can be expressed as,

$$\mathbf{Z} = (\mathbf{C_0 W}). \tag{4.3}$$

Therefore, using the first few k eigenvectors can also be referred to using the first few k PCs.

In this study, the number of PCs, k, that were retained to estimate the denoised data were such that specified amount of percentage variance could be realized from them. Specifically, the two methods LSC–PCA and OLS–PCA were evaluated at two levels of percentage variance, 95% and 97%.

Mathematics of PCA

Given our data matrix $\mathbf{C} \in \mathbb{R}^{R \times p}$, the aim is to find a linear combination of the columns of matrix $\mathbf{C}$ with maximum variance [15]. Let us denote these linear combinations by $\mathbf{Z}$, i.e., the first PC Z_1 is defined as,

$$\begin{aligned} Z_1 &= \mathbf{C}a_1 \\ &= \sum_{i=1}^{p} a_{i1}\mathrm{C}(t_i), \end{aligned} \tag{4.4}$$

where a_1 is a vector of constants coefficients, $\begin{bmatrix} a_{11} & a_{21} & \dots & a_{p1} \end{bmatrix} \in \mathbb{R}^{1 \times p}$.

In general, the k^{th} PC can be expressed as,

$$\begin{aligned} Z_k &= \mathbf{C}a_k \\ &= \sum_{i=1}^{p} a_{ik}\mathrm{C}(t_i). \end{aligned} \tag{4.5}$$

Equation (4.4) is subject to "var(Z_1) is maximum", and therefore Eq. (4.5) is subject to "var(Z_k) is maximum". Further, the PCs are such that they are uncorrelated with each other (i.e., cov$(Z_k, Z_l) = 0 \mid k > l \geq 1$), and a_k has a unit length (i.e., $a_k^T a_k = 1$) [53, 15].

The variance of the first PC is given by,

$$\begin{aligned} \mathrm{var}(Z_1) &= \langle Z_1^2 \rangle - \langle Z_1 \rangle^2 \\ &= \sum_{i,j=1}^{p} a_{i1}a_{j1}\langle \mathrm{C}(t_i)\mathrm{C}(t_j) \rangle - \sum_{i,j=1}^{p} a_{i1}a_{j1}\langle \mathrm{C}(t_i) \rangle \langle \mathrm{C}(t_j) \rangle \\ &= \sum_{i,j=1}^{p} a_{i1}a_{j1}S_{ij} \\ &= a_1^T \mathbf{S} a_1, \end{aligned} \tag{4.6}$$

where $\mathbf{S}$ is the sample covariance matrix, and S_{ij} is the covariance of $\mathrm{C}(t_i)$ and $\mathrm{C}(t_j)$, respectively given by,

$$\mathbf{S} = \frac{1}{R-1}\mathbf{C_0^T C_0} \tag{4.7}$$

and

$$S_{ij} = \langle \mathrm{C}(t_i)\mathrm{C}(t_j)\rangle - \langle \mathrm{C}(t_i)\rangle\langle \mathrm{C}(t_j)\rangle = \sigma_{\mathrm{C}(t_i)\mathrm{C}(t_j)}. \tag{4.8}$$

Identifying the linear combination with maximum variance, var(Z_1), is equivalent to maximizing the equation below [15],

$$a_1^T \mathbf{S} a_1 - \lambda_1(a_1^T a_1 - 1), \tag{4.9}$$

where λ_1 is a Lagrange multiplier. Differentiating Eq. (4.9) with respect to a_1 and equating to zero gives,

$$\begin{aligned} &\mathbf{S}a_1 - \lambda_1 a_1 = 0 \\ &(\mathbf{S} - \lambda_1 \mathbf{I_p}) a_1 = 0 \\ &\implies \mathbf{S}a_1 = \lambda_1 a_1, \end{aligned} \tag{4.10}$$

where $\mathbf{I_p}$ is a $\mathbf{p} \times \mathbf{p}$ identity matrix. Equation (4.10) implies that a_1 is a unit-norm eigenvector, whereas λ_1 is the corresponding eigenvalue of the covariance matrix $\mathbf{S}$ [15]. λ_1 is the largest eigenvalue of $\mathbf{S}$, and thus the first pricipal component Z_1 accounts for the largest of variances in the data [53].

The second vector of coefficients, $a_2 = \begin{bmatrix} a_{12} & a_{22} & \dots & a_{p2} \end{bmatrix} \in \mathbb{R}^{\mathbf{1}\times\mathbf{p}}$, is found by maximizing the variance, var(Z_2), of the of the second PC, Z_2 —under the constraints cov(Z_2, Z_1) $= 0$ and $a_2^T a_2 = 1$. The variance of Z_2, and the covariance of Z_1 and Z_2 are respectively given by,

$$\mathrm{var}(Z_2) = a_2^T \mathbf{S} a_2, \tag{4.11}$$

and

$$\begin{aligned} \mathrm{cov}(Z_1,\, Z_2) &= a_1^T \mathbf{S} a_2 \\ &= \lambda_1 a_1^T a_2 \qquad \text{(using Eq. (4.10)).} \end{aligned} \tag{4.12}$$

Maximizing var(Z_2) under the aforementioned constraints equates to [15],

$$a_2^T \mathbf{S} a_2 - \lambda_2(a_2^T a_2 - 1) - \phi a_2^T a_1, \tag{4.13}$$

where λ_2 and ϕ are Langrage multipliers. Differentiating Eq. (4.13) with respect to a_2 gives,

$$\mathbf{S}a_2 - \lambda_2 a_2 - \phi a_1^T a_1 = 0. \tag{4.14}$$

Multiplying through Eq. (4.14) on the left by a_1^T we obtain,

$$a_1^T \mathbf{S}a_2 - \lambda_2 a_1^T a_2 - \phi a_1^T a_1 = 0. \tag{4.15}$$

From Eq. (4.15) we notice that $a_1^T \mathbf{S}a_2 = \lambda_2 a_1^T a_2 = 0$ and $a_1^T a_1 = 1$, and thus $\phi = 0$. Equation (4.15) can then be written as,

$$\begin{gathered} \mathbf{S}a_2 - \lambda_2 a_2 = 0 \\ \implies (\mathbf{S} - \lambda_2 \mathbf{I_p}) a_2 = 0. \end{gathered} \tag{4.16}$$

Equation (4.16) implies that a_2 is also an eigenvector of the covariance matrix $\mathbf{S}$, whereas λ_2 is the corresponding eigenvalue. λ_2 is the second largest eigenvalue of $\mathbf{S}$, and thus the second pricipal component, Z_2, accounts for the second largest of variances in the data [53].

Generally, the variance of the k^{th} PC Z_k is given by,

$$\text{var}(Z_k) = a_k^T \mathbf{S} a_k, \tag{4.17}$$

and maximizing var(Z_k) under the aforementioned constraints, ($\{\text{cov}(Z_k, Z_l) = 0 \mid k > l \geq 1\}$, and $\{a_k^T a_k = 1 \mid k = 1, 2, \cdots, p\}$), we get the k^{th} eigenvector (a_k) and the corresponding eigenvalue (λ_k). In the same way as the eigenvectors and eigenvalues for the first and second PC, λ_k is the k^{th} largest eigenvalue of the covariance matrix $\mathbf{S}$, and the k^{th} PC, var(Z_k), accounts for the k^{th} largest fraction of the variance in the data [53].

In total, with our data matrix, $\mathbf{C} = \begin{bmatrix} \mathrm{C}(t_1) & \mathrm{C}(t_2) & \dots & \mathrm{C}(t_p) \end{bmatrix} \in R \times p$, we define the PCs as, $\mathbf{Z} = \begin{bmatrix} \mathrm{Z}_1 & \mathrm{Z}_2 & \dots & \mathrm{Z}_\mathrm{p} \end{bmatrix} \in R \times p$, such that,

$$\mathbf{Z} = \mathbf{CV}, \tag{4.18}$$

where,

$$\mathbf{V} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{p1} \\ a_{12} & a_{22} & \dots & a_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \dots & a_{pp} \end{bmatrix}_{p \times p}, \tag{4.19}$$

is an orthogonal matrix, whose columns are the eigenvectors of the of $\mathbf{S}$.

The original matrix can be re-estimated from the PCs by,

$$\begin{aligned} \hat{\mathbf{C}} &= \mathbf{Z}\mathbf{V}' \\ &= (\mathbf{C}\mathbf{V})\mathbf{V}' \end{aligned} \tag{4.20}$$

If the PCs are calculated using the mean-centred data, $\mathbf{C_0}$, (i.e., $\mathbf{Z} = \mathbf{C_0}\mathbf{V}$), then Eq. (4.20) becomes,

$$\hat{\mathbf{C}} = (\mathbf{C_0}\mathbf{V})\mathbf{V}' + \bar{C}. \tag{4.21}$$

Employing only the first few eigenvectors, $\mathbf{W} = \mathbf{V}(:, 1:k)$, Eq. (4.20) then matches Eq. (4.2) as below,

$$\begin{aligned} \hat{\mathbf{C}} &= (\mathbf{C_0}\mathbf{W})\mathbf{W}' + \bar{C} \\ &= \hat{\mathbf{A}}, \end{aligned} \tag{4.22}$$

and this gives the denoised lower dimensional estimates of the original data set $\mathbf{C}$.

4.1.1 Simulation studies

Simulation data design

In addition to the CFN radiotracer simulations used in section 3.1.1, a simulation of ^{11}C-PiB was also performed in this section. ^{11}C-PiB is a reversible Aβ-binding radiotracer, and it is the radiotracer of the clinical data in both Chapter 3 and the current Chapter (Chapter 4). Therefore, for simulation studies, the methods are evaluated for two radiotracers, CFN and ^{11}C-PiB, in this Chapter. In this Chapter, the methods are compared in terms of DVR, instead of BP_{ND}. It is noted earlier in Eq. (2.41) that BP_{ND} is related to DVR by $BP_{ND} = DVR - 1$.

The set of kinetic parameters — exchange rate constants K_1, $k_{2\text{–}4}$ and the non-displaceable

distribution volume (V_{ND}) — below was used to make the first simulation data set (CFN radiotracer). These kinetic parameters were adopted from a previous study [52]. These are also the same range of parameters adopted by other studies such as [16].

- 11 values of DVR were set in the range of [0.0, 4.0], covering a range of [0.0, 0.35] for k_3.
- $[K_1 \;\; k_4] = [0.1835\ \text{mL}\,\text{cm}^{-1}\,\text{min}^{-1} \quad 0.115\ \text{min}^{-1}]$.
- $[k_2 \;\; k_3] = [K_1/V_{ND} \quad (DVR-1)\cdot k_4]$.
- $V_{ND} \quad = 1.59\ \text{mL}\,\text{cm}^{-1}$.

The above parameters are similar to the ones in Chapter 3, except for that now there is DVR instead of BP_{ND}. Simulated data were made for a 60 min, 27-frame scan (6×0.1667 min, 3×0.5 min, 5×1 min, 5×2.5 min, and 8×5 min). Two-tissue compartment PET data were simulated using a clinically measured plasma time-activity curve. The reference region was formed with a minimal noise level, and $DVR = 1$. As per the physiological assumptions of the reference region, the delivery (K_1) and clearance (k_2) rates were set identical to those of the target tissues.

For ^{11}C-PiB simulations, the kinetic parameters, K_1, $k_{2\text{–}4}$ and V_{ND}, below were used to simulate two-tissue compartmental model PET data. These parameters covers a range of parameters in one of the pioneering studies of ^{11}C-PiB [26].

- 11 values of DVR were set in the range of [0.0, 4.0], covering a range of [0.0, 0.045] for k_3.
- $[K_1 \;\; k_2] = [0.262\ \text{mL}\,\text{cm}^{-1}\,\text{min}^{-1} \quad 0.121\ \text{min}^{-1}]$.
- $[k_3 \;\; k_4] = [(DVR-1)\cdot k_4 \quad 0.015]$.
- $V_{ND} \quad = K_1/k_2$.

The scan period and frame for ^{11}C-PiB simulation was made similar to that of clinical ^{11}C-PiB PET images used in this study. These were 70 min, 25-frame scan (6×0.1667 min, 3×0.3333 min, 2×1 min, 2×3 min, and 12×5 min). The reference region was formed in a similar manner as for CFN; minimal noise level; $DVR = 1$; and K_1 and k_2 identical to those of the target tissues.

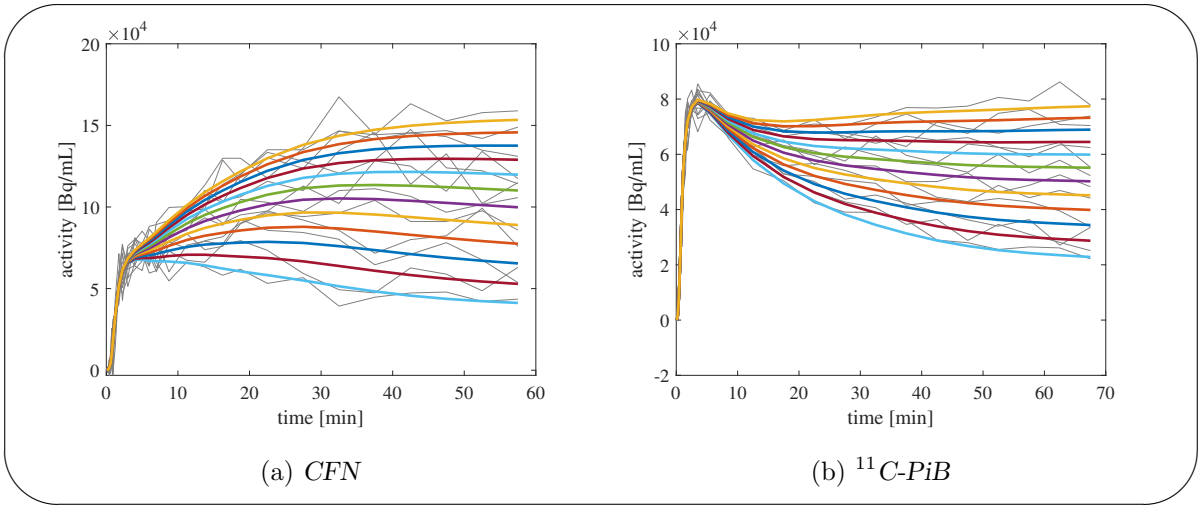
For each of the two radiotracers, a noise-free tTAC was formed for each DVR value. Statistical noise was added to the noise-free tTACs to form noisy tTACs. For each noise-free tTAC, 1024 noisy tTACs were simulated as a slice of 32×32 voxels. Four methods, LSC–PCA, OLS–PCA, LSC and conventional OLS-based LGA, were then used to estimate the DVR values from the noisy tTACs, and the results were compared. The true value for the dissociation rate of the reference region (k_2^R) used to construct the simulation data was used for all methods. The noise model used here is the same as descibed in section 3.1.1.

Results

Figure 4.1 presents the simulated tTACs, mimicking the two radiotracers, 4.1(a) CFN and 4.1(b) ^{11}C-PiB. Noise-free tTACs are represented by the thick curves, whereas the thin curves represent the noisy tTACs. For each noise-free curve, out of the 1024 simulated noisy curves, only one is shown in each of the subfigures in Fig. 4.1. In both Figs. 4.1(a) and 4.1(b), the lowermost thick curve denotes the reference region, and the upper 11 curves denote the 11 DVR values. DVR estimations from ^{11}C-PiB appeared to be sensitive to noise in comparison to CFN, and therefore, ^{11}C-PiB tTACs were simulated with lower noise levels than CFN, as it can be seen in Fig. 4.1. Lower radioactivity measurements in ^{11}C-PiB could be a contributing factor to the raised sensitivity to noise.

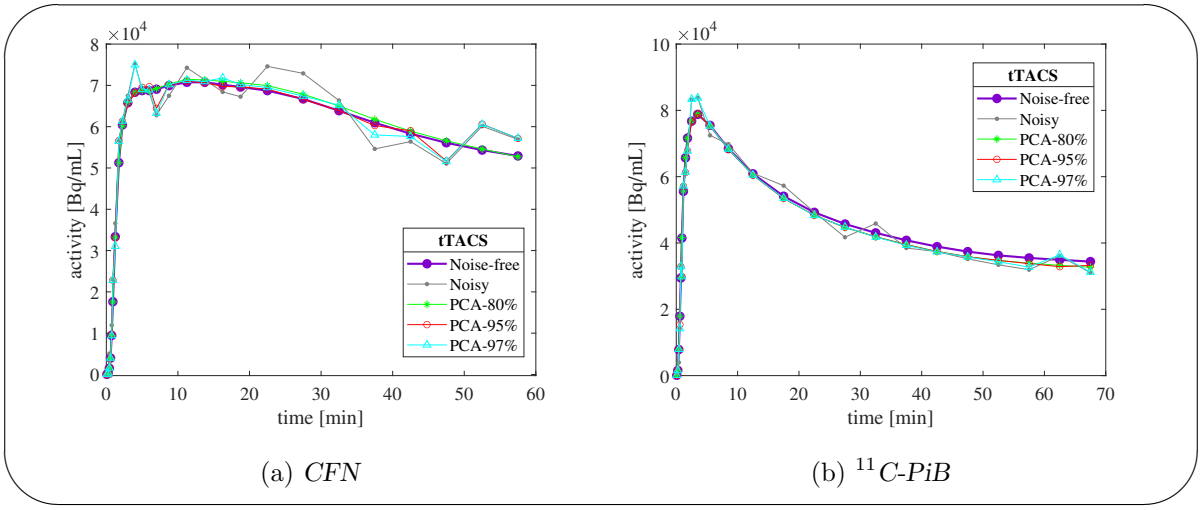
Figure 4.2 shows examples of noisy tTACs, and their corresponding PCA-denoised tTACs. Three denoised tTACs are shown for each radiotracer, and they represent PCA-denoising with a number of PCs that explains 80%, 95% and 97% of variance in the data, respectively. For CFN, the tTACs denoised with 95% and 97% variances are still noisy to a noticeable extent. As the number of PCs is lowered to explain only 80% variance, it can be seen that the tTACs are further denoised. One still needs to be careful not to lower the number of PCs to a greater extent. The tTACs would seem well-denoised, however spatial information will be lost and this cannot be inferred from the tTACs. For ^{11}C-PiB, the tTACs are denoised to a greater extent in comparison to CFN. This could be, in part, due to that they consists lower noise levels.

Figure 4.3 compares the percentage recovered DVR in the averages of the values estimated from the simulated noisy tTACs using four methods, LSC–PCA, LSC, OLS–PCA and OLS. In other words, the difference from 100% denotes the percentage bias. For PCA denoising, CFN



(a) *CFN* (b) ^{11}*C-PiB*

Figure 4.1: *Simulated tTACs for the radiotracers, (a) CFN and (b)* ^{11}*C-PiB. In each figure, the thick and thin curves represent the noise-free and noisy tTACs, respectively. The upper 11 thick curves correspond to the 11 DVR values, whereas the 12th and lowest curve correspond to the time-activity curve of the reference region.*



(a) *CFN* (b) ^{11}*C-PiB*

Figure 4.2: *Simulated noisy tTACs and their corresponding PCA-denoised tTACs for the two radiotacers, (a) CFN and (b)* ^{11}*C-PiB. Three denoised tTACs are shown for each radiotracer. The three denoised tTACs correspond to PCA with a number of PCs that explains 80%, 95% and 97% variance, respectively.*

tTACs were denoised with 95% and 97% variances for the results in Figs. 4.3(a) and 4.3(b), respectively. The same is true for ^{11}C-PiB in Figs. 4.3(c) and 4.3(d). As expected, LSC–PCA estimates have smaller error bars than LSC estimates. This is especially visible in the figures for ^{11}C-PiB, Figs. 4.3(c) and 4.3(d). This demonstrates PCA's effectiveness of reducing the variance. PCA's variance reduction can also be seen for OLS–PCA estimates. With regards to Figs. 4.3(c) and 4.3(d), it can be seen that OLS–PCA is greatly affected by the choice of the percentage variance that should be accounted for in PCA denoising. With only 2% difference of the explained variance, the difference in OLS–PCA estimates in Figs. 4.3(c) and 4.3(d) is clearly visible.

Table 4.1 shows some of the average values of the estimated *DVR* for CFN raditracer, with tTACs denoised with 95% variance PCs for PCA methods. The data in Tab. 4.1 correspond to some of the values plotted in Fig. 4.3(a). Table 4.1 also confirms reduced variance in both PCA-based methods, and also the difference in bias reduction of these two PCA-based methods, LSC–PCA and OLS–PCA, i.e., smaller bias for LSC–PCA compared to OLS–PCA.

Table 4.1: *Averege DVR values (and standard deviations in brackets) estimated from the noisy tTACs by the four methods, LSC–PCA, LSC, OLS–PCA and OLS, for CFN radiotracer. These data correspond to Fig. 4.3(a), in which PCA methods are denoised with 95% variance PCs.*

True *DVR*	**1.250**	**2.350**	**2.900**	**3.450**	**4.00**
			Estimated *DVR*		
LSC-PCA	1.249 (0.063)	2.356 (0.167)	2.895 (0.246)	3.456 (0.324)	3.992 (0.394)
LSC	1.249 (0.068)	2.344 (0.190)	2.881 (0.288)	3.452 (0.372)	4.022 (0.518)
OLS-PCA	1.212 (0.061)	2.233 (0.152)	2.719 (0.219)	3.212 (0.277)	3.656 (0.338)
OLS	1.211 (0.065)	2.212 (0.165)	2.683 (0.227)	3.169 (0.296)	3.616 (0.395)

4.1.2 Clinical ^{11}C-PiB PET data studies

Study design

A cohort of 12 (11 Aβ-negative and 1 Aβ-positive) subjects was included in this study. The subjects underwent 70 min ^{11}C-PiB PET scan of 25 frames (6 $\times$ 0.1667 min, 3 $\times$ 0.3333 min, 2 $\times$ 1 min, 2 $\times$ 3 min, and 12 $\times$ 5 min). The scan procedure was carried out according to the Alzheimer's Disease Neuroimaging Initiative (ADNI) ^{11}C-PiB standards. The study was approved by the Ethics Committee of Kindai University Hospital, and written informed consent was obtained from all participants. Obtained PET data were voxels of size 2.1 $\times$ 2.1 $\times$ 3.4 mm^3, in arrays of 128 voxels $\times$ 128 voxels $\times$ 47 slices. *DVR* parametric images were then generated from the PET data using the LSC–PCA, OLS–PCA, LSC, and OLS methods.

In addition to visual inspections, the parametric images were compared numerically in terms of the contrast between the four main gray matter cortices (frontal, temporal, occipital, and parietal) and white matter (corona radiata). The contrast index was calculated by the equation,

$$\mathcal{C} = \frac{\mu_W - \mu_G}{\sigma_W + \sigma_G}, \tag{4.23}$$

as in [55]. μ_G and μ_W denote the mean *DVR* of the gray and white matter regions, respectively, whereas σ_G and σ_W denote the standard deviations of *DVR* in the gray and white matter regions, respectively. All calculations were performed in MATLAB (The MathWorks, Inc., Natick, MA, USA).

Results

Figure 4.4 shows a comparison of the medians of contrast between gray matter cortices and white matter for *DVR* images obtained by the four methods (LSC–PCA, OLS–PCA, LSC and OLS), calculated for 11 Aβ-negative subjects. Statistical differences of these values are expressesd in Tab. 4.2. In the left figure, 4.4(a), the data were denoised with PCs that explained 95% variance for the PCA-based methods. The four gray matter regions show a similar trend, with the contrast of LSC–PCA and OLS–PCA images being the highest, as expected. As similar results were observed for LSC–PCA and OLS–PCA in the simulation results for ^{11}C-PiB data denoised with 95% of variance, here similar contrast is also observed between LSC–PCA and OLS–PCA. The figure on the right, 4.4(b), shows contrast values of a single subject, and for which the tTACs were denoised with 97% of variance for the PCA methods. Numerical values

for this figure are given in Tab. 4.3. Similar to simulation results, here it is observed again that OLS–PCA can perform poorly in comparison to LSC–PCA if more PCs are used. This serve to inform that, more precaution is required for OLS–PCA to balance between bias and variance reduction. On the other hand, for LSC–PCA, one needs only to consider throwing away a few PCs that could help with variance reduction, and not worry about the bias.

The Friedman Test Post-Hoc multiple comparison p-values of the difference between the contrasts of the images obtained using the four methods are shown in Tab. 4.2. The contrasts being compared in Tab. 4.2 are those used to plot Fig. 4.4(a). According to these statistical values, the contrasts of both LSC–PCA and OLS–PCA images are significantly different from those of LSC and OLS.

Table 4.2: *Friedman test and Post-Hoc multiple comparisons of the p-values of the contrast. Tabulated numerical values correspond to the p-values of the difference between the two methods in the corresponding rows and columns.*

Friedman Test (p value = 0.001)			
Post-Hoc Multiple Comparisons p-values			
	OLS-PCA	LSC	OLS
LSC–PCA	0.656	0.005	< 0.001
OLS–PCA		0.025	< 0.001
LSC			0.086

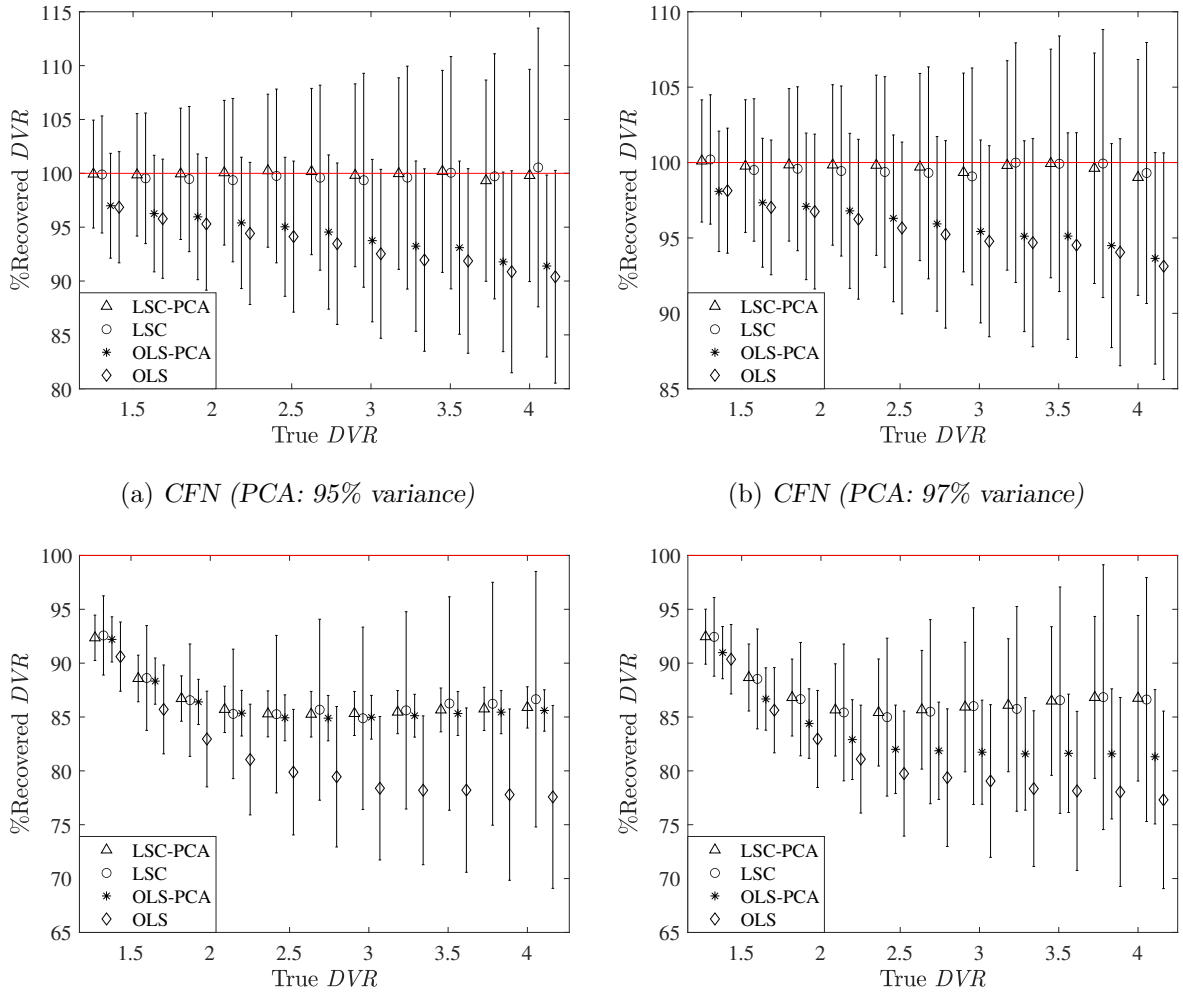
Figure 4.5 shows axial slices of the *DVR* parametric images obtained using the four methods. PCA-based methods were denoised with the number of PCs that explained 95% variance for these images. The two images in the upper row are from two Aβ-negative subjects, while the bottom row is from the Aβ-positive subject. In the figure, noise components in the images estimated by the LSC method can be seen appearing as random noisy pixels/regions with high *DVR* values throughout the slice (see the black arrows in Fig. 4.5). Images obtained by the LSC–PCA method exhibit a rather fair distribution of the *DVR* values with no sharp increases

Table 4.3: *Contrast values for a single Aβ-negative subject, which are plotted in Fig. 4.4(b).*

	Contrast			
	Frontal	Temporal	Occipital	Parietal
LSC-PCA	1.162	1.030	1.110	1.049
OLS-PCA	1.030	0.939	0.962	0.847
LSC	1.007	0.915	0.919	0.789
OLS	0.905	0.831	0.813	0.624

in the pixels/regions. The images obtained by OLS–PCA appear similar to those for LSC–PCA. This was expected since similar trends were observed between LSC–PCA and OLS–PCA for *DVR* estimates in simulation results and also for contrast values, for the tTACs denoised with 95% variance. The LSC images thus appear slightly thick (in comparison to the LSC–PCA images), which can hinder the contrast between the brain structures. For the OLS-based images, the lowest *DVR* estimates were observed as expected.

Figure 4.6 shows a single slice of *DVR* image for an Aβ-negative subject from which the contrasts in Tab. 4.3 were calculated, for each of the four methods. The difference between LSC–PCA and OLS–PCA can again be seen in this figure, with the OLS–PCA image being biased in comparison to that of LSC–PCA.



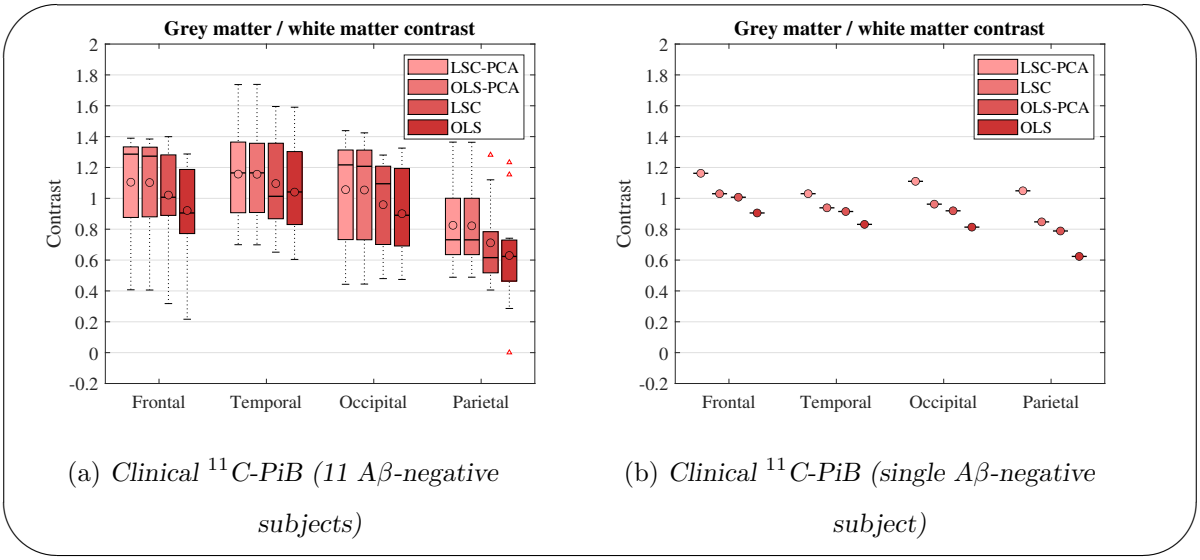
(a) *CFN (PCA: 95% variance)*

(b) *CFN (PCA: 97% variance)*

(c) ^{11}*C-PiB (PCA: 95% variance)*

(d) ^{11}*C-PiB (PCA: 97% variance)*

Figure 4.3: *Percentage recovered DVR of the average ($N = 1024$) DVR values estimated from the noisy tTACs by the four methods, LSC–PCA, LSC, OLS–PCA and OLS. The error bars denote the standard deviations. In the upper panel, (a) and (b), CFN tTACs were denoised with PCA at 95% and 97% variance respectively. The same applies for* ^{11}*C-PiB in the lower panel, (c) and (d).*



(a) *Clinical* ^{11}*C-PiB (11 Aβ-negative subjects)*

(b) *Clinical* ^{11}*C-PiB (single Aβ-negative subject)*

Figure 4.4: *(a) Box plots of the DVR contrast between gray matter regions, frontal, temporal, occipital and parietal cortices, and white matter region, corona radiata, calculated for the 11 Aβ-negative subjects. For the PCA methods, the tTACs were denoised with 95% variance. The circles in the box plots denote the means, whereas the horizontal lines denote the medians. (b) Plots of the DVR contrast of a single subject. For these plots, for the PCA methods, the brain volume was denoised with 97% of variance.*

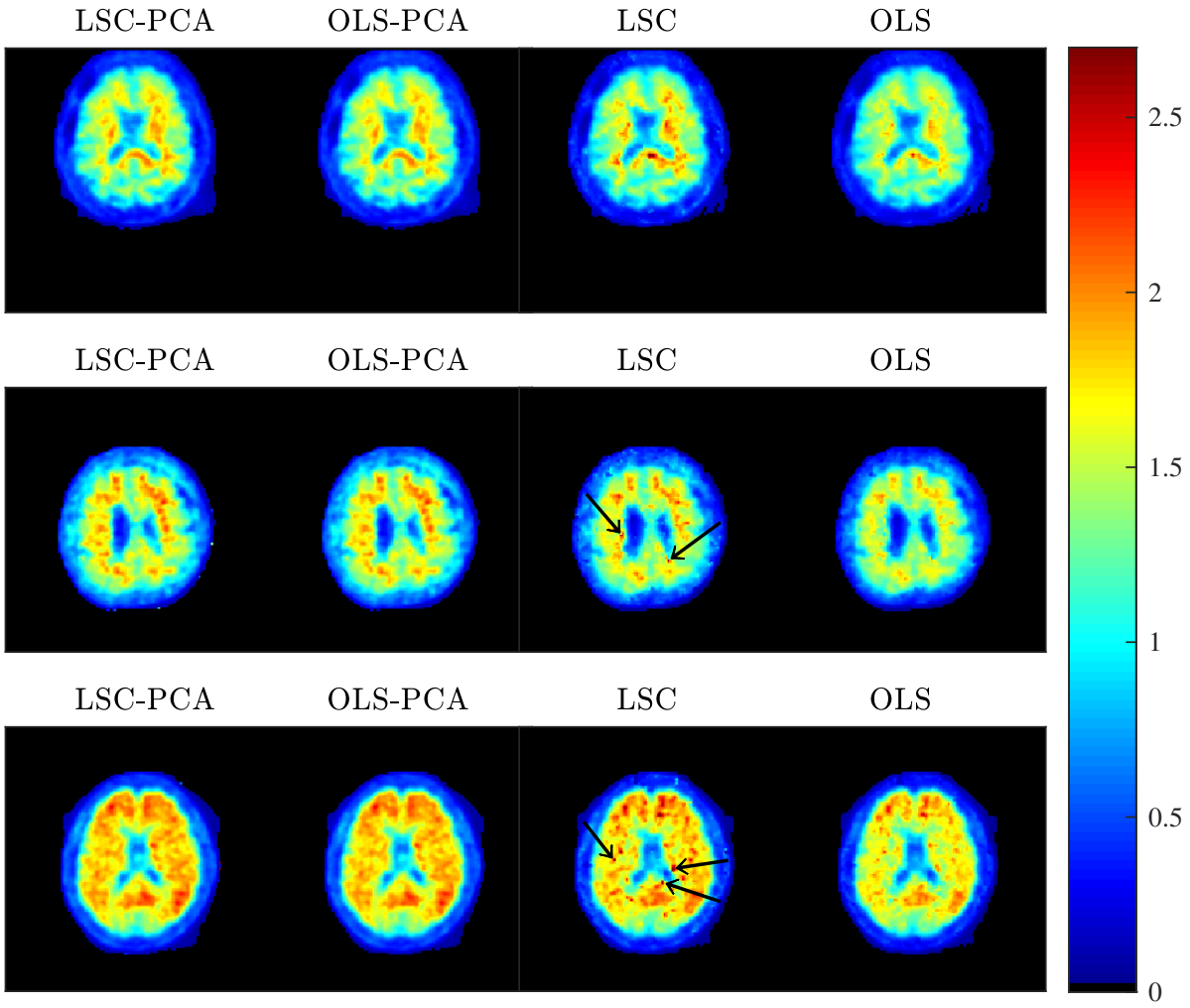


Figure 4.5: *DVR parametric images as obtained by the four methods, LSC–PCA, OLS–PCA, LSC and OLS. The two upper rows are for Aβ-negative subjects, whereas the lowermost row is for the Aβ-positive subject. For PCA-based methods, the tTACs were denoised with PCs that explained 95% of variance.*

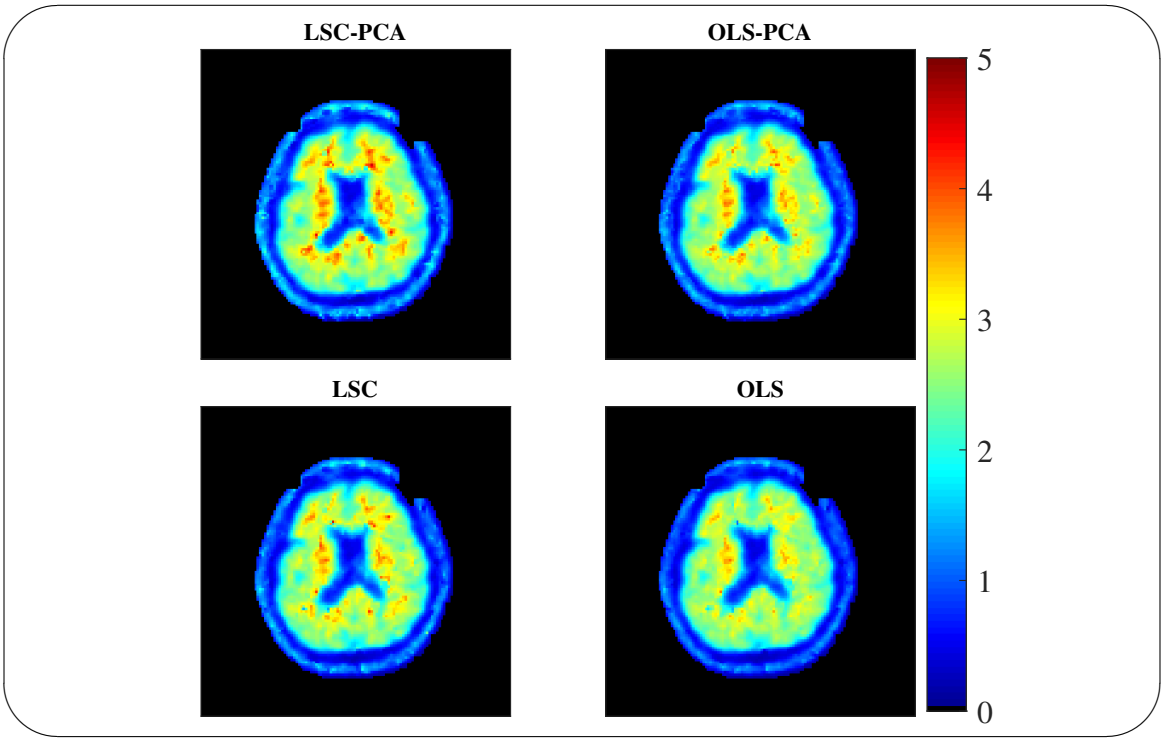


Figure 4.6: *DVR parametric images as obtained by the four methods, LSC–PCA, OLS–PCA, LSC and OLS. In this case, the tTACs were denoised with PCs that explained 97% of variance for the PCA-based methods.*

4.2 Correlated component analysis

Like PCA, CorrCA is also a feature extraction method. CorrCA operates by identifying components that are maximally correlated between repetitions in multivariate data [19]. Specifically, CorrCA maximizes the ratio of between-repetition to within-repetition covariance. In the context of our application to PET data, our repetitions are along the slices dimension. CorrCA will therefore maximize the ratio of between-slices to within-slices covariance. This ratio is generally referred to as inter-subject correlation (ISC) [19]. Specifically in our context, it translates to inter-slice correlation.

Consider a set of dynamic PET brain volume data arranged as an array of size $q \times p \times M$, where M denotes the number of slices, q denotes the number of voxels in a slice, and p denotes the number of time points (frame numbers). CorrCA identifies directions in the p-dimensional space along which the tTACs maximally correlate between M slices, with correlation measured across q voxels.

Let us put it into perspective in comparison to PCA applied to the same data. For the PCA case, the brain volume will be represented as an $R \times p$ array, where R is a product of q and M, and denotes the **total number of voxels in the whole brain volume**. PCA returns a set of p-dimensional vectors which successively capture the variance in the data in a descending order. Similarly, CorrCA returns a set p-dimensional vectors which successively capture the ISC in a descending order. CorrCA also formulates into an eigenvalues and eigenvectors problem, from which the p-dimensional orthogonal vectors (and correlated components) are found as the eigenvectors, and ISCs as the eigenvalues.

Maximizing the between-slices to within-slices covariances maximizes the mean-over-variance across slices, which has been asserted to define a signal-to-noise ratio [19]. CorrCA capitalize on the ability to simultaneously operate "within individual slices (within-slices covariances)" and "across all slices (between-slices covariances)", without arranging the slices in one plane. That means operating directly on the $q \times p \times M$ array, unlike for PCA where the slices are arranged in one plane side-by-side to form the $R \times p$ array. This means that CorrCA is a dual operation, exploiting through all slices whilst also treating each slice individually. This can be seen as advantageous over PCA, which means CorrCA can possibly provide a better noise filter

method over PCA. This therefore forms the motivation to introduce CorrCA to PET parametric imaging and assess the performance of CorrCA in comparison to PCA. For CorrCA only two correlated components were retained to denoise the tTACs.

For a brief mathematical outline of CorrCA as presented in [19], let the observed noisy tTACs be $\mathrm{x}_i^j = [\mathrm{c}_i^j(t_1), \mathrm{c}_i^j(t_2), \cdots, \mathrm{c}_i^j(t_p)]^T$, where $i = 1, \cdots, q$ and $j = 1, \cdots, M$. (^{T} denotes a transpose). Thus, $\mathrm{x}_i^j \in \mathbb{R}^{p\times 1}$ is the ith voxel in the jth slice. This slice can be represented as,

$$\mathrm{x}_i^j(t) = \mathbf{A}\mathrm{y}_i^j + \varepsilon_i^j, \tag{4.24}$$

where $\mathbf{A} \in \mathbb{R}^{p\times k}$ $(k \leq p)$ is a projection matrix to be constructed by CorrCA; $\mathrm{y}_i^j \in \mathbb{R}^{p\times 1}$ is the coefficient vector; and $\varepsilon_i^j \in \mathbb{R}^{p\times 1}$ is the residuals vector.

The jth slice can be written as,

$$\mathbf{C}^j = \begin{bmatrix} \mathrm{c}_1^j(t_1) & \mathrm{c}_1^j(t_2) & \dots & \mathrm{c}_1^j(t_p) \\ \mathrm{c}_2^j(t_1) & \mathrm{c}_2^j(t_2) & \dots & \mathrm{c}_2^j(t_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{c}_p^j(t_1) & \mathrm{c}_p^j(t_2) & \dots & \mathrm{c}_p^j(t_p) \end{bmatrix}_{q\times p}, \tag{4.25}$$

and the whole brain data as,

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^1 & \mathbf{C}^2 & \dots & \mathbf{C}^M \end{bmatrix}_{q\times p\times M}. \tag{4.26}$$

The objective of CorrCA is to find a linear combination of the p measurements (y_i^j), defined as,

$$\mathrm{y}_i^j = \mathrm{v}^T\mathrm{x}_i^j, \tag{4.27}$$

such that the correlation across M slices is maximized —where $\mathrm{v} \in \mathbb{R}^{p\times 1}$ is a projection vector. The ISC (ρ) is defined as the ratio of the between-subject covariance, $\mathbf{r_B}$, and the within-subject covariance, $\mathbf{r_W}$, which can be expressed as,

$$\rho = \frac{1}{M-1}\frac{\mathbf{r_B}}{\mathbf{r_W}}. \tag{4.28}$$

The term $1/(M-1)$ is such that $\rho \leq 1$. In our context, the between-slices covariance can be interpreted as a sum over all pairs of slices, and within-slices covariance can be interpreted as

a sum over all slices, as in the equations below,

$$\begin{aligned}\mathbf{r_B} &= \sum_{j=1}^{M}\sum_{k=1,\,k\neq j}^{M}\sum_{i=1}^{q}(\mathrm{y}_i^j-\bar{\mathrm{y}}^j)(\mathrm{y}_i^k-\bar{\mathrm{y}}^k)\\ \mathbf{r_W} &= \sum_{j=1}^{M}\sum_{i=1}^{q}(\mathrm{y}_i^j-\bar{\mathrm{y}}^j)(\mathrm{y}_i^j-\bar{\mathrm{y}}^j),\end{aligned} \tag{4.29}$$

where $\bar{\mathrm{y}}^j = \frac{1}{q}\sum_{i=1}^{q}\mathrm{y}_i^j$ is the sample-mean of the projected data for the jth slice. Substituting Eq. (4.27) into Eq. (4.29) gives,

$$\begin{aligned}\mathbf{r_B} &= \mathrm{v}^T\mathbf{R_B}\\ \mathbf{r_W} &= \mathrm{v}^T\mathbf{R_W},\end{aligned} \tag{4.30}$$

where $\mathbf{R_B}$ and $\mathbf{R_W}$ are the between-slice and within-slice covariance matrices of x_i^j, respectively given by,

$$\mathbf{R_B} = \sum_{i=1}^{q}\sum_{j=1}^{M}\sum_{k=1,\,k\neq j}^{M}(\mathbf{x}_i^j-\bar{\mathbf{x}}_*^j)(\mathbf{x}_i^k-\bar{\mathbf{x}}_*^k)^T \tag{4.31}$$

$$\mathbf{R_W} = \sum_{i=1}^{q}\sum_{j=1}^{M}(\mathbf{x}_i^j-\bar{\mathbf{x}}_*^j)(\mathbf{x}_i^j-\bar{\mathbf{x}}_*^j)^T. \tag{4.32}$$

Substituting Eq. (4.30) into Eq. (4.28) gives,

$$\rho = \frac{1}{M-1}\frac{\mathrm{v}^T\mathbf{R_B}\mathrm{v}}{\mathrm{v}^\mathrm{T}\mathbf{R_W}\mathrm{v}}. \tag{4.33}$$

To maximize ρ, we rearrange and equate Eq. (4.33) to zero and differentiate with respect to v^T, i.e.,

$$\begin{aligned}&\frac{\partial}{\mathrm{v}^T}\left(\rho\,\mathrm{v}^\mathrm{T}\mathbf{R_W}\mathrm{v}-\mathrm{v}^T\mathbf{R_B}\mathrm{v}=0\right)\\ \implies &\rho\,\mathbf{R_W}\mathrm{v}=\mathbf{R_B}\mathrm{v}.\end{aligned} \tag{4.34}$$

Equation (4.34) means that v is an eigenvector of ${\mathbf{R_W}}^{-1}\mathbf{R_B}$, and ρ is the corresponding eigenvalue —under the assumption that $\mathbf{R_W}$ is an invertible matrix. It is therefore that ISC is maximized by projecting the data onto the eigenvector of ${\mathbf{R_W}}^{-1}\mathbf{R_B}$ with the largest eigenvalue [19]. In a similar manner as for PCA, these eigenvector and eigenvalue can be labeled as v_1 and ρ_1, denoting the first eigenvector and eigenvalue, respectively. Further in a similar way to PCA, successive eigenvectors $(\mathrm{v}_2,\cdots,\mathrm{v}_p)$ and corresponding eigenvalues $(\rho_2,\cdots,\rho_p)$ captures

the ISC in a decreasing order. To find all the eigenvectors and eigenvalues, maximal ISC is summed over all components. The components are required to be uncorrelated, which means each component captures "different and independent" characteristics of the data [19].

In entirety, Eq. (4.33) can therefore be expressed as,

$$\mathbf{J}(\mathbf{V}) = \sum_{d}^{p} \rho_d = \sum_{d}^{p} \frac{\mathrm{v}_d^T \mathbf{R_B}\, \mathrm{v}_d}{\mathrm{v}_d^T \mathbf{R_W}\, \mathrm{v}_d}. \tag{4.35}$$

Thus, CorrCA maximize Eq. (4.35) with respect to $\mathbf{V} = [\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_p]$, subject to $\mathrm{v}_d^T \mathbf{R_W}\, \mathrm{v}_c = 0$ for $c \neq d$. The solution to Eq. (4.35), corresponding to Eq. (4.34) is,

$$\mathbf{R_B V} = \mathbf{R_W V \Lambda}. \tag{4.36}$$

Here, $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$ denotes a diagonal matrix consisting of the ISC values (ρ_d) in the diagonals. These ISC values are eigenvalues of the eigenvectors v_d which defines the projections $\mathrm{y}_d = \mathrm{v}_d^T \mathrm{x}$. The first projection, along the first eigenvector, provides the highest ISC across M slices. Successive projections are uncorrelated with each other and captures ISC in decreasing order, i.e., the second projection correspond to the direction with second largest ISC and so forth, until the last projection which corresponds to the direction with the smallest ISC. [19].

In matrix format, Eq. (4.24) can be written as,

$$\mathbf{X}|_{\mathbf{p \times q \times M}} = (\mathbf{A}|_{\mathbf{p \times k}})\, (\mathbf{Y}|_{\mathbf{k \times q \times M}}) + \varepsilon|_{\mathbf{p \times q \times M}}. \tag{4.37}$$

The least-squares estimate of $\mathbf{A}$ is,

$$\hat{\mathbf{A}} = \mathbf{R_W v} (\mathbf{v}^T \mathbf{R_W v})^{-1}. \tag{4.38}$$

Analogous to using only a few PCs in PCA, here, using only the first few k ($k \leq p$) correlated components with the highest ISC values in Eqs. (4.37) and (4.38), denoised tTACs can be estimated as per Eq. (4.24) as,

$$\hat{\mathbf{C}} = \hat{\mathbf{X}} = \hat{\mathbf{A}} \hat{\mathbf{Y}}. \tag{4.39}$$

For mean-centred input data, $\mathbf{C_0}$, we get that,

$$\hat{\mathbf{C}} = \hat{\mathbf{A}} \hat{\mathbf{Y}} + \bar{\mathbf{C}}, \tag{4.40}$$

where $\bar{\mathbf{C}}$ is the mean of $\mathbf{C}$.

In short, it can be inferred from Eqs. (4.34) and (4.36) that CorrCA performs PCA on the matrix of the ratio of between-slices to within-slices, $\mathbf{R_B R_W^{-1}}$, whose the eigenvalues define the ISC.

CorrCA was computed using the MATLAB codes provided by Parra and colleagues [19]. The code can be found at *http://parralab.org/corrca*.

4.2.1 Simulation studies

Simulation data design

The simulation parameters and conditions used for CorrCA assessment in this section are the same with those used in section 4.1.1 for PCA studies. However, here a further step was taken to evaluate the methods at different noise levels. The magnitudes of the noise level were calculated as the percentage of the ratios of the "standard deviation of the noisy tTACs" to the "mean of the noise-free tTAC" for the portion of the time range used for the *DVR* estimation. For CorrCA assessment, the focus is mostly on LSC-CorrCA against LSC-PCA. However, a brief analysis of LSC–CorrCA and OLS–CorrCA, demonstrates (for OLS–CorrCA) the similar issue observed with OLS–PCA that one needs to be extra careful when choosing the number components to retain to denoise the tTACs. The results reported for LSC-CorrCA are mainly for the data denoised with only two components (correlated components). This is to assess if LSC-CorrCA with only two correlated components can achieve better or similar results with LSC–PCA with many PCs.

Results

Figure 4.7 shows a simulated noise-free tTAC, a corresponding noisy tTAC, and three denoised tTACs, for the two radiotracers, CFN 4.7(a) and ^{11}C-PiB 4.7(b). Two of the denoised tTACs were denoised by CorrCA with 2 and 5 components, as denoted by the subscripts in the legend. For PCA, examples of tTACs denoised with many components are already shown in section 4.1.1 (Fig. 4.2). In this figure the PCA-denoised tTACs were denoised with 2 PCs. The tTACs denoised by CorrCA with two components appears to to give the best estimate of the true (noise-free) tTAC. The tTAC denoised by PCA with two PCs appear to overestimate the true curve in Fig. 4.7(a) for CFN, but appears to underestimate the one for ^{11}C-PiB in 4.7(b). Overall all three denoised tTACs appears to be considerably denoised. However, especially for PCA, it is known that denoising the tTACs with only two PCs leads to information loss which

leads to low quality images. That means looking at the tTACs alone is not enough to conclude whether a denoising method will lead to improved images.

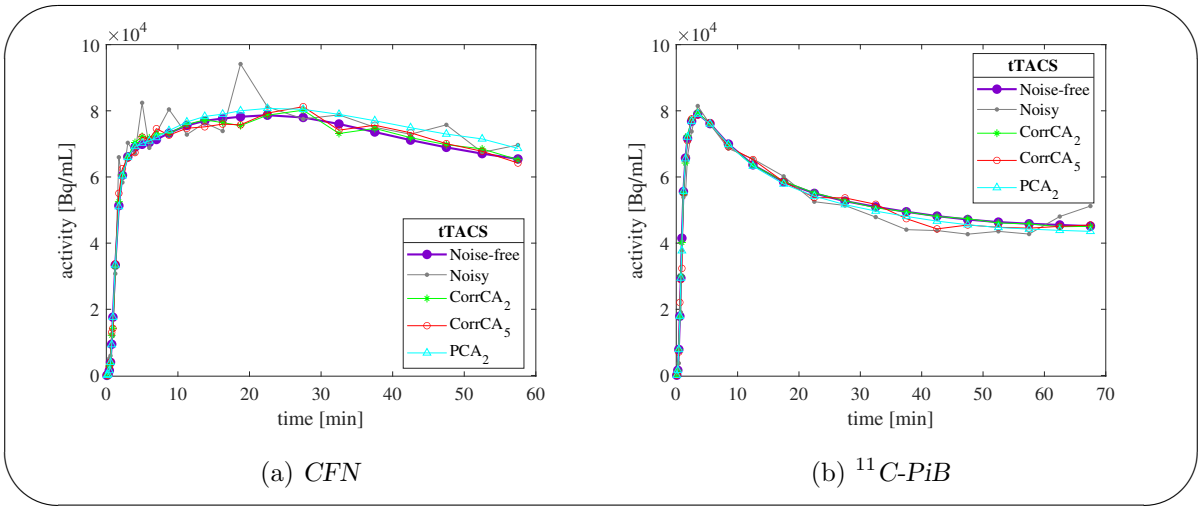


Figure 4.7: *Simulated noisy tTACs and their corresponding CorrCA- and PCA-denoised tTACs for the two radiotacers, (a) CFN and (b) ^{11}C-PiB. Three denoised tTACs are shown for each radiotracer – two denoised by CorrCA and one denoised by PCA. The subscripts denote the number of components used to denoise the corresponding curves.*

Figure 4.8 shows the percentage bias in the averages of *DVR* values estimated from the noisy tTACs by four methods, LSC–CorrCA, LSC–PCA, LSC, and OLS. The percentage bias is the difference from the 100% mark, i.e., the closer the plot is to 100% the less biased it is. Here the focus is on LSC–CorrCA and LSC–PCA. It can be seen in both Figs. 4.8(a) and 4.8(b) that the variance is drastically reduced in the LSC–CorrCA estimates. This is mostly because only 2 components were used to denoise the data for LSC–CorrCA, in comparison to many PCs (which could go up to 10) used to retain 95% variance for LSC–PCA. Estimates from tTACs denoised by PCA with 2 PCs would also have smaller variances. It then remains to evaluate the clinical images to assess which method can hold much information in a smaller number of components.

In Fig. 4.9, it is shown for both radiotracers that, when all the correlated components are used, CorrCA returns the original data, and therefore the *DVR* estimates of LSC–CorrCA are identical to those of LSC. This is also a known characteristic of PCA [16] , i.e., when all PCs

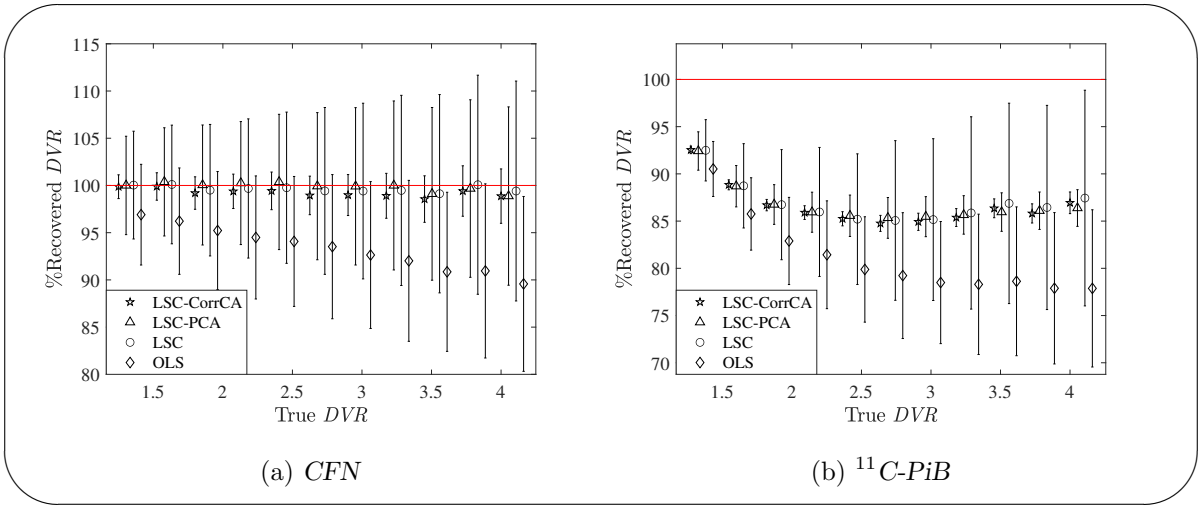


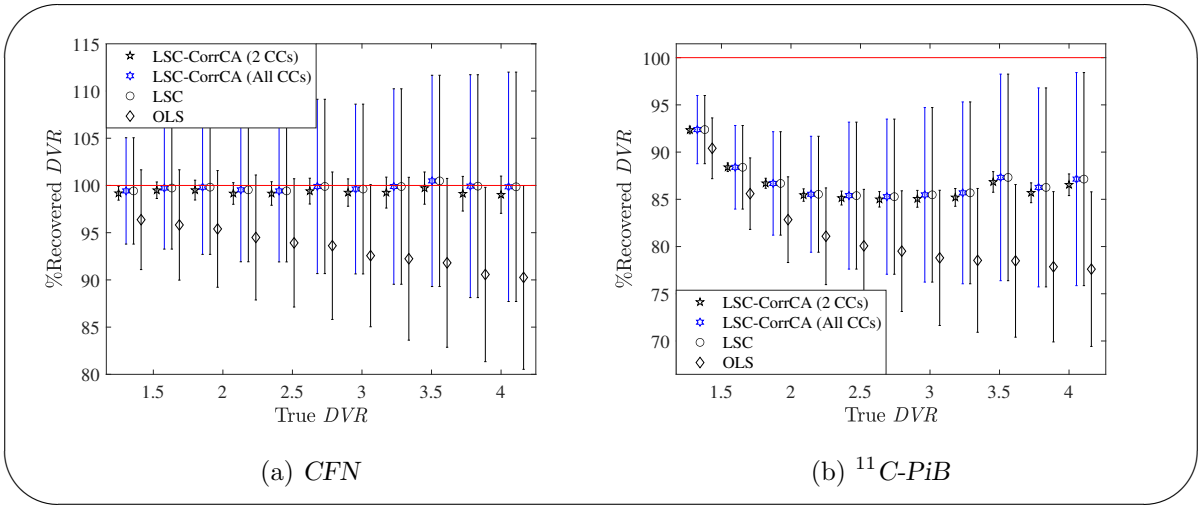
Figure 4.8: *Percentage recovered DVR of the average ($N = 1024$) DVR values estimated from the noisy tTACs by four methods, LSC–CorrCA, LSC–PCA, LSC and OLS. The error bars denote the standard deviations. In both figures, (a) and (b), the tTACs were denoised with two components for LSC–CorrCA. For LSC–PCA, the tTACs were denoised with the number of PCs accounting for 95% of variance in the data.*

are used PCA returns the original data.

Figure 4.10 shows the slices estimated from the noisy tTACs by four methods, LSC–CorrCA, LSC–PCA, LSC and OLS. In this figure, it can be seen that the LSC–CorrCA estimates show better estimates than other methoods, in agreements with the smaller variances observed in Fig. 4.8. The noise-free slices are shown in Fig. 4.11.

Figure 4.12 compares the *DVR* values estimated by LSC–CorrCA and OLS–CorrCA, in which the tTACs were denoised with 5 components. It is revealed here that if many components are used, it raises the bias in the OLS–CorrCA estimates, similar to what was observed for OLS–PCA. This further puts LSC to be the regression method of choice for LGA-based PET parametric imaging.

In Fig. 4.13, the four methods, LSC–CorrCA, LSC–PCA, LSC and OLS are compared at different noise levels. In Fig. 4.13(a), the estimates are plotted individually for each noise level. In Fig. 4.13(b), the estimates at all noise levels are put together in a box plot. It is seen again



(a) *CFN* (b) ^{11}*C-PiB*

Figure 4.9: *Percentage recovered DVR of the average ($N = 1024$) DVR values estimated from the noisy tTACs by four methods, LSC–CorrCA (with two correlated components), LSC–CorrCA (with all correlated components), LSC and OLS. The error bars denote the standard deviations. Figures (a) and (b) shows the estimates for CFN and ^{11}C-PiB radiotracers, respectively.*

here how LSC–CorrCA has reduced variances.

In Fig. 4.13, however, the minimum number of PCs was set to three. This means that, if the number of PCs that accounts for 95% of variance are less than 3, 3 PCs are used, and if the number of PCs that accounts for 95% of variance are more than three, then that number of PCs are used. Figure 4.14 is the equivalent of Fig. 4.13 without the restriction of minimum number of PCs of 3. The effect of this condition can be seen for the first three lower noise levels in Fig. 4.14(a), in which LSC–PCA shows an underestimation in comparison to all other methods —including OLS-based LGA. The underestimation in LSC–PCA-based estimates is also reflected in Fig. 4.14(b). For the first three lower noise levels, the number of PCs that accounted for 95% of variance is one. This observation provides further insight into the notion that using a few number of PCs leads to information loss. A similar trend was observed in [16], in which 1 PC-based estimates appeared slightly biased compared to those of 2 PCs. It is therefore important to always consider including more than 1 PC when denoising by PCA.

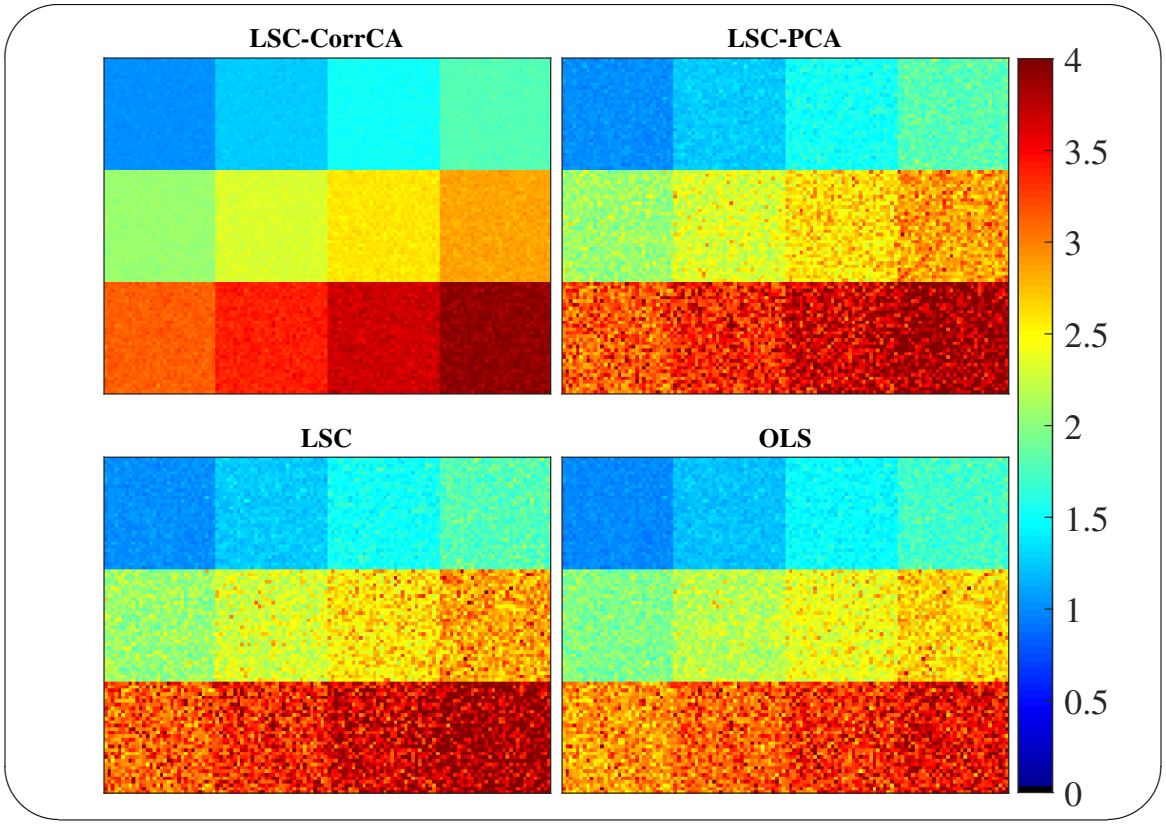


Figure 4.10: *A display of the slices obtained by four methods, LSC-CorrCA, LSC-PCA, LSC and OLS from simulated noisy tTACs for CFN radiotracer (DVR = 4). Each color block represents a slice of 32 × 32 voxels, which makes a total of 1024 tTACs. The tTACs were denoised with 2 components for LSC–CorrCA, and with 95% variance for LSC–PCA.*

4.2.2 Clinical ^{11}C-PiB PET data studies

Study design

The clinical data described and used for analysis in section 4.1.2 are the same data used for this section. Under this section, the focus is completely on LSC–CorrCA vs LSC–PCA.

Results

Figure 4.15 compares the box plots of the contrast between gray and white matter for images obtained by two methods, LSC–CorrCA and LSC–PCA. The tTACs were denoised with 2 components for LSC–CorrCA, and with 95% variance for LSC–PCA. The contrast was calculated for 11 Aβ-negative images, as it was done in section 4.1.2 for Fig. 4.4(a). Some difference could be seen in the mean (circle marks) and medians (horizontal lines) in the box plots of the two

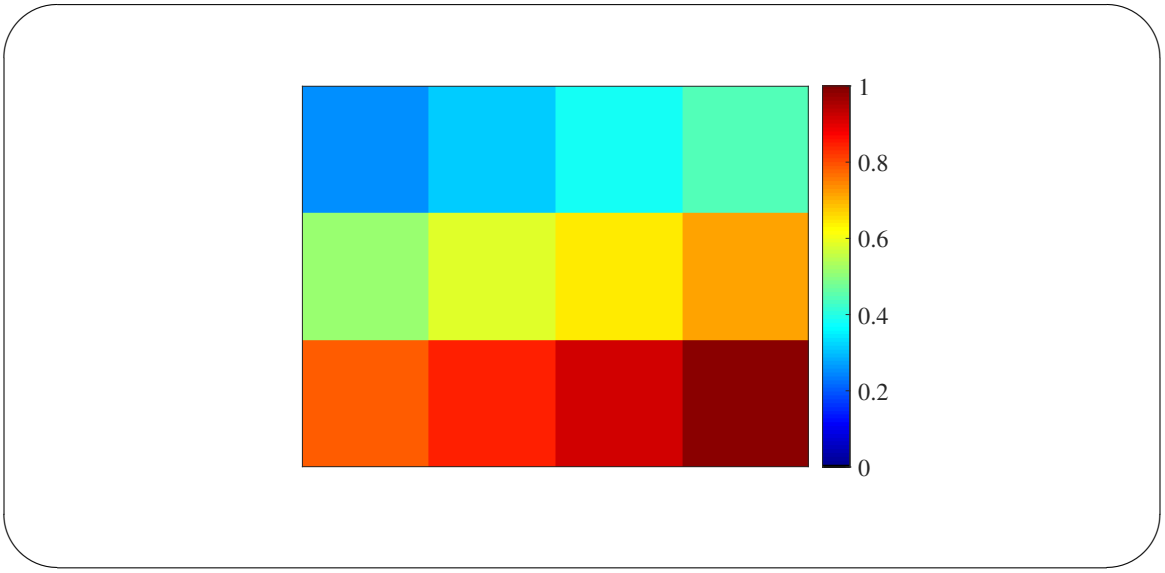


Figure 4.11: *A display of the simulated noise-free slices.*

methods in Fig. 4.15. However the p value of their statistical differences across the four regions as calculated by the Friedman Post-Hoc multiple comparison is 0.076, slightly over the desired 0.05 statistical significance mark. Numerical values for Fig. 4.15, mean values ans standard deviations are shown in Tab. C.1.

Table 4.5 shows the computation time of four methods, LSC–CorrCA, LSC–PCA, LSC and OLS. These data show another advantage aspect of the LSC–CorrCA method compared to LSC–PCA, in which LSC–CorrCA takes shorter computation time compared to LSC–PCA. As expected, both LSC–CorrCA and LSC–PCA took longer computation time compared to LSC. This is because of the additional task of data denoising, by CorrCA and PCA, in addition to regression parameter estimation by LSC. The OLS approach required the shortest computation time, as expected as well.

Figure 4.16 compares the images obtained by two methods, LSC–CorrCA and LSC–PCA, for two subjects, Aβ-negative and Aβ-positive. For the Aβ-negative subject, the LSC–CorrCA image appears to show higher DVR estimates. For the Aβ-positive subject, DVR estimates across the slice appear about the same for LSC–CorrCA and LSC–PCA, with LSC–PCA image showing a rather smoother distribution. For LSC–CorrCA images, it is worth noting that they were all

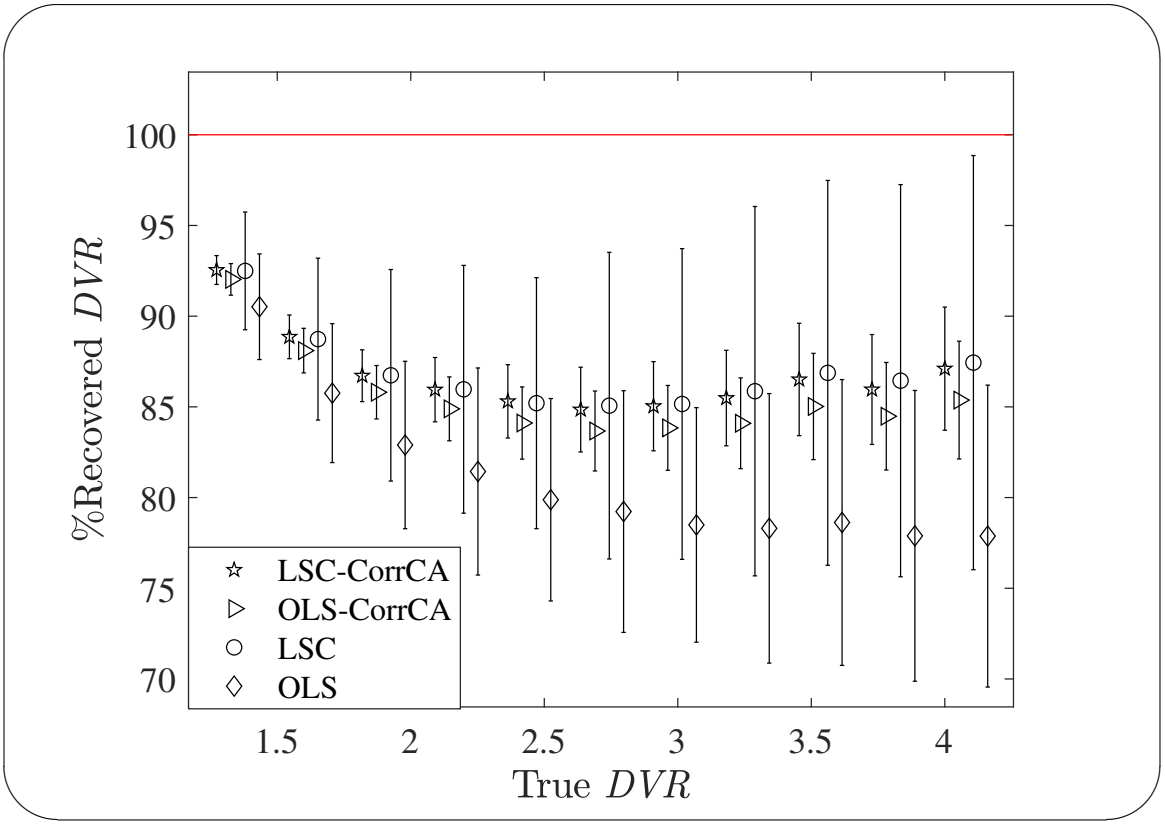


Figure 4.12: *Comparison of the percentage recovered in the average ($N = 1024$) DVR values estimated from the noisy tTACs by four methods, LSC–CorrCA, OLS–CorrCA, LSC and OLS. For these results, the tTACs were denoised with 5 components for both LSC–CorrCA and OLS–CorrCA.*

estimated with 2 components, and an optimization for the number of components could make for an improvement of the images. Overall, these results show that LSC–CorrCA with only 2 components achieved about the same or better results with LSC–PCA with many components. This means that CorrCA contains much information in a smaller number of components, in comparison to PCA. For PCA, the number of PCs that accounted for 95% variance in the 12 clinical data sets (11 Aβ-negative subjects and 1 Aβ-positive subject) used in this study were all greater than 2, ranging from 4 to 8.

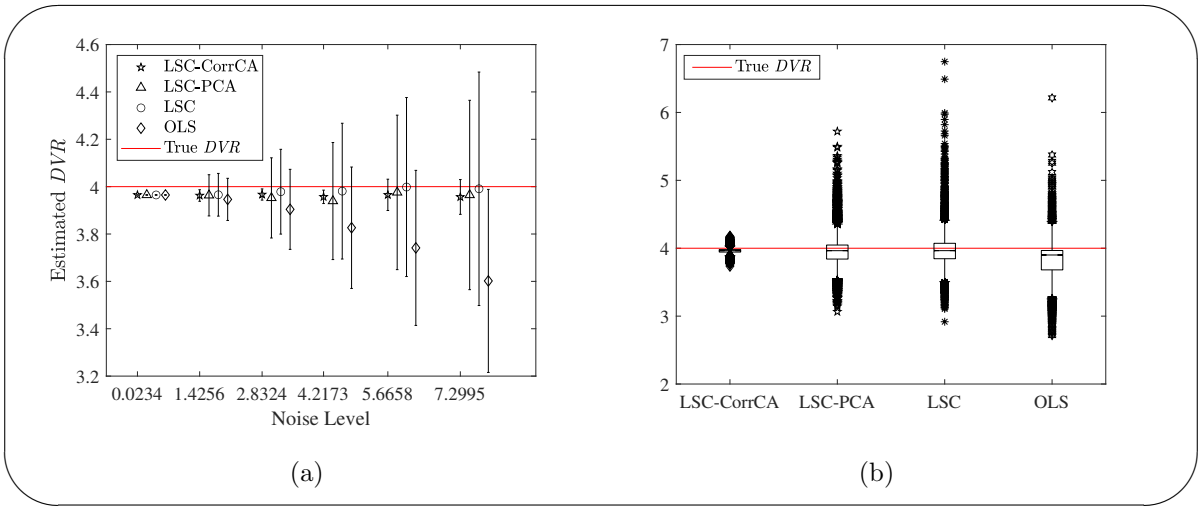


Figure 4.13: *DVR estimates of CFN radiotracer. (a) A comparison of DVR (DVR = 4) estimation at different noise levels by four methods, LSC–CorrCA, LSC–PCA, LSC and OLS. (b) Box plots of the estimated DVR (DVR = 4) values. These plots includes the values estimated at all noise levels. The tTACs were denoised with two components for LSC–CorrCA. For LSC–PCA, the tTACs were denoised with the number of PCs accounting for 95% of variance in the data (minimum number of PCs set to three).*

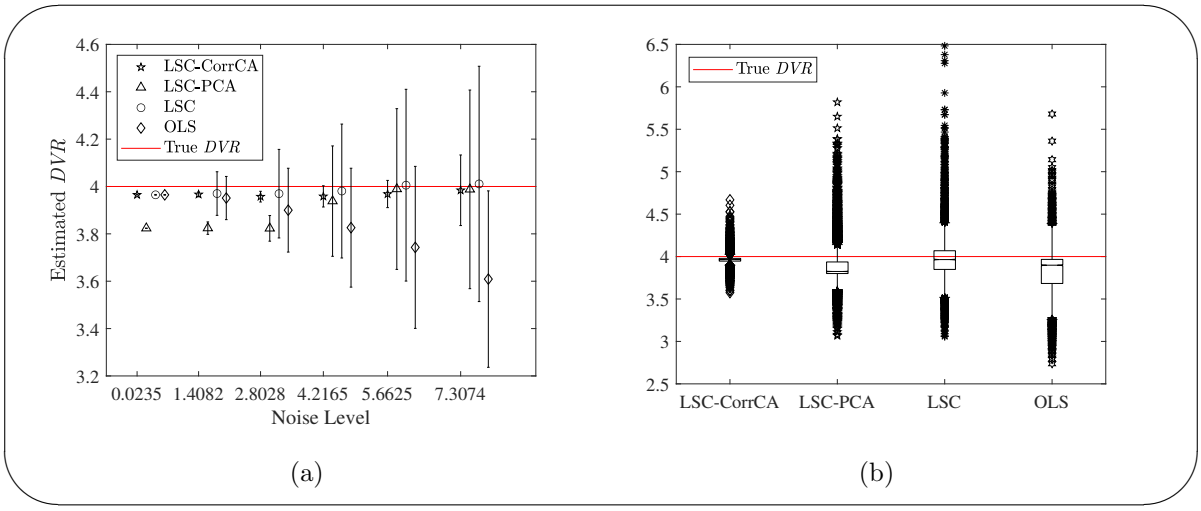


Figure 4.14: *DVR estimates of CFN radoitracer. (a) A comparison of DVR (DVR = 4) estimation at different noise levels by four methods, LSC–CorrCA, LSC–PCA, LSC and OLS. (b) Box plots of the estimated DVR (DVR = 4) values. These plots includes the values estimated at all noise levels. The tTACs were denoised with two components for LSC–CorrCA. For LSC–PCA, the tTACs were denoised with the number of PCs accounting for 95% of variance in the data.*

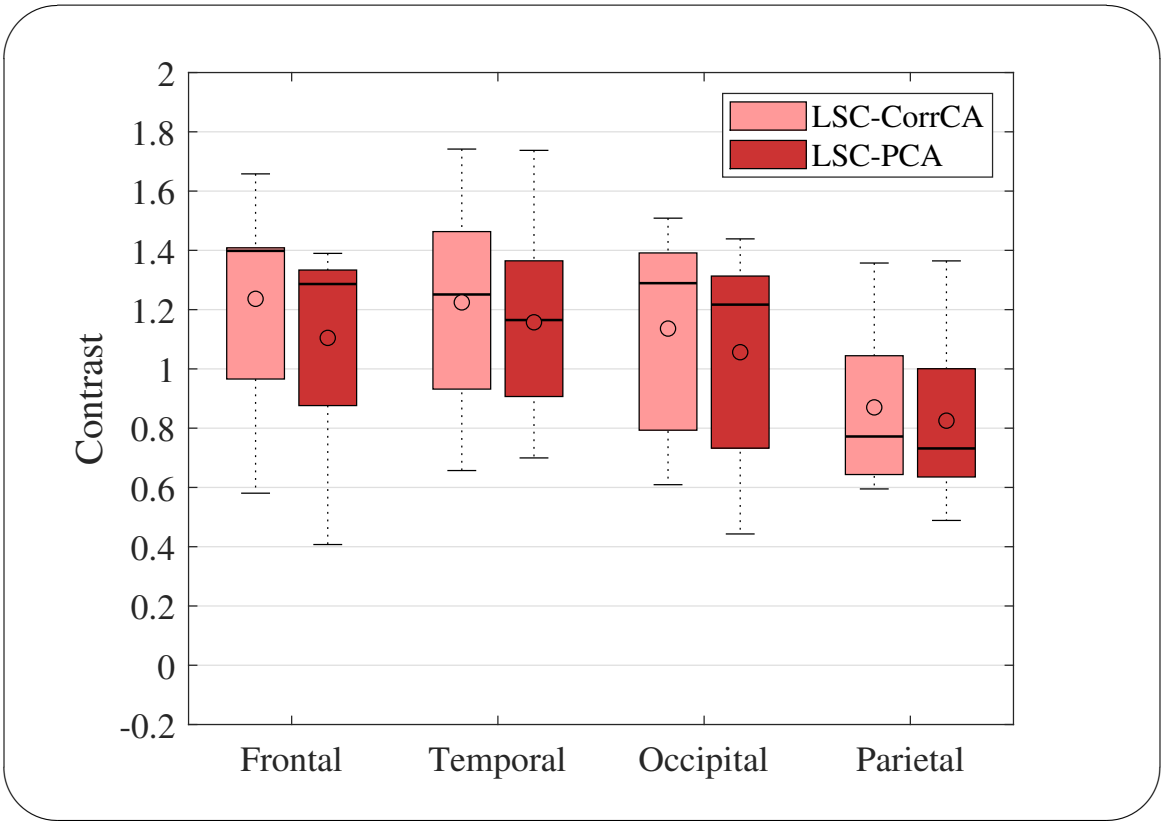


Figure 4.15: *Box plots of the DVR contrast between gray matter regions, frontal, temporal, occipital and parietal cortices, and white matter region, corona radiata, calculated for the 11 Aβ-negative subjects for the images obtained by two methods, LSC–CorrCA and LSC–PCA. For LSC–CorrCA, the tTACs were denoised using two components. For the LSC–PCA, the tTACs were denoised with 95% variance. The circles in the box plots denote the means, whereas the horizontal lines denote the medians.*

Table 4.4: *Numerical values for the contrast between gray and white matter regions of the images obtained by two methods, LSC–CorrCA and LSC–PCA.*

	Contrast			
	Frontal	Temporal	Occipital	Parietal
LSC-CorrCA	1.236 ± 0.341	1.224 ± 0.360	1.136 ± 0.349	0.870 ± 0.266
LSC-PCA	1.105 ± 0.329	1.157 ± 0.333	1.056 ± 0.351	0.825 ± 0.296

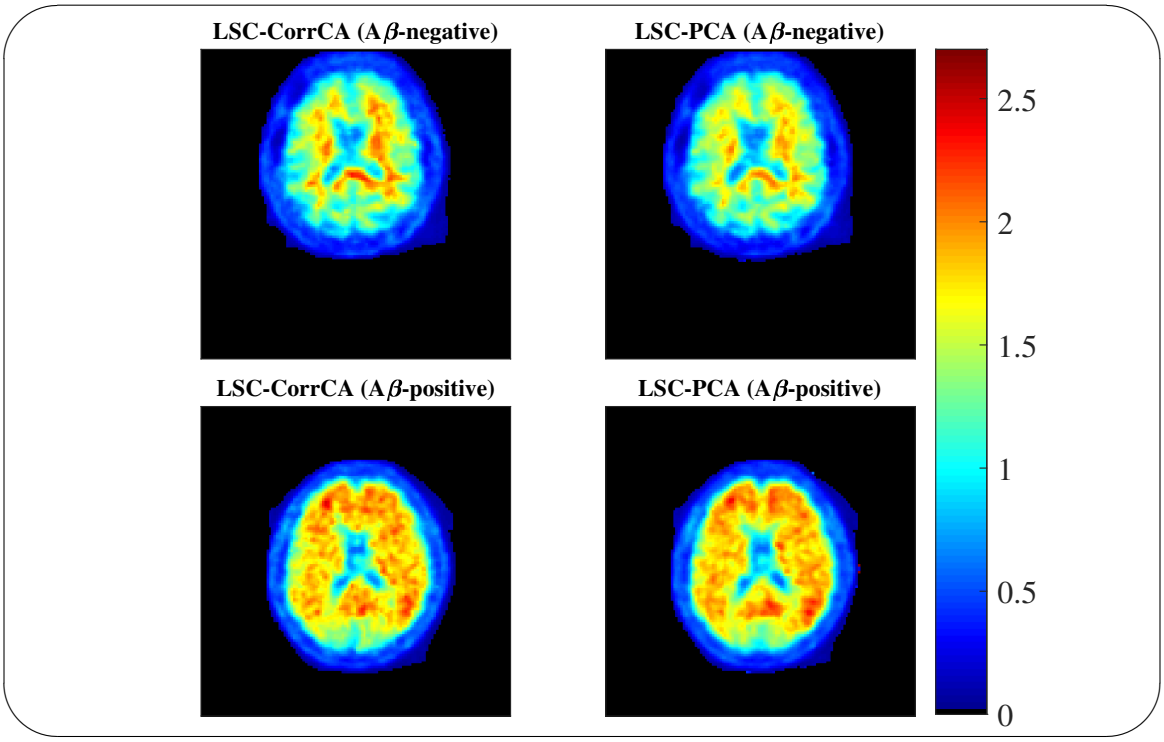


Figure 4.16: *DVR parametric images of a Aβ-negative and Aβ-positive subject, obtained by two methods, LSC–CorrCA and LSC–PCA. The tTACs were denoised with two components for LSC–CorrCA, and with 95% variance for LSC–PCA.*

Table 4.5: *Means (± standard deviations) of computation times taken by each of the four methods, LSC–CorrCA, LSC–PCA, LSC, and OLS, to obtain a whole DVR image of size 128 × 128 × 47 —calculated for the 11 Aβ-negative subjects and 1 Aβ-positive subject (i.e., N = 12).*

Computation time (s)			
LSC–CorrCA	**LSC–PCA**	**LSC**	**OLS**
9.249 ± 1.811	10.678 ± 1.225	7.514 ± 0.879	2.570 ± 0.165

4.3 Discussion

The results obtained in section 4.1 demonstrate that the LSC–PCA method improves the contrast of parametric images, in comparison with LSC. With LSC and PCA combination, LSC reduces the bias, and PCA could help fend off the effects of the variations in the estimates, resulting in parametric images with both minimal bias and variance. OLS–PCA was also found to have reduced bias and improved contrast, but this was found to be greatly affected by the number of PCs used to denoise the tTACs. For the CFN and ^{11}C-PiB simulations, it was seen that the OLS–PCA estimates could get biased if one is not careful in determining the number of PCs. In this case, one should then rely on LSC for bias reduction. This should therefore highlight the importance of adopting a less-biased regression method, LSC, even when PCA tTACs denoising method is applied. This is important especially when one is unsure of the number of PCs to use, which is usually the case in practical settings.

PCA has been shown to reduce bias using only a few PCs [16]. Using a minimal number of PCs might result in information loss, leading to lower contrasts. This means that information is lost with the number of PCs discarded, i.e. the fewer the number of PCs used the more information is lost. The question of how many PCs to retain has no exact answer. It mostly depends on the purpose of the study in question. For PET imaging, to maintain the spatial resolution of the images, it is necessary to retain sufficient variance to reconstruct denoised tTACs. This can always be set by determining the percentage variance to be accounted for in the denoised data, and all the PCs that add up to that value can then be retained. For OLS–PCA it is therefore that using a few PCs leads to information being lost. On the other hand, using many PCs reintroduces the noise in the data, and the estimates get biased. This is what LSC–PCA is addressing; to allow many PCs to be used without risking bias.

The CorrCA denoising approach was applied to PET parametric imaging for the first time, and assessed against PCA. CorrCA is one of the many various approaches to canonical correlation analysis (CCA) of multidata set, developed to identify linear combinations of signals that are reliably reproduced across the data. In our context of a single PET image, CorrCA operates by maximizing the ratio of the correlations measured within and across the slices in a PET image. This is in contrast to PCA which focuses on the variance. Here, CorrCA is employed as a temporal filter of tTACs, similar to the manner in which PCA is used. By using only a

few dimensions with the largest ISCs, one could get rid of the noise in the data along with the remaining components with smaller ISCs, whilst preserving the essential characteristics of the data.

The simulation results showed that both LSC–CorrCA and LSC–PCA reduced the variance in the *DVR* estimates, in comparison to LSC. This was indicated by smaller error bars in the plotted figures. LSC–CorrCA estimates showed the smallest error bars amongst all methods, suggesting that it will give clinical images with improved contrasts. It was also shown that OLS–CorrCA estimates get biased if more components are used, similar to OLS–PCA; this informs that LSC should be the preferred regression method to estimate the LGA slope.

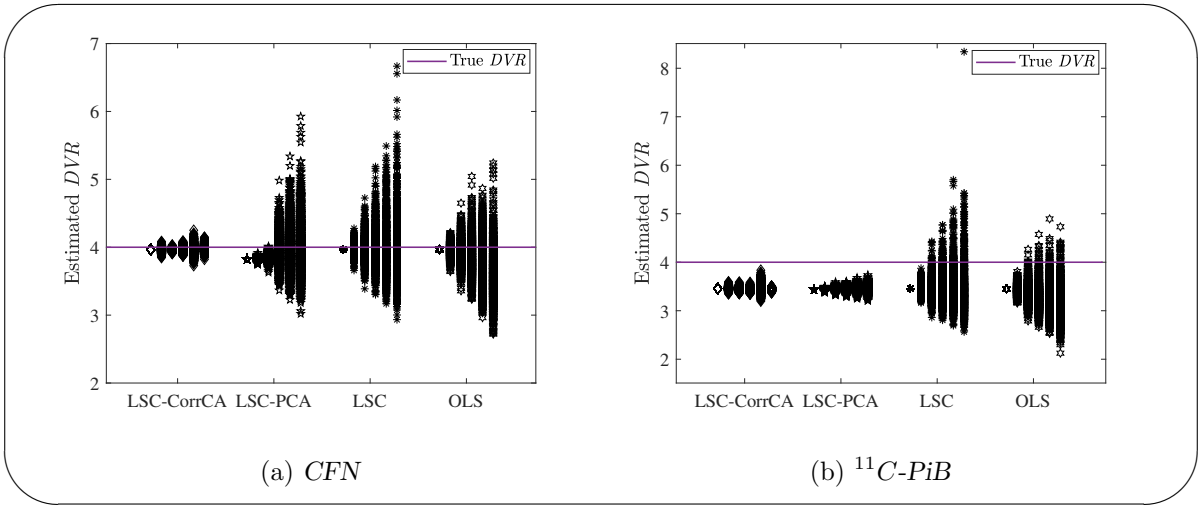
The contrast of LSC–CorrCA images were observed to be higher than those for LSC–PCA images, as expected based on the simulation results. However, this difference was not statistically significant in terms of the calculated p-value. These findings based on the calculated contrasts are in agreement with the visuals of the displayed images, as not much of a difference could be seen between the images obtained by the two methods, LSC–CorrCA and LSC–PCA. It should however be noted that LSC–CorrCA with only two components was able to achieve seemingly better or similar results with LSC–PCA with many PCs. This could be that CorrCA capture much information in a less number of components in comparison to PCA. This will require further analysis to be established as such. As for computation time, LSC–CorrCA outperformed LSC–PCA.

Another aspect of LSC–CorrCA is depicted in Fig. 4.17. Figure 4.17 shows scatters of *DVR* estimates at different noise levels for four methods, LSC–CorrCA, LSC–PCA, LSC, and OLS. In comparison to other methods, the variance of LSC–CorrCA estimates seems to be unaffected by noise level, as the dispersion in the scattered *DVR* estimates does not show an increase with noise level. This signifies a positive feature for LSC–CorrCA. Other methods show a clear increase in the dispersion of *DVR* estimates with respect to noise level. For LSC–CorrCA method however, the scattered *DVR* estimates show some irregularities with respect to noise level, with some estimates at lower noise levels exhibiting larger dispersion than those at higher noise levels. This implies some instability in the LSC–CorrCA method. For CFN radiotracer, the effect of this observation appears to be not of concern in comparison to LSC–PCA. For ^{11}C-PiB radio-

tracer, however, it appears these estimates with unexpected larger variances attains variances larger than those for LSC–PCA. The root cause of this instability resulting in random increases in the dispersion of LSC–CorrCA estimates is unclear at this point, and it will be studied in future researches.

In the approach proposed in this study, the number of correlated components to keep was determined with $\hat{\mathbf{A}}$ and $\hat{\mathbf{Y}}$, which are used in Eqs. (4.39) and (4.40). As explained earlier, this approach is equivalent to performing PCA on the matrix $\mathbf{R_B R_W^{-1}}$. In other words, this approach is as well equivalent to using a few, k, principal vectors in $\mathbf{v^T}$ in Eqs. (4.27) and (4.38). Therefore, the ISC values correspond to the variances explained by the eigenvectors of $\mathbf{R_B R_W^{-1}}$ in this PCA operation.

An alternative method is to perform the truncation in $\mathbf{R_B}$ and $\mathbf{R_W}$ in Eq. (4.36) [19]. In doing this, the resulting $\hat{\mathbf{A}}$ and $\hat{\mathbf{Y}}$ will also have reduced dimensions. This approach seemed to have reduced the irregularities observed in Fig. 4.17. However, this approach required a larger number of dimensions (correlated components) for better contrasts to be obtained in clinical brain images. Based on the analysis of real data contrast values, this approach also seemed to fit better with a method based on shuffle statistics [19] for determining the number of relevant correlated components. Shuffle statistics is used to test for nonlinearity of time series data. In short, the idea of the shuffle statistics method is to generate surrogate data that are consistent with the null hypothesis, but also similar to the original data in terms of time and spatial correlation [19]. For CorrCA, in the context of our PET brain images, the null hypothesis states that there is no shared one-dimensional subspace in which the data of at least two slices are correlated. A shuffle statistics technique called *random circular shift* [19] is used to test a significant ρ based on the consideration that there is no correlation between slices after a random circular shift for each slice [19]. However, the number of correlated components that gave better contrast values were smaller than those estimated by the shuffle statistics-based approach —hence, overestimation. Overall, the approach used in this study (above paragraph) appears better than the approach described in this paragraph. However, a thorough analysis of these two approaches will be necessary in the future.



(a) *CFN* (b) ^{11}*C-PiB*

Figure 4.17: *Scatter of DVR estimates at 6 different noise levels (DVR = 4). (a) Scatter of DVR values estimated by four methods, LSC–CorrCA, LSC–PCA, LSC and OLS, from noisy CFN tTACs. (b) Scatter of DVR values estimated by four methods, LSC–CorrCA, LSC–PCA, LSC and OLS, from noisy* ^{11}*C-PiB tTACs. The tTACs were denoised with two components for LSC–CorrCA. For LSC–PCA, the tTACs were denoised with the number of PCs accounting for 95% of variance in the data.*

Chapter 5

Summary and future directions

This study aimed at improving the LGA-based parametric estimates by reducing both the bias and variance. In the first part of the study (Chapter 3), an alternative linear regression method, LSC, was employed, and its viability was assessed for BP_{ND} estimation. The results showed that LSC-based estimates are the least biased in comparison to those of MRTM2 and conventional OLS-based LGA. Specifically, LSC reduced the bias in the estimates of BP_{ND}, up to 12% and 28% difference compared with MRTM2 and conventional OLS-based LGA, respectively. In comparison to the standard OLS-based LGA, LSC reduced the bias at a slight expense of increased variation. However, in total, LSC providied a better trade-off between bias and variability, making it a promising tool for LGA-based PET parametric imaging.

Having noted the slightly increased variances in the LSC-based estimates in comparison with OLS-based estimates, the second part of the study (Chapter 4) employed two tTACs denoising techniques. These are PCA and CorrCA. Since LSC-based estimates are already least-biased, the two tTACs denoising methods were employed to reduce the variance in parametric estimates. Both PCA and CorrCA operates by finding the most important dimensions. These approaches were assessed in terms of DVR estimation. The results obtained demonstrate that both LSC–CorrCA and LSC–PCA methods improved the contrast of the parametric images. This was supported by a reduction in the variation of the estimates observed in the simulation results. The contrast of LSC–CorrCA images were observed to be slightly higher than those for LSC–PCA images, as expected based on the simulation results. On the other hand, OLS–CorrCA and OLS–PCA were also found to improve the contrast of parametric images. However, more caution is needed for the OLS-based method because if more components are used, the estimates

could get biased. This therefore demonstrates that it is important to always consider using LSC as a regression method even when the data are denoised with CorrCA or PCA. Thus, this forms the basis to encourage the adoption of LSC-based approaches, LSC–CorrCA and LSC–PCA, in which the complementary capabilities of LSC and PCA or CorrCA assures that both bias and variance will be reduced, unlike in OLS–CorrCA or OLS–PCA in which one always risks bias. These findings should be useful in PET parametric imaging for appropriate interpretation of radiotracer binding in the brain. These are complementary approaches in which LSC mainly serve to reduce the bias, and PCA and CorrCA are employed to minimise the variance in the estimates.

For LSC, a deeper look into weighting techniques could provide for improved estimates. The appropriateness of the weight function used in this study is demonstrated in section 3.1. However, they are estimates, and there is still a room for improvements. Improved weighting could allow LSC to be a stand-alone approach, which means the data will no longer need to be pre-treated with methods like PCA or CorrCA prior to parameter estimation by LSC.

Simulation studies showed that the number of components required to adequately denoise the tTACs could vary depending on factors such as the radiotracer or noise level. Therefore, for PCA and CorrCA, further studies could be to explore the methods to determine the number of components to be retained to denoise the data. This is especially worth a while for CorrCA since it is showing the smallest variations (in simulation data) and the highest contrasts (in clinical data); the aim will be to clearly reflect these observations in the displayed parametric images. Investigation and comparison to other tTACs denoising methods would also make good grounds for further studies.

Appendix A

Reducing the least-squares cubic equation to commonly known regression methods under specific assumptions

Let us consider the earlier expression of the LSC slope in Eq. (3.48) (Chapter 3) as written below,

$$b = \frac{\sum W_i^2 \left(\frac{U_i}{w(Y_i)} + \frac{bV_i}{w(X_i)} - \frac{r_i V_i}{\alpha_i} \right) V_i}{\sum W_i^2 \left(\frac{U_i}{w(Y_i)} + \frac{bV_i}{w(X_i)} - \frac{br_i U_i}{\alpha_i} \right) U_i}. \tag{A.1}$$

This solution is general, and various other regression methods, such as OLS and ODR, can be obtained only as special cases of this equation. As demonstrated in [13], the sections below explore through some of the special cases which reduce to some of the commonly known regression methods.

Below is an earlier expression in Eq. (3.35), and here it is noted for eased referencing.

$$W_i = \frac{w(X_i)w(Y_i)}{b^2 w(Y_i) + w(X_i) - 2br_i\alpha_i}. \tag{A.2}$$

A.1 Only Y_i is subject to errors (i.e., X_i is error-free):

In this case, $w(X_i) \to \infty$, and $r_i \to 0$. The expression in Eq. (A.2) then reduces to $W_i = w(Y_i)$; and thus Eq. (A.1) solves to,

$$b = \frac{\sum w(Y_i)U_iV_i}{\sum w(Y_i)U_i^2}. \tag{A.3}$$

This solution is equivalent to the ussual weighted OLS regression of Y on X. Removing the weight component, Eq. (A.3) then reduces to the expression derived for OLS in Eq. (2.56) as,

$$\begin{aligned} b &= \frac{\sum U_iV_i}{\sum U_i^2} \\ &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \end{aligned} \tag{A.4}$$

A.2 Only X_i is subject to errors (i.e., Y_i is error-free):

In this case, $w(Y_i) \to \infty$, and $r_i \to 0$. The expression in Eq. (A.2) then reduces to $W_i = \dfrac{w(X_i)}{b^2}$; and thus Eq. (A.1) solves to,

$$b = \frac{\sum w(X_i)V_i^2}{\sum w(X_i)U_iV_i}. \tag{A.5}$$

This solution is equivalent to the ussual weighted OLS regression of X on Y, and only the X variable is being weighted. Removing the weight component, it reduces to the OLS equation below, which is obtained by considering only the measurement errors in the X variable;

$$\begin{aligned} b &= \frac{\sum V_i^2}{\sum U_iV_i} \\ &= \frac{\sum (Y_i - \bar{Y})^2}{\sum (X_i - \bar{X})(Y_i - \bar{Y})} \end{aligned} \tag{A.6}$$

A.3 Errors present in both X and Y variables:

If we assume that $w(X_i) = w(Y_i) = c$ (constant) such that, $\dfrac{w(X_i)}{w(Y_i)} = 1$, and $r_i = 0$, W_i in Eq. (A.2) becomes,

$$W_i = \frac{c}{b^2 + 1}. \tag{A.7}$$

The expression in Eq. (A.1) then reduces to,

$$\begin{aligned} b &= \frac{\sum \left(\frac{a}{b^2+1}\right)^2 \left(\frac{U_i}{a} + \frac{bV_i}{a}\right) V_i}{\sum \left(\frac{a}{b^2+1}\right)^2 \left(\frac{U_i}{a} + \frac{bV_i}{a}\right) U_i}, \\ &= \frac{\sum \left(U_iV_i + bV_i^2\right)}{\sum \left(U_i^2 + bU_iV_i\right)} \end{aligned} \tag{A.8}$$

which gives,

$$b^2 \sum U_i V_i - b\left(\sum V_i^2 - \sum U_i^2\right) - \sum U_i V_i = 0. \tag{A.9}$$

Solving for b from Eq. (A.9) using the standard quadratic approach gives,

$$\begin{aligned} b &= \frac{(\sum V_i^2 - \sum U_i^2) + \sqrt{\left(\sum V_i^2 - \sum U_i^2\right)^2 + 4\left(\sum U_i V_i\right)^2}}{2\sum U_i V_i} \\ &= \frac{(S_{yy} - S_{xx}) + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}}. \end{aligned} \tag{A.10}$$

Equation (A.10) is identical to Eq. (2.62) for ODR, which is obtained by minimizing the sum of the squares of the perpendicular distances between the observed data points and the estimated straight line. Again it is shown here that this solution is obtained by assuming that the variances of the X and Y errors are equal, since in cases where the variances are known, the weights (which were here are assumed to be equal) are estimated as the inverse of the variances.

Appendix B

Noise model

The noise model utilised in this study is adopted from previous studies [6, 8, 11, 16]. As per the description in [11], given the radioactivity concentration, C_T, of the radiotracer in the ROI, if c is the count measured at mid-time t of the frame with "frame length" Δt, then we have that,

$$c = \frac{C_T \cdot d \cdot \Delta t}{a}, \tag{B.1}$$

where a is a scaling factor, and d is a decay factor given by, $d = \left(\frac{1}{2}\right)^{t/T}$, where T is the half-life of a radionuclide. The error in C_T can be expressed as,

$$\Delta C_T = \frac{a \cdot \Delta c}{d \cdot \Delta t}, \tag{B.2}$$

where, $\Delta c = \sqrt{c}$, is the error in c. So, the simulated noisy tTACs are then given by

$$\bar{C_T} = C_T(1 + cv \cdot G(0,1)), \tag{B.3}$$

where, $cv = \dfrac{\Delta C_T}{C_T} = \sqrt{\dfrac{a}{C_T \cdot d \cdot \Delta t}}$, is the variation coefficient, and $G(0,1)$ is a random number of a Gaussian distribution with mean 0, and standard deviation 1.

Appendix C

Contrast evaluation of simulation data

For contrast comparison between LSC–PCA and OLS–PCA in simulation data for ^{11}C-PiB, a set of hypothetical white matter regions were simulated —in addition to the gray matter regions simulated in Section 4.1.1. For ^{11}C-PiB Aβ-negative images, white matter regions can have high retention of the radiotracer due to nonspecific retention [25, 56], and slow accumulation and slower clearance [25]. Therefore, the white matter regions were simulated with the values of K_1 (0.190 mL cm^{-1}) and k_2 (0.100 min^{-1}) lower than those of the gray matters, and DVR values high than those of the gray matters. Thus, to simulate white matter with high DVR values than the garay matters, the 11 unit vector, $[0.4 : 0.055 : 0.95]$, was added to the 11 unit vector of the gray matter DVRs. This gave the lowest white matter DVR a value of 1.6727 (the lowest for gray matter is 1.2727), and the highest of 4.95 (the highest for gray matter is 4). The ratio of these gray and white matter DVR values are within the ranges reported in a previous study [57]. The scan period and number of frames for ^{11}C-PiB white matter simulations were set to be similar to those of the gray matter, and thus similar to those of clinical ^{11}C-PiB PET images used in this study.

Similar with gray matter simulations, a noise-free tTAC was formed for each white matter DVR value. Statistical noise was added to the noise-free tTACs to form noisy tTACs. For each noise-free tTAC, 1024 noisy tTACs were simulated as a slice of 32 $\times$ 32 voxels. Four methods, LSC–PCA, OLS–PCA, LSC, and conventional OLS-based LGA, were then used to estimate the DVR values from the noisy tTACs, and the results were compared.

The simulated gray and white matter data for ^{11}C-PiB were denoised together by PCA. For each data set (1024 simulated tTACS for each of the 11 true DVRs), DVR estimation from the noisy tTACs were performed thrice at 9 percentage variance level of PCA, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% (i.e., 27 times) for for LSC–PCA and OLS–PCA methods. This was to assess the performance of the two methods over a wide range of percentage variance, since in a real clinical practice the actual percentage variance to be retained in PCA denoising is not known. The three estimates at each variance level were differed by a slight change of $< 5\%$ in the magnitude of noise level. The magnitude of the noise level refers to the percentage ratio of the means of "standard deviation in the noisy tTACs to the mean of the noise-free tTAC".

To have equal set of data for all methods, the LSC and OLS estimates were performed 27 times at noise levels corresponding to those of the PCA-based methods. The mean and standard deviations were then calculated for 27 estimates for each 1024 simulated tTACs for each true DVR value.

Figure C.1 compares the results of contrast between the simulated gray and white matter regions. The comparisons include all calculated DVR values (three times at 9 percentage variance levels for LSC–PCA and OLS–PCA, and at 27 different noise levels for LSC and OLS). Statistical comparisons of these contrast are shown in table C.1, using the Friedman test [58] and Post-Hoc multiple comparison. A statistical significant difference is observed between LSC–PCA and all other methods. On the other hand, no statistical significant difference is observed between OLS–PCA and LSC.

The significant p value observed between LSC–PCA and OLS–PCA in simulation results could be an indication that the same can be obtained for clinical data. This can be investigated with a larger cohort of clinical ^{11}C-PiB brain data.

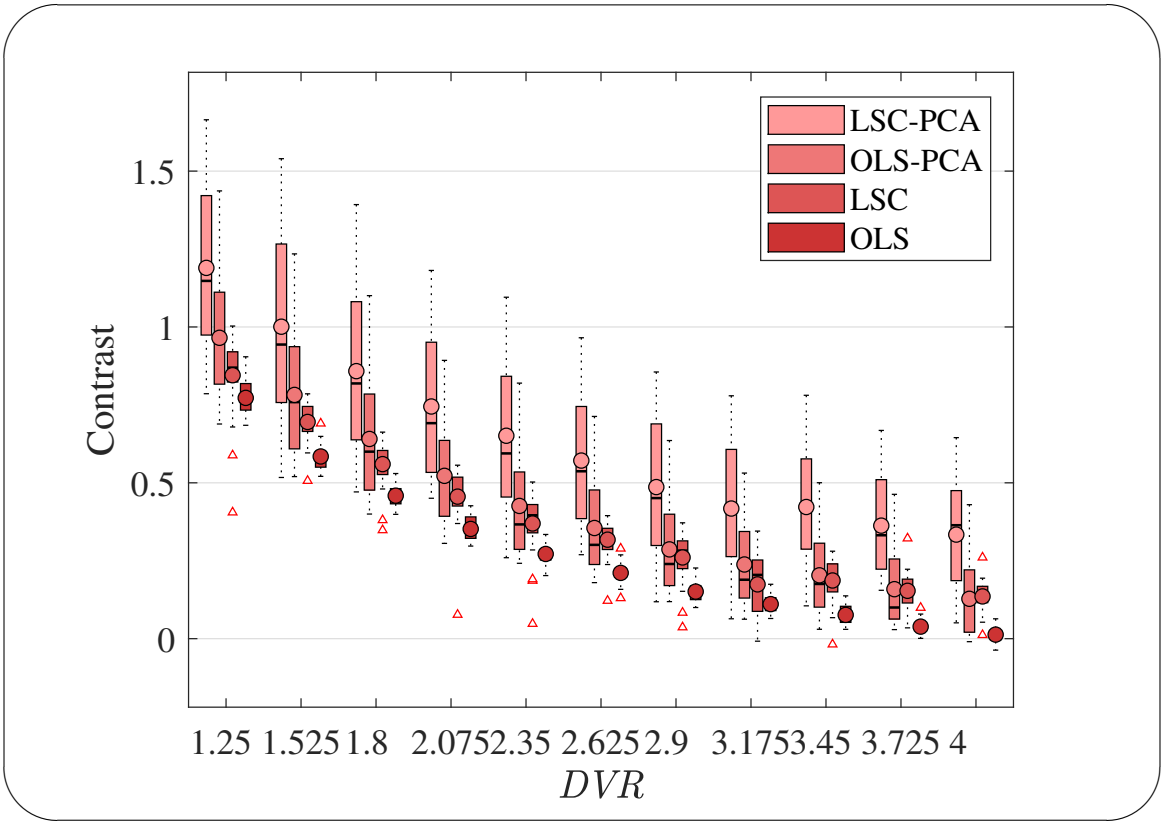


Figure C.1: *Box plots of the DVR contrast between the hypothetical simulated gray and white matter regions, calculated for all DVR values estimated at different percentage variance levels for LSC–PCA and OLS–PCA, and at 27 different noise levels for LSC and OLS. The circles in the box plots denote the means, whereas the horizontal lines denote the medians.*

Table C.1: Friedman test and Post-Hoc multiple comparisons of the p-values of the contrast values of simulated ^{11}C-PiB data plotted in figure C.1. Tabulated numerical values correspond to the p-values between the two methods in the corresponding rows and columns.

Friedman Test (p value $<$ 0.001)			
Post-Hoc Multiple Comparisons p-values			
	OLS-PCA	LSC	OLS
LSC–PCA	0.041	0.022	$<$ 0.001
OLS–PCA		0.804	$<$ 0.009
LSC			0.018

Bibliography

[1] D. L. Bailey, M. N. Maisey, D. W. Townsend, and P. E. Valk, *Positron Emission Tomography: Basic Sciences*. Springer, 2005.

[2] J. Logan, J. S. Fowler, N. D. Volkow, A. P. Wolf, S. L. Dewey, D. J. Schlyer, R. R. MacGregor, R. Hitzemann, B. Bendriem, S. J. Gatley, *et al.*, "Graphical analysis of reversible radioligand binding from time—activity measurements applied to [N-^{11}C-methyl]-(–)-Cocaine PET studies in human subjects," *Journal of Cerebral Blood Flow & Metabolism*, vol. 10, no. 5, pp. 740–747, 1990.

[3] J. Logan, J. S. Fowler, N. D. Volkow, G.-J. Wang, Y.-S. Ding, and D. L. Alexoff, "Distribution volume ratios without blood sampling from graphical analysis of pet data," *Journal of Cerebral Blood Flow & Metabolism*, vol. 16, no. 5, pp. 834–840, 1996.

[4] M. Slifstein and M. Laruelle, "Effects of statistical noise on graphic analysis of pet neuroreceptor studies," *Journal of Nuclear Medicine*, vol. 41, no. 12, pp. 2083–2088, 2000.

[5] J. Logan, "A review of graphical methods for tracer studies and strategies to reduce bias," *Nuclear Medicine and Biology*, vol. 30, no. 8, pp. 833–844, 2003.

[6] Y. Kimura, M. Naganawa, M. Shidahara, Y. Ikoma, and H. Watabe, "PET kinetic analysis—pitfalls and a solution for the logan plot," *Annals of Nuclear Medicine*, vol. 21, no. 1, pp. 1–8, 2007.

[7] A. Abi-Dargham, D. Martinez, O. Mawlawi, N. Simpson, D.-R. Hwang, M. Slifstein, S. Anjilvel, J. Pidcock, N.-N. Guo, I. Lombardo, *et al.*, "Measurement of striatal and extrastriatal dopamine d1 receptor binding potential with [^{11}C] NNC 112 in humans: validation and reproducibility," *Journal of Cerebral Blood Flow & Metabolism*, vol. 20, no. 2, pp. 225–243, 2000.

[8] J. Logan, J. S. Fowler, N. D. Volkow, Y. S. Ding, G.-J. Wang, and D. L. Alexoff, "A strategy for removing the bias in the graphical analysis method," *Journal of Cerebral Blood Flow & Metabolism*, vol. 21, no. 3, pp. 307–320, 2001.

[9] M. Ichise, H. Toyama, R. B. Innis, and R. E. Carson, "Strategies to improve neuroreceptor parameter estimation by linear regression analysis," *Journal of Cerebral Blood Flow & Metabolism*, vol. 22, no. 10, pp. 1271–1281, 2002.

[10] M. Ichise, J.-S. Liow, J.-Q. Lu, A. Takano, K. Model, H. Toyama, T. Suhara, K. Suzuki, R. B. Innis, and R. E. Carson, "Linearized reference tissue parametric imaging methods: application to [^{11}C] DASB positron emission tomography studies of the serotonin transporter in human brain," *Journal of Cerebral Blood Flow & Metabolism*, vol. 23, no. 9, pp. 1096–1112, 2003.

[11] J. Varga and Z. Szabo, "Modified regression model for the logan plot," *Journal of Cerebral Blood Flow & Metabolism*, vol. 22, no. 2, pp. 240–244, 2002.

[12] D. York, N. M. Evensen, M. L. Martínez, and J. De Basabe Delgado, "Unified equations for the slope, intercept, and standard errors of the best straight line," *American Journal of Physics*, vol. 72, no. 3, pp. 367–375, 2004.

[13] D. York, "Least-squares fitting of a straight line," *Canadian Journal of Physics*, vol. 44, no. 5, pp. 1079–1086, 1966.

[14] D. York, "Least squares fitting of a straight line with correlated errors," *Earth and Planetary Science Letters*, vol. 5, pp. 320–324, 1968.

[15] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.

[16] A. Joshi, J. A. Fessler, and R. A. Koeppe, "Improving pet receptor binding estimates from logan plots using principal component analysis," *Journal of Cerebral Blood Flow & Metabolism*, vol. 28, no. 4, pp. 852–865, 2008.

[17] Y. Kimura, M. Senda, and N. M. Alpert, "Fast formation of statistically reliable FDG parametric images based on clustering and principal components," *Physics in Medicine & Biology*, vol. 47, no. 3, p. 455, 2002.

[18] J. Axelsson and J. Sörensen, "The 2d hotelling filter-a quantitative noise-reducing principal-component filter for dynamic PET data, with applications in patient dose reduction," *BMC Medical Physics*, vol. 13, no. 1, pp. 1–17, 2013.

[19] L. C. Parra, S. Haufe, and J. P. Dmochowski, "Correlated components analysis-extracting reliable dimensions in multivariate data," *arXiv preprint arXiv:1801.08881*, 2018.

[20] P. K. Shigwedha, T. Yamada, K. Hanaoka, K. Ishii, Y. Kimura, and Y. Fukuoka, "A strategy to account for noise in the x-variable to reduce underestimation in logan graphical analysis for quantifying receptor density in positron emission tomography," *BMC Medical Imaging*, vol. 20, no. 1, pp. 1–8, 2020.

[21] H. Watabe, Y. Ikoma, Y. Kimura, M. Naganawa, and M. Shidahara, "PET kinetic analysis—compartmental model," *Annals of Nuclear Medicine*, vol. 20, no. 9, pp. 583–588, 2006.

[22] A. Joshi, *Improved quantitative methods for multiple neuropharmacological non-invasive brain PET studies*. PhD thesis, University of Michigan, 2008.

[23] E. Miele, G. P. Spinelli, F. Tomao, A. Zullo, F. De Marinis, G. Pasciuti, L. Rossi, F. Zoratto, and S. Tomao, "Positron emission tomography (PET) radiotracers in oncology–utility of ^{18}F-fluoro-deoxy-glucose (FDG)-PET in the management of patients with non-small-cell lung cancer (NSCLC)," *Journal of Experimental & Clinical Cancer Research*, vol. 27, no. 1, pp. 1–10, 2008.

[24] S. Yu, "Review of ^{18}F FDG synthesis and quality control," *Biomedical Imaging and Intervention Journal*, vol. 2, no. 4, 2006.

[25] W. E. Klunk, H. Engler, A. Nordberg, Y. Wang, G. Blomqvist, D. P. Holt, M. Bergström, I. Savitcheva, G.-F. Huang, S. Estrada, *et al.*, "Imaging brain amyloid in alzheimer's disease with pittsburgh compound-b," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 55, no. 3, pp. 306–319, 2004.

[26] 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," *Journal of Cerebral Blood Flow & Metabolism*, vol. 25, no. 11, pp. 1528–1547, 2005.

[27] J. J. Vaquero and P. Kinahan, "Positron emission tomography: current challenges and opportunities for technological advances in clinical and preclinical imaging systems," *Annual Review of Biomedical Engineering*, vol. 17, pp. 385–414, 2015.

[28] A. K. Shukla and U. Kumar, "Positron emission tomography: An overview," *Journal of Medical Physics/Association of Medical Physicists of India*, vol. 31, no. 1, p. 13, 2006.

[29] E. D. Morris, C. J. Endres, K. C. Schmidt, B. T. Christian, R. F. Muzic, and R. E. Fisher, "Kinetic modeling in positron emission tomography," *Emission Tomography: The Fundamentals of PET and SPECT. Academic, San Diego*, 2004.

[30] R. Koeppe, V. Holthoff, K. Frey, M. Kilbourn, and D. Kuhl, "Compartmental analysis of [^{11}C] flumazenil kinetics for the estimation of ligand transport rate and receptor distribution using positron emission tomography," *Journal of Cerebral Blood Flow & Metabolism*, vol. 11, no. 5, pp. 735–744, 1991.

[31] Y. Ikoma, H. Watabe, M. Shidahara, M. Naganawa, and Y. Kimura, "PET kinetic analysis: error consideration of quantitative analysis in dynamic studies," *Annals of Nuclear Medicine*, vol. 22, no. 1, pp. 1–11, 2008.

[32] M. Ichise, J. R. Ballinger, H. Golan, D. Vines, A. Luong, S. Tsai, and H. F. Kung, "Noninvasive quantification of dopamine d2 receptors with iodine-123IBF [SPECT," *Journal of Nuclear Medicine*, vol. 37, no. 3, pp. 513–520, 1996.

[33] M. Yaqub, N. Tolboom, R. Boellaard, B. N. van Berckel, E. W. van Tilburg, G. Luurtsema, P. Scheltens, and A. A. Lammertsma, "Simplified parametric methods for [^{11}C] pib studies," *Neuroimage*, vol. 42, no. 1, pp. 76–86, 2008.

[34] R. B. Innis, V. J. Cunningham, J. Delforge, M. Fujita, A. Gjedde, R. N. Gunn, J. Holden, S. Houle, S.-C. Huang, M. Ichise, *et al.*, "Consensus nomenclature for in vivo imaging of reversibly binding radioligands," *Journal of Cerebral Blood Flow & Metabolism*, vol. 27, no. 9, pp. 1533–1539, 2007.

[35] M. A. Mintun, M. E. Raichle, M. R. Kilbourn, G. F. Wooten, and M. J. Welch, "A quantitative model for the in vivo assessment of drug binding sites with positron emission tomography," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 15, no. 3, pp. 217–227, 1984.

[36] L. Farde, L. Eriksson, G. Blomquist, and C. Halldin, "Kinetic analysis of central [^{11}C] raclopride binding to d2-dopamine receptors studied by PET—a comparison to the equilibrium analysis," *Journal of Cerebral Blood Flow & Metabolism*, vol. 9, no. 5, pp. 696–708, 1989.

[37] Y. Zhou, W. Ye, J. R. Brašić, A. H. Crabb, J. Hilton, and D. F. Wong, "A consistent and efficient graphical analysis method to improve the quantification of reversible tracer binding in radioligand receptor dynamic pet studies," *Neuroimage*, vol. 44, no. 3, pp. 661–670, 2009.

[38] D. Feng, S.-C. Huang, Z. Wang, and D. Ho, "An unbiased parametric imaging algorithm for nonuniformly sampled biomedical system parameter estimation," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 512–518, 1996.

[39] X. Dai, Z. Chen, and J. Tian, "Performance evaluation of kinetic parameter estimation methods in dynamic FDG-PET studies," *Nuclear Medicine Communications*, vol. 32, no. 1, pp. 4–16, 2011.

[40] R. Wehr and S. R. Saleska, "The long-solved problem of the best-fit straight line: Application to isotopic mixing lines," *Biogeosciences*, vol. 14, no. 1, pp. 17–29, 2017.

[41] J. Fox, *Applied regression analysis and generalized linear models*. Sage Publications, 3 ed., 2015.

[42] J. M. Wooldridge, *Introductory Econometrics: A Modern Approach*. Cengage Learning, 6 ed., 2015.

[43] W. M. Stock J, *Introduction to Econometrics*. Pearson Addison Wesley, 2008.

[44] P. T. Boggs and J. E. Rogers, "Orthogonal distance regression," *Contemporary Mathematics*, vol. 112, pp. 183–194, 1990.

[45] R. Carroll and D. Ruppert, "The use and misuse of orthogonal regression in linear errors-in-variables models," *The American Statistician*, vol. 50, no. 1, pp. 1–6, 1996.

[46] R. Deakin and M. Hunter, "Fitting a line of best fit to correlated data of varying precision." `http://www.mygeodesy.id.au/documents/Fitting\%20a\%20straight\%20line.pdf`, 2017. Accessed: 2019-05-27.

[47] W. E. Deming, *Statistical Adjustment of Data*. Wiley, 1943.

[48] N. Shammas, "Hp 39gii regression: Part iv least-squares relative error regression." `http://www.namirshammas.com/HP39gII/HP39gII\_part4.pdf`, 2013. Accessed: 2018-12-26.

[49] C. Tofallis, "Least squares percentage regression," *Journal of Modern Applied Statistical Methods*, vol. 7, no. 2, pp. 526–534, 2008.

[50] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied regression analysis: a research tool*. New York: Springer, 2 ed., 2001.

[51] A. Cottrell, "Regression analysis: basic concepts." `http://users.wfu.edu/cottrell/ecn215/regress.pdf`. Accessed: 2018-12-26.

[52] C. J. Endres, B. Bencherif, J. Hilton, I. Madar, and J. J. Frost, "Quantification of brain μ-opioid receptors with [^{11}C] carfentanil: reference-tissue methods," *Nuclear Medicine and Biology*, vol. 30, no. 2, pp. 177–186, 2003.

[53] I. Jolliffe, "Principal component analysis, 2nd edn. series: Springer series in statistics, vol. XXIX, 487," *Springer, NY*, p. 28, 2002.

[54] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[55] T. Yamada, Y. Kimura, K. Fujii, S. Watanabe, T. Nagaoka, M. Nemoto, K. Hanaoka, H. Kaida, C. Hosokawa, and K. Ishii, "Noise reduction algorithm for amyloid imaging preserving image resolution–quantitative evaluation using clinical images," *Human Amyloid Imaging Conference*, 2018.

[56] R. A. Sperling, P. S. LaViolette, K. O'Keefe, J. O'Brien, D. M. Rentz, M. Pihlajamaki, G. Marshall, B. T. Hyman, D. J. Selkoe, T. Hedden, *et al.*, "Amyloid deposition is associated with impaired default network function in older persons without dementia," *Neuron*, vol. 63, no. 2, pp. 178–188, 2009.

[57] C. Hosokawa, K. Ishii, Y. Kimura, T. Hyodo, M. Hosono, K. Sakaguchi, K. Usami, K. Shimamoto, Y. Yamazoe, and T. Murakami, "Performance of 11c-pittsburgh compound b pet binding potential images in the detection of amyloid deposits on equivocal static images," *Journal of Nuclear Medicine*, vol. 56, no. 12, pp. 1910–1915, 2015.

[58] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the american statistical association*, vol. 32, no. 200, pp. 675–701, 1937.