STATISTICAL CONSIDERATIONS IN THE KINETIC ANALYSIS OF PET-FDG BRAIN TUMOUR STUDIES.

A thesis submitted to the National University of Ireland, Cork
for the degree of Doctor of Philosophy

David Hawe

Supervisor: Prof. Finbarr O'Sullivan
Head of department: Dr. Michael Cronin

Department of Statistics
College of Science, Engineering and Food Science
National University of Ireland, Cork

May 2016

NATIONAL UNIVERSITY OF IRELAND, CORK
DEPARTMENT OF
STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Science, Engineering and Food Science for acceptance a thesis entitled "**Statistical Considerations in the Kinetic Analysis of PET-FDG Brain Tumour Studies.**" by **David Hawe** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: May 2016

Supervisor: _____
Prof. Finbarr O'Sullivan

Readers: _____
Prof. John Hinde

_____
Dr. Michael Cronin

NATIONAL UNIVERSITY OF IRELAND, CORK

Date: **May 2016**

Author: **David Hawe**

Title: **Statistical Considerations in the Kinetic Analysis of PET-FDG Brain Tumour Studies.**

Department: **Statistics**

Degree: **Ph.D.** Convocation: **June** Year: **2016**

Signature of Author

Contents

List of Tables

List of Figures

Abstract

Dynamic positron emission tomography (PET) imaging can be used to track the distribution of injected radio-labelled molecules over time in vivo. This is a powerful technique, which provides researchers and clinicians the opportunity to study the status of healthy and pathological tissue by examining how it processes substances of interest. Widely used tracers include ^{18}*F-fluorodeoxyglucose, an analog of glucose, which is used as the radiotracer in over ninety percent of PET scans. This radiotracer provides a way of quantifying the distribution of glucose utilisation in vivo. The interpretation of PET time-course data is complicated because the measured signal is a combination of vascular delivery and tissue retention effects. If the arterial time-course is known, the tissue time-course can typically be expressed in terms of a linear convolution between the arterial time-course and the tissue residue function. As the residue represents the amount of tracer remaining in the tissue, this can be thought of as a survival function; these functions been examined in great detail by the statistics community. Kinetic analysis of PET data is concerned with estimation of the residue and associated functionals such as flow, flux and volume of distribution. This thesis presents a Markov chain formulation of blood tissue exchange and explores how this relates to established compartmental forms. A nonparametric approach to the estimation of the residue is examined and the improvement in this model relative to compartmental model is evaluated using simulations and cross-validation techniques. The reference distribution of the test statistics, generated in comparing the models, is also studied. We explore these models further with simulated studies and an FDG-PET dataset from subjects with gliomas, which has previously been analysed with compartmental modelling. We also consider the performance of a recently proposed mixture modelling technique in this study.*

Acknowledgements

I would like to thank first and foremost Finbarr for giving me the opportunity to do this thesis. It has been an enjoyable experience that I would undertake again in an instant. I thank Finbarr for his kindness, generosity, patience and understanding in helping me along the way.

I would like to thank my examiners Prof. John Hinde and Dr. Michael Cronin for their time and insightful questions and comments.

The animised image data in this thesis, was generated at the University of Washington PET Imaging Center on a GE Advance scanner. Use of these data was supported in part by the Science Foundation Ireland under the MI-2007 and 11/PI/1027 grants. Work on the data was supervised by Finbarr O'Sullivan, Department of Statistics, UCC. The use of these data outside of this Thesis, requires explicit permission from investigators at the University of Washington. Details can be had from Professor F. OSullivan.

I would also like to thank Michael Cronin, Supratik Roy and Eric Wolsztynski for allowing me to audit some of their classes and the members of the diagnostic imaging group at UCC, Jian Huang, Janet O'Sullivan, Sarah Murphy, Yanhui Zhang and Tian Mao for help I have received during this research. Thanks also to Francisco Hernandez-Fernandez, who deserves special mention.

Many friends have been made in UCC during the completion of this thesis. I would like to thank them all for making coming to college something to look forward to.

Finally I would like to thank my mother for all of the support throughout this work.

Chapter 1

Introduction

In recent times, positron emission tomography (PET) has become a vital tool in the medical imaging of cancer and as time progresses, its uses are growing into other fields. A PET scan is conducted by injecting a subject with a radiolabelled molecule (radiotracer) and imaging the distribution of this radiotracer in the body. Common radiotracers used in practice include radiolabelled glucose, which allows energy consumption in the body to be observed [62]. The majority of the data used in this thesis have been generated using a glucose radiotracer called ^{18}F-fluorodeoxyglucose (FDG). This is generated by replacing one hydroxyl group in glucose with an ^{18}F isotope [2].

PET is considered quite versatile as the number of substances which have been adapted to study various processes in the body, is large and continues to grow. PET quantifies the metabolism of substrates, which is how PET is distinguished from other imaging modalities that simply provide an image of structures in the subject being studied. As cancers are generally aggressive and grow at an alarming rate, one can often differentiate between healthy and cancerous tissue on an FDG-PET image. In simple terms, one can visualise glucose uptake (energy use) with an FDG-PET scan. This is important as glucose uptake is a prognostic indicator of many cancers due to the accelerated rate of cell growth associated with some malignancies. It is also worth highlighting that metabolism of the tracer can be used to confirm the presence of cancer but further work is still being conducted to examine the way tracers are absorbed by tissues.

There are however a number of disadvantages to PET. Firstly, the cost of doing

PET imaging is substantial compared to other medical imaging procedures. It is extremely labour intensive, the radiotracers have to be manufactured using complicated and expensive equipment and once the radiotracer is manufactured, it is subject to rapid deterioration, dependent on its half-life. The half-lives for the tracers used vary with the most widely used tracers in PET imaging have half lives of between 2 minutes and 108 minutes. It is also the case that the clarity of the images (resolution) is dependent on the dose of the tracer taken by the subject, however increasing the dose results in increased toxicity. For this reason, anatomical detail (resolution) is not as clear on a PET scan as say a magnetic resonance (MR) scan. Another issue in PET is that when one looks at a PET image, it presents a combination of flow and retention. To obtain a clearer picture of what is occurring in the tissue, it is important to separate these two components. It is this problem which leads to the work conducted in this thesis on PET kinetic analysis. This field is discussed by Phelps *et al.* [63] which introduces the '*measurement of local cerebral glucose metabolic rate in humans*', using FDG.

This first chapter of this thesis provides an introduction to the basic science behind PET and medical imaging and is organised as follows: In Section 1.1, a review of some of the uses of PET in research and clinical practice are presented, with a view to highlighting the motivation behind this work. In Section 1.2, a description of other fields of medical imaging are reviewed. This will not be a comprehensive treatment of medical imaging but just a brief introduction, designed to give the reader an idea of what medical imaging is and how different imaging techniques are used together to answer the questions clinicians seek to answer. For more details the reader is referred to the literature. This naturally leads to an outline of PET, the PET scanning procedure, the data analysis aspects of PET and a discussion of the software used in this modelling, which are presented in Section 1.3. The contribution of this thesis is then discussed in Section 1.4. This includes a brief discussion on brain imaging with PET, PET residue analysis, compartmental models and nonparametric regression. We conclude the chapter with some discussion in Section 1.5.

1.1 Motivation

PET scanning has become standard practice for cancer patients in the developed world. The FDG tracer is the backbone of PET cancer imaging (as it is used in over ninety percent of PET scans), however others have been developed with specific cancers in mind. For example ^{18}F-fluorocholine imaging has shown some promise in prostate cancers [55]. In clinical cancer care PET can be thought of as having mainly three distinct uses; (i) diagnosis and staging, (ii) treatment response and (iii) recurrence assessment, which are discussed by Choi *et al.* and Eisenhauer *et al.* [12, 19], and also outline guidelines for judging response using PET.

Diagnosis and staging is determining whether disease is present or not and if it is present how far the disease has progressed. Although PET can be used to diagnose cancer, it is usually the case that cancer is initially detected by a different imaging modality before a PET scan is conducted due to the cost of a PET scan. PET can provide potentially better information on this front than traditional imaging modalities as it measures the functional processes, such as glucose uptake, that are taking place in tumours and lesions instead of just providing a picture of the anatomical structures in the field of view. PET can better identify issues such as necrosis in tumours, that other imaging techniques might not detect.

The response of cancerous tissues can be analysed following a course of treatment such as chemotherapy. If a mass remained in the subject, a PET scan can allow one to quantify if the treatment has had any benefit by assessing whether a cancerous region is using less glucose after treatment. A cancerous mass may not have decreased in size but may be more active or may have increased in size and be less active. PET can provide this information, while magnetic resonance (MR) or x-ray computer tomography (CT) scans typically provide the shape and size of the mass. It is worth noting, however that there are exceptions to this such as functional MR, which can provide information on bodily functions such as blood volume.

PET scans can be conducted on the entire body of a subject and so are ideal for recurrence assessment. This means the primary and any potential secondary

sites (metastases) can be examined simultaneously.

Another advantage PET imaging presents is that generally it can be used very cost effectively. In the past, when a patient presented with a cancerous mass there was often very little alternative but to carry out exploratory surgery or a biopsy to determine if the mass was benign or malignant. This takes a toll on a patients well-being and in addition to the risk of releasing cancerous cells that might get distributed elsewhere in the body, and also has a high monetary cost. However, nowadays a PET scan, which is relatively inexpensive compared to exploratory surgery can be carried out and almost as much, or in some cases more, information obtained from it.

1.2 Medical Imaging

Although PET is the focus of this thesis, it is not used in isolation. A patient who undergoes a PET scan will likely undergo some other imaging procedures such as a CT [17] scan using a device similar to the GE advance PET/CT scanner shown in Figure 1.1. Other subjects may undergo an MR scan [44, 51] along with a PET scan.

Figure 1.1: *General Electric Advance PET Scanner [11].*

PET studies which involve other imaging modalities will be presented later in this thesis and so some of the most widely used medical imaging procedures are

discussed here before discussing PET. We begin by explaining what we mean by medical imaging. This is useful as it sets the stage for where PET imaging falls in the clinicians assortment of tools. A medical imaging procedure is the process of generating pictures of the human body for clinical purposes (medical procedures seeking to reveal, diagnose or examine disease) or medical science (including the study of normal anatomy and physiology) [8].

If you meet an ordinary person on the street and mention medical imaging to them, they will naturally think of an X-ray. An X-ray is a form of electromagnetic radiation. When the term is used it often refers to a medical imaging procedure, discussed by Spiegel [72], which involves aiming a source of X-rays at a subject and examining how much of these X-rays are absorbed by the body or, more accurately, how much of the source is attenuated by the body. This technique is very common in a modern hospital and it can be used to diagnose ailments from broken bones to lung cancer. However, the method is not perfect as it involves exposing the subject to a relatively small dose of radiation and the image produced is two dimensional and of low resolution. The downsides of this procedure is offset by the ease and cost effectiveness of the technique.

When the detail produced by an X-ray image is not sufficient, X-ray computer tomography (CT) can be used. This technique involves a larger dose of X-rays to produce a more detailed two or three dimensional image of the whole body or part of it. Similar to an X-ray, the CT image is generated by examining the capability of the ROI to attenuate X-rays. However, as the dose is higher, an image with improved anatomical detail (higher resolution) can be obtained. CT has a wide variety of uses. To illustrate this point we simply state that there are over sixty million CT scans conducted in the United States annually [17]. Similar rates for CT scanning are likely to occur in Europe but this information is not readily available. These scans are conducted in areas varying from emergency medicine to cancer diagnosis. CT scans are relatively user friendly for both the patient and physician [17]. It does however have a question mark hanging over its safety. The subject being studied is exposed to a high dose of radiation during the scan. A single CT scan can give a radiation dose as much as 400 chest X-rays and it is estimated by some authors that, in healthy subjects, a CT scan gives the subject a one in eighty chance of developing cancer. This is disputed amongst

the radiography community with many arguments put forward by both sides. Different perspectives of their arguments are presented in [15, 17, 49, 69]. CT scanning has other uses such as creating an attenuation image for PET scanning. This is done using a low dose CT scan to minimise the toxicity that the subject being imaged is exposed to.

MR imaging is described by Brown [44] as a technique where strong magnetic fields are used to spin the electrons contained within a subject. This movement excites hydrogen atoms which give off a specific radio frequency. The magnetic field is switched on and off many times during the scan and the rate at which the atoms stop emitting the radio frequency of interest, generated by the spinning is used to distinguish between different types of tissue. These scans provide exquisite anatomical detail and are sometimes conducted alongside PET so that a high resolution image of the area under study as well as the activity measured by the PET scan can be examined by clinicians. This method has less risk associated with it than CT imaging and is frequently in place of CT imaging in the case of pregnant women.

Ultrasonography [67] is another medical imaging technique, which is used mainly for imaging soft tissues and the unborn. This technique is similar to how bats navigate as it directs pulses of ultrasound at the ROI in the subject being studied and this is then reflected back to the source where it is detected and an image is constructed. There are advantages and disadvantages to this type of imaging. Firstly, there is no harmful radiation involved and this is why ultrasonography is used to image the unborn. Secondly, it is also cheap. However, it has drawbacks, for example, it is not possible to image parts of the body behind dense tissue such as bone. A recent review of ultrasonography is available in [67].

Electrocardiography (EKG) [32] and electroencephalography (EEG) [53] are techniques which are considered as part of medical imaging, but are quite different to most medical imaging techniques. In the case of EEG, detectors are placed on the subjects head to detect electrical impulses in the brain. This provides an insight into how the brain works. With improvements in other imaging modalities, EEG is not as widely used as it was previously. However, it provides excellent temporal resolution, which other imaging procedures cannot as the data have a continuous form. This is due to the fact that EEG continuously measures the

electrical activity in the brain. EKG is similar to EEG, but instead the electrodes are attached to the subjects chest and the area of interest is usually the heart.

Nuclear medicine is a field within medical imaging which detects radioisotope emissions from radiotracers injected, ingested, inhaled or consumed in some way by the subject to produce images of functional body processes. PET is a nuclear medicine imaging technique which has already been mentioned and will be discussed in detail shortly. It is worth noting that nuclear imaging is not limited to PET. Single photon emission computer tomography (SPECT) was widely used for purposes similar to which a PET scanner is used for in a modern clinical setting [81]. SPECT uses substances such as Iodine-123 for tumour imaging. This type of scanning tool was in a sense a predecessor to PET and is based on similar principles as it is a nuclear medicine imaging technique, but SPECT is based on a single photon emission unlike PET, where two coincidental photons are detected by the scanner.

While this work focuses on PET, it is obvious that these other techniques are also vital in medical science and generally PET scans are conducted alongside these techniques and not in isolation. We will now discuss PET imaging procedures in more detail.

1.3 PET

PET is a medical imaging procedure that produces a three-dimensional image or picture of functional processes in the body [75]. The system detects pairs of gamma rays emitted indirectly by a positron-emitting radioisotope (radiotracer), which is introduced into the body on a biologically active molecule [40]. Three-dimensional images of tracer concentration within the body are then constructed by computer analyses. As PET measures information which involves metabolism of the radiotracer *in vivo*, it adds a different type of information to the structural information available from say an MR scan.

The concept of emission tomography was introduced by Kuhl and Edwards [40] in the late 1950's. Annihilation is a key part of emission tomography. Once the positron is emitted from the subject being studied, it meets an electron and these two particles destroy each other as they are anti particles. This reaction

produces two photons which move away from each other at 180 degrees and it is these photons the scanner detects. This work was then improved further by Phelps and Ter-Pogossian [76] at Washington University and Sweet and Brownell [73] at Massachusetts General Hospital who were responsible for the first use of an annihilation for medical imaging. These technologies have been refined and improved over the decades to give the equipment which forms the basis of the modern PET scanners, which are used to generate images such as the pediatric patient imaged with both PET and CT, which is shown in Figure 1.2. Where medical images are superimposed like this they are said to be coregistered. In this case a pediatric patient is imaged with ^{18}F FDG.

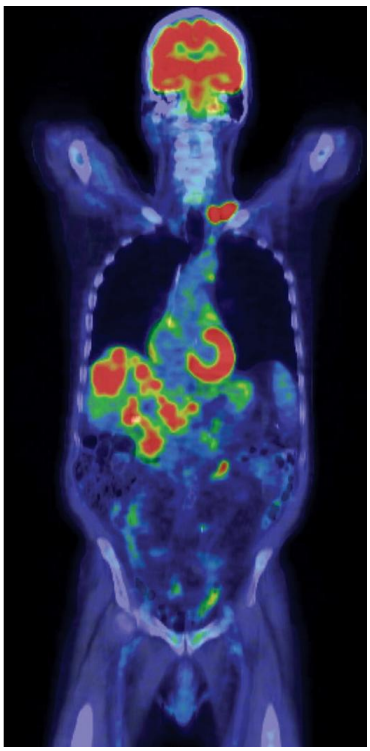


Figure 1.2: *PET-CT image of the head and torso a pediatric patient [7].*

Over the decades PET has advanced and in 2012 Wacholtz [78], published a report discussing the growing role for PET in clinical practice. The report highlights that until recently, PET had been envisioned and employed as a research tool, particularly in the study of neurophysiology. In the last few decades, however, PET is being more widely used in clinical settings.

The PET scanning procedure can be divided into three stages. Firstly, the

radiotracer must be manufactured for injection to the patient. Secondly, once this is done the patient can be placed in the scanner and the scan can begin. Finally, once the scan is completed the data from the scan has to be analysed. These three components of PET data generation and analysis will now be discussed.

1.3.1 Radiotracers

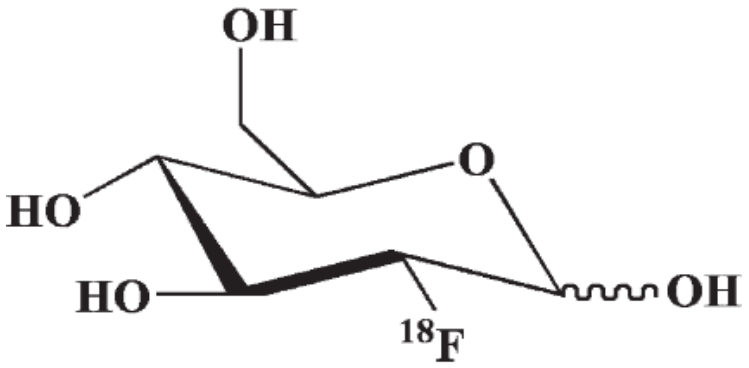


Figure 1.3: *Schematic of an FDG atom showing presence of* ^{18}F *[64].*

As was already mentioned, PET is conducted because a physician or scientist wants to know how a particular region of the body is using a substance of interest. To achieve this the molecule of interest is labelled with a positron emitting radionuclide and then given to the subject being studied which allows its distribution to be estimated *in vivo*. The rate of uptake of these tracers can be examined to determine if there is an abnormal use of them within an ROI. In the case of cancer, FDG (illustrated in Figure 1.3) can be used as an analogue for glucose, which is a marker for energy usage by tissue.

There are two complications that must be noted when using FDG. Firstly, FDG transforms into fluorodeoxyglucose-6-phosphate (FDG-6-P), which can revert back to FDG during the scan. This leads to complications in understanding the tracer uptake in tissue. Secondly, Figure 1.3 shows the structure of an FDG molecule and the readers attention is drawn to the presence of the ^{18}F isotope

in the diagram. FDG has the chemical formula $C_6H_{11}\ ^{18}FO_5$ while the chemical formula for glucose is $C_6H_{12}O_6$. One can see here an atom of oxygen and an atom of hydrogen are replaced by the ^{18}F atom. In a sense this is not ideal, however this limitation is not hugely significant as the information gained from an FDG scan has been shown to have an impact on diagnosis staging and recurrence assessment for lung, head and neck, colorectal, breast, ovarian, cervical, uterine, gastroesophageal, pancreatic, testicular and thyroid cancers, along with melanoma and lymphoma [21, 55]. Also, the fact that FDG has a long half-life of approximately 108 minutes makes it very useful in practice.

The relationship between how glucose and how FDG behave in tissue is the subject of much discussion in the literature [34]. Spence *et al.* [71] discuss this in the context of brain tumours by comparing FDG and ^{11}C glucose. FDG and glucose enter cells in the same way, however FDG is not metabolised in the same way as normal glucose or ^{11}C glucose. Although this is more practical as FDG has a longer half life than ^{11}C glucose, it is at the expense of gaining the information of how glucose behaves exactly. While ^{11}C glucose can be used instead of FDG, and it more accurately reflects how glucose behaves in the body, it is rarely used in a clinical setting due to its relatively short half-life, which makes it expensive.

This outline of radiotracers, should not be considered an in depth treatment but merely an introduction to the subject. For more detailed discussion the reader is referred to [31].

1.3.2 Scanning Procedure

Figure 1.1 shows a typical scanner in routine clinical use. This particular scanner is a GE advance PET scanner. One can see the bed like area at the front of the scanner, where the patient lies. The detector rings are contained in the large structure to the rear with a cylindrical internal shape. This process of conducting a PET scan using a scanner such as this is illustrated in Figure 1.4 for the case of a small animal study.

Depending on the type of study being conducted, a subject undergoing a PET imaging procedure will arrive at the scanner and may be injected with the tracer immediately or may be placed in the scanner sometime before injection.

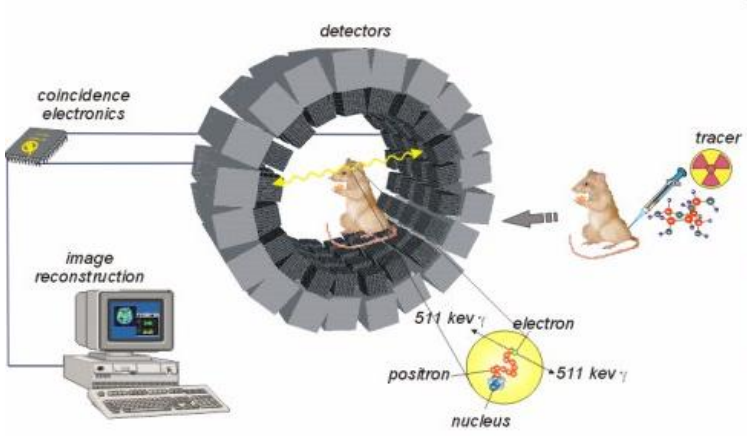


Figure 1.4: *Illustration of PET scan procedure [29].*

The differences arise because there are different half-lives for different tracers and because different types of scanning protocols are implemented to test different hypotheses.

It is worth mentioning again that PET scan measurements are made up of emissions over a time window. For example, a study might involve measuring the number of emissions over a five minute window. If the time-bin is very short, there may be one or no emission from a region over a time-bin. This is not very useful as the quality of the information obtained depends on the number of emissions. Having such a small number of emissions in a region of interest would lead to very noisy ROI data which would be difficult to analyse. Also, knowing that one emission came from a region in a very short period of time is of little practical use.

There are two different ways to gather the emission data used in practice and these are called static imaging and dynamic imaging. Static images consist of one single time-bin where the patient will generally wait a period of time between injection of the tracer and being placed in the scanner to allow the material being injected to distribute throughout the body and reach equilibrium. Here only one time-bin is used to generate a single PET image.

If more than one snapshot is taken, it is referred to as dynamic imaging. Here, the patient is imaged numerous times in order to generate a picture of how the

tracer is distributed throughout the body over time. The patient is most likely placed in the scanner prior to the injection of the radiotracer. Also, if more than one image is being taken while the patient is in the scanner and the time-bins are too long then not enough information about the changes over time is obtained.

In this case, the kinetics of how the tracer moves between blood and tissue are of interest. Arterial sampling using a catheter may be conducted before and during the scan to measure the volume of radiotracer in arterial blood as is the case in [51, 52, 71]. Having an arterial input function (AIF) is important for the analyses conducted on scan images and if it is not generated by arterial sampling, there are other methods which can be used to generate this data from the dynamic PET image [33, 56].

In spite of this drawback, there are many advantages to dynamic scanning. These include gaining a better understanding of how the tracer distributes throughout the body over time. It allows a better understanding of when an equilibrium occurs, i.e. if the tracer in tissue has reached a steady state. Another benefit is that if a new tracer is under study, the distribution of the tracer over time can provide information to researchers as to what protocol should be used for static imaging. It must be noted that static imaging is standard practice in a clinical setting. This is mainly to do with the time and financial costs associated with dynamic scanning and also that static imaging is usually sufficient for the questions clinicians seek to answer in practice.

For both dynamic and static imaging, the patient lies still in the scanner for the duration of the scanning procedure. The tracer within the patient will be emitting positrons. These will travel a 'tiny' distance (less than one millimeter) prior to meeting an electron. As the positron and electron are antiparticles, once they meet an annihilation occurs and two photons in the form of gamma rays are given off. These photons move away from each other in opposite directions (180° from each other). If these two photons are simultaneously detected by the detectors, we can infer that the annihilation must have occurred along the line connecting the two detectors. This line is referred to as the line of response (LOR). It is these photons coincidentally hitting the scintillators (material that fluoresces when it comes into contact with ionising radiation) contained in the detector rings of the scanner that the PET scan measures (the scanner ignores

non temporal pairs) [20]. From here the data can be reconstructed and sinograms (two dimensional matrices) containing the counts from the scan be generated [20]. Figure 1.5A (ellipse) illustrates a subject undergoing PET imaging. The center

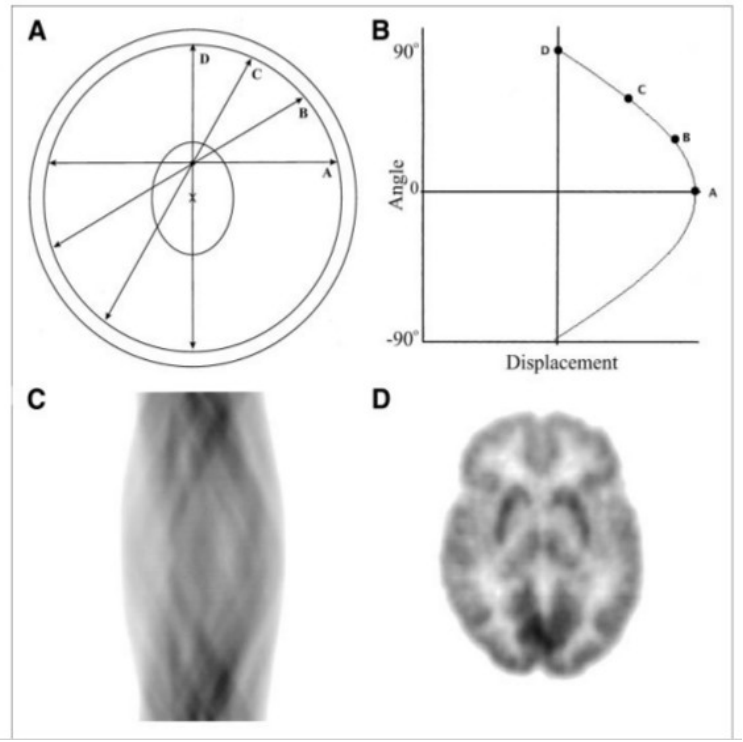


Figure 1.5: *Illustration of PET scan reconstruction [20].*

of the PET gantry is noted in the figure with an *'x'*. A tracer which is absorbed by the tumour has been administered to the patient. An emission occurs in the subject which leads to a pair of photons detected on the LOR A in Figure 1.5A. The LOR is characterised by the angle of orientation of the LOR and the shortest distance between the LOR and the center of the gantry. This is shown in Figure 1.5B where the angle of orientation ($0°$ in case of the LOR marked A) is plotted on the y-axis, and the shortest distance between the LOR and the center of the gantry is plotted on the x-axis. Three other LORs (B, C, and D) are shown for three other emissions. These points are also plotted in Figure 1.5B. If a large number of LORs are plotted from the same point (or pixel), the resulting graph is half of a sine wave turned on its side, as shown in Figure 1.5B. This is appropriately referred to as a sinogram.

As these data are made up of these sinograms, they need to be transformed

to Cartesian co-ordinates. Details of this transformation will be presented later in this volume, but an illustration of what an image looks like before and after this transformation are presented in Figure 1.5C and 1.5D respectively.

Technology has advanced recently and new PET scanners use a technique known as time of flight where the small time difference between the detection of the two photons is used to help estimate the location where the emission occurred, i.e. where on the line of response the emission occurred. This technology should continue to advance in years to come and lead to better quality images.

1.3.3 Structure of PET Data

As was mentioned in the previous section, once the counts have been stored by the scanner, these have to undergo a transformation to be in the form of coordinates and activity. The basis for this is the Radon transform [82]. Iterative expectation-maximization algorithms are now the preferred method of reconstruction. These algorithms compute an estimate of the likely distribution of annihilation events that led to the measured data, based on statistical principles [57]. The PET reconstruction procedure will be outlined and then a brief discussion on the post reconstruction data will be presented.

The Poisson Nature of Annihilation Detection

As has already been mentioned, the data is generated by detecting two photons being emitted from an annihilation occurring between a positron and an electron. We need to determine two pieces of information about the emissions. The location and the time that each detected annihilation occurred. Scanners are set up so that coincidental detections must be at approximately 180 degrees to each other and happen within nanoseconds of each other. If this does not happen the data are discarded.

In trying to gain an understanding of the distribution of the data one considers a Bernoulli trial where there are two possible outcomes, usually called success or failure. Consider an emission being detected to be success and an emission not being detected as failure. There are a large number of of annihilations but only a small proportion of these are actually detected and accepted by the scanner

(i.e. probability of detection is small). In this case the Binomial distribution (the distribution which describes the probability of having a number of successes or failures in Bernoulli trials) can be approximated by the Poisson distribution [66].

We start with a Poisson process with rate r. As the emissions are assumed independent, and can either be detected or not detected, we can categorise emissions as (those detected and those not detected): detected, type 1 with probability p and missed type 0 with probability $1-p$, where $p \in (0,1)$ is a parameter. We want to consider the type 1 and type 0 random points separately. For this reason, this new random process is sometimes referred to as splitting the original Poisson process. In some applications, the type 1 points are accepted while the type 0 points are rejected. Thus, the new random process is also referred to as thinning the original Poisson process [66].

This can be described more succinctly by saying if all the annihilations, which occur, are Poisson then the detected annihilations are also Poisson. That is to say that if $X \sim Poisson(\lambda)$ and Y is a subset of X, then Y is also Poisson as we have simply reduced the sample space or thinned the Poisson process. The fact that the data are Poisson will be useful later in this work when choosing weights for fitting models to the data. It is also useful to note that the sum of Poisson variables is also a Poisson variable.

Reconstructed Data

PET image reconstructions have been examined in [16, 35, 45, 55, 57, 77]. In spite of the fact that PET has been in use for decades, there is no canonical way to do reconstruction. Techniques used by different machines vary and include filtered back projection as well as expectation maximisation algorithms. The basis for these methods is CT reconstruction. As this thesis is primarily focused on analysing reconstructed data and this section is background material, we will only focus on the least squares reconstruction, which is known as filter back projection. We will also consider a simple two dimensional case rather than the three dimensional case, as the two dimensional case is generalisable to three dimensions and is much more understandable.

We have already stated that not every emission is detected by the scanner.

There will be a number of these photons attenuated by the tissue. This is similar to how an X-ray works. The attenuation characteristics of the tissue determine what proportion of emissions make it to the detectors. The probability of detection is a function of the line integral of attenuation on the LOR. This can be thought of as a sum of the attenuation $\gamma(s)$ at every point on the LOR

$$p_l = e^{-\int_l \gamma(s)ds}.$$

We also note that at each location s in the field of view the expected number of annihilations $\lambda(s)$ is Poisson. The collection of observed counts

$$N = (n_1, n_2, ...n_l),$$

are Poisson variables (thinned) with rates

$$(p_1\mu_1, p_2\mu_2, .., p_l\mu_l),$$

where p_i is the probability of detection.

We now consider a two dimensional image similar to what you would obtain from a particular slice and time-bin in a PET study. This could have say 128×128 voxels (three dimensional pixel). We have already seen $\mu_l = \int_l \lambda(s)ds$. We now express the data as a linear model i.e. $\mu_l = \int_l \lambda(s)ds = x_l^T\lambda$, where x_l is the integral along l. Values of x_l can be obtained by placing a point source at each voxel in the field of view and obtaining a measurement (nowadays this is approximated using CT images).

Let n_l^* be the attenuation corrected count $n_l^* = \frac{n_l}{p_l}$, and $N^* = (n_1^*, n_2^*, ...n_l^*)$. Then the form of the linear model is obtained as

$$N^* = X\lambda + \varepsilon,$$

where $X = (x_1, x_2, ...x_l)^T$ and ε is a zero mean error term without constant variance. That means the expectation of n_l^* is $x_l^T\lambda$ with variance $\frac{x_l^T\lambda}{p_l}$. Using weighted least squares squares an estimate for λ called $\hat{\lambda}$ is found as

$$\hat{\lambda} = (X^TX)^{-1}X^TN^*.$$

The use of least squares type methods has been the basis of reconstruction for some time. Although others have proposed alternative methods, which may be

more advantageous, discussing this is outside the scope of this work. For further details the reader is referred to [16, 35, 45, 55, 57, 77].

In practice, the filter back projection solution is smoothed by convolution with a Gaussian kernel. This can be thought of as multiplying the filter back projection by a matrix S_h.

$$\hat{\lambda}_h = S_h(X^T X)^{-1} X^T N^* = S_h r * b = r S_h * b,$$

where $*$ denotes a convolution, $r = (X^T X)^{-1}$ and $b = X^T N^*$. It is

Correcting the Data for Volume and Duration

A window of time where the number of emissions is counted from a PET scan can last from a few seconds to several minutes. There will likely be more emissions in an interval of say five minutes over one of five seconds. Accordingly when analysing a PET scan the time-bins can be thought of in terms of thinning the Poisson process. Dividing each measured value by the duration of its time-bin takes account for this. It is also necessary to take account of the volume being examined. For example, if one examines a large region and a small region there will be more emissions in the larger region as it has a larger volume. This is compensated for by using a unit of radioactivity such as the Becquerel (Bq) per unit volume. One Bq is the equivalent to one nucleus decay per second. In the case of PET the data are usually written as kilo-Becquerels per cubic centimeter or (KBq cm^{-3}).

Correcting the Data for Decay

In Boyce and di Prima [10] radioactivity is discussed in the context of differential equations. If one has a source of radiation N, which is decaying as follows:

$$\frac{dN}{dt} = \frac{1}{\tau} N, \quad \tau < 0$$

one can solve this equation as

$$N(t) = N_0 e^{\frac{t}{\tau}},$$

where $\frac{1}{\tau} < 0$ is the decay constant and, in the case of radioactive decay, it can be related to the half-life of the radioisotope by $t_{\frac{1}{2}} = \tau \log \frac{1}{2}$. This is easily found by solving

$$\frac{1}{2} = e^{\frac{t}{\tau}}$$

for t. The half-life plays a key role in deciding the dose administered to the subject being imaged. In the case of ^{15}O a large dose is injected as the half-life is approximately two minutes and injecting a smaller dose would not yield an image with useful resolution. In the case of FDG, which has a half-life of almost two hours, smaller doses are sufficient.

If we consider the PET scan counts, are gathered over a collection of time-bins whose intervals are defined by $(s, e]$, where s and e are the start and end times of the time-bin measured in seconds, we can convert this to decay corrected activity as follows

$$Z = \alpha \frac{e^{\frac{(s+e)\tau}{2}}}{v(e-s)} \hat{\lambda}_h = a\hat{\lambda}_h,$$

where v is volume and α is called the calibration factor, which is usually available from the scanner manufacturer.

Mean and Variance of PET Data

The linearity of the reconstruction allows us write down an expression for the mean and the variance of the reconstructed data Z.

$$E(Z) = aE(\hat{\lambda}_h) = aS_h(X^TX)^{-1}X^TE(N^*) = aS_h\lambda$$

$$Var(Z) = a^2S_h(X^TX)^{-1}X^TVar(N^*)X(X^TX)^{-1}S_h^T.$$

$E(N^*) = X\lambda$ and $Var(N^*)$ is a diagonal matrix with entries $\frac{x_l^T\lambda}{p_l}$. It is worth noting that if the $Var(N^*)$ is proportional to the identity, then

$$Var(Z) = a^2\sigma^2S_h(X^TX)^{-1}S_h^T.$$

It has been shown that if we let $C = aS_h\lambda$, the resolution limited expected decay corrected activity values, then we can represent the random deviation of Z from C by

$$Z - C \approx aC^{\frac{1}{2}}\sigma_h\varepsilon,$$

where ε is a standardised second order stationary process defined over the imaging domain, whose power spectrum is simply related to the resolution filter. Thus, reconstructed activities have a variance proportional to the mean, i.e.

$$Var(z_i) = \sigma^2 a^2 E(z_i).$$

This result allows us to use the quasi-Poison nature of the data to our advantage in modelling.

Analysis After Reconstruction

Once this process has been completed, it is extremely wasteful to simply look at (eyeball) the PET image and do no quantitative analysis. If one were to eyeball Figure 1.6, which shows the coronal, sagittal and transverse views of an FDG brain study, and do no further analysis, then the best information available from the scan may not be attained. This would be regrettable as the cost of a PET scan in Ireland varies between €1,206 and €2,500 [4, 23, 37] and given this significant cost, one would hope to get as much benefit from the procedure as possible. However if the financial cost was not an issue, then it would still make sense to do a quantitative analysis of the PET scan as information that can be derived by this analysis could have an impact on treatment and outcome. Such an analysis can determine if there is a higher volume of blood in one region over another and can answer some other questions, that could have an impact on treatment. In cancer, if tumours have developed their own vascular network, is of interest and this is an important prognostic indicator. Hence estimating blood volume is key here. Given the importance of the questions PET seeks to answer and the significant financial cost associated with doing a PET scan, all available useful information should be extracted from the scan.

The work in this thesis is based on dynamic scans where the kinetics of tracers over time are of interest. We consider a subject who has been imaged with PET. We may wish to know the activity in a particular region of the subject. The activity at each voxel and at each time point in the ROI can be calculated. This gives n time-courses where n is the number of voxels in the ROI. The number of time points is the number of snapshots taken by the scanner. As there may be a large number of these time-courses, it is customary to take a measure of

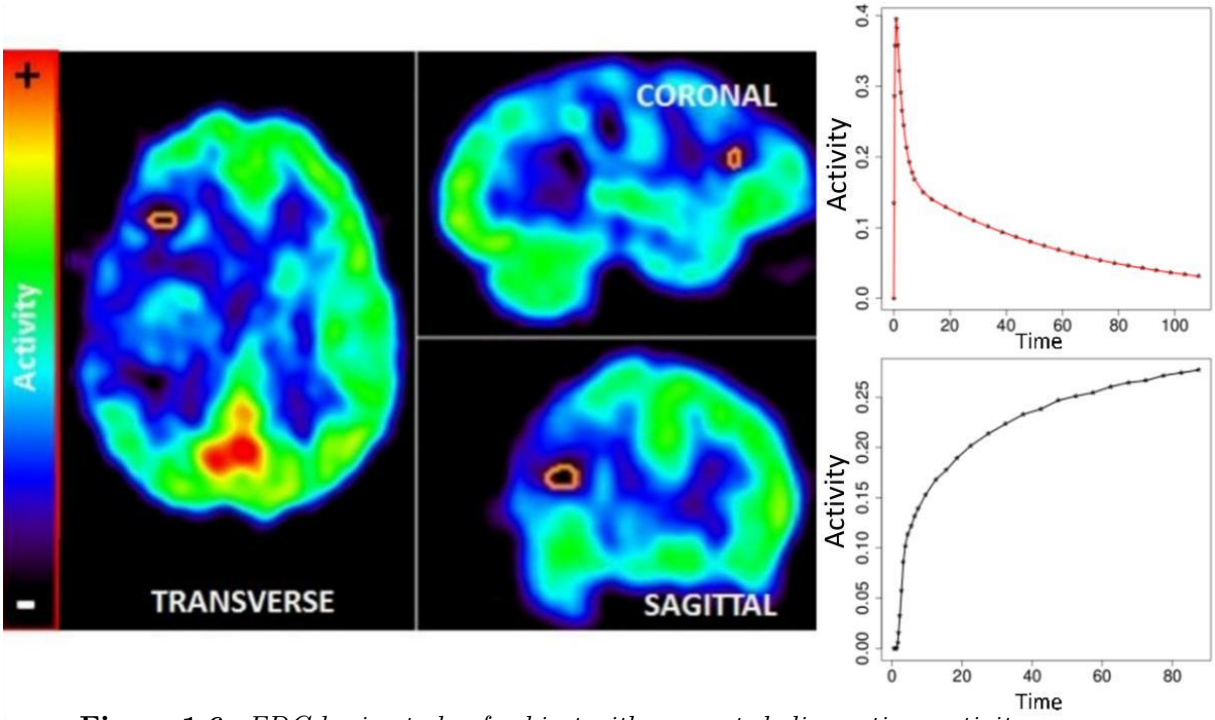


Figure 1.6: *FDG brain study of subject with a resected glioma,time activity curve and AIF are also shown.*

centrality of all these time-courses to represent the region as a whole. This is usually the mean, however in some datasets, with a small number of voxels, it may be more appropriate to use the median. Taking averages like this reduces the noise in the system. An example of an average PET time-course (called time activity curve) is shown in Figure 1.6. This is one of the key components required to begin kinetic analysis of PET data.

Kinetic analysis of PET data is the focus of this thesis. To aid in this task, the activity in the arterial blood is usually measured during the PET scan. An example of an AIF which is used in this work is also shown in Figure 1.6. This AIF has been generated by arterial sampling. This is where a catheter is inserted into an artery (usually radial artery) of the subject being imaged and blood is drawn from the patient. This is then examined to determine the activity in the blood. However, having to conduct arterial sampling is seen as a disadvantage by many clinicians. This has led to methods being developed to generate the radioactivity in arterial blood without arterial sampling by generating the AIF from the image, such as those proposed by Huang and O'Sullivan and O'Sullivan *et al.* [33, 58].

The software used in this work to view images is AMIDE [42] (A Medical Image Data Examiner), which is a free software tool used for multimodality medical image analysis. AMIDE allows PET, CT, SPECT and MR scans to be viewed, manipulated and ROI time course data to be exported. There are a number of tools provided by AMIDE such as cropping, Gaussian filtering, principal component analysis and perhaps most importantly the ability to draw ROIs on the images and subsequently extract these data. Three dimensional ROIs can be drawn using ellipsoids, elliptic cylinder's, boxes, isocontours, or by hand. Data can then be extracted from these ROIs and statistics can be generated for these regions. AMIDE was used to select ROIs that were of interest to be examined in this chapter. Later an alternative software package Alice will also be introduced, and this package serves a similar function.

The extracted data from the PET scan was then analysed using a common freely available statistical software system called R [65] version 2.14.1. R is a free software environment for statistical computing and graphics (R development Core Team, 2008). As well as having many built in functions, R also allows the user to write their own code to analyse data. These two pieces of software were vital in the work contained in this thesis.

1.4 Technical Contribution of this Thesis

The primary goal in kinetic analysis is to separate the flow and retention characteristics of the radiotracer. As has already been mentioned, the AIF is measured by arterial sampling in our case and this represents the activity in the arterial blood flowing around the body. What remains is the estimation of the amount of tracer in the tissue. We consider the tracer in tissue in terms of the residue function (which will be called $R(t)$ later in this work) and this is what we seek to estimate. This residue function, which describes the proportion of radiotracer remaining in tissue over time, allows quantities to be estimated which have prognostic value to the subject being imaged. For example, the initial value in the residue is flow and the final value represents the metabolic flux.

As dynamic PET data contains two components, flow and retention, this can be thought of as the atoms of radiotracer being in blood or in tissue. The fact that

the atoms can be in distinct states leads to compartmental models being used to examine the kinetics of the system. Those familiar with epidemiology will know of compartmental models for subjects in different states, such as susceptible, infected, or recovered. In Chapter 2, compartmental models used in kinetic analysis and Markov modelling will be reviewed. Also, a novel Markov formulation of the most widely used and well established compartmental model will be presented. The relationship between the two formulations will be established and shown to be equivalent in a limiting case. This Markov approach has not previously been studied in the context of PET data.

Nonparametric estimation is a statistical method that allows the form of a model to be varied based on the observed data. Nonparametric models differ from parametric models in that the structure of the model may not be known *a priori* but is estimated from the dataset. A nonparametric model is often employed in cases where data is complex or the distributions underlying the data are unknown. An example of a nonparametric estimator is a histogram, which is used to estimate a probability distribution. This can be improved on by using kernel smoothing. In this work, we will be interested in nonparametric estimators for the residue function in PET imaging [28, 38, 59, 60].

In Chapter 3 a nonparametric model for PET data is introduced. Techniques for model comparison will be presented in the case where models are nested or non-nested. In the case of non-nested models, a cross-validation approach is emphasised as traditional analysis of variance techniques may not be appropriate in this situation. The concept of this work has already been published in Hawe *et al.* [28]. This is the first time a nonparametric method based on a piecewise constant residue has been published. This will be applied to PET data and comparisons made between this and the compartmental model. Simulation studies to examine the efficacy of the various models is undertaken. This includes looking at rates of convergence for functionals and examines the test statistics used in linear regression in this nonlinear case.

Waxman *et al.* [79] describe guidelines on brain PET imaging which are briefly reviewed here, as the data used in this thesis is on brain imaging with FDG-PET of subjects with gliomas. In the brain, glucose provides approximately 95 percent of the energy needed for the brain to function [79]. Waxman *et al.* discuss

the relationship between glucose metabolism and neuronal activity highlighting that "FDG is suitable for imaging regional cerebral glucose consumption with PET since it accumulates in neuronal tissue depending on facilitated transport of glucose and hexokinase mediated phosphorylation. Therefore changes in neuronal activity induced by disease are reflected in an alteration of glucose metabolism". Simply put, if one has a hyper or hypoglycemic tumour, one can visualise higher or lower than usual uptake respectively. The relationship between brain energy consumption and neuronal activity is mentioned here, but there are other uses of FDG-PET in brain imaging. For these the reader is referred to [71, 79].

In Chapter 4, various analyses including the models mentioned in Chapter 2 and 3 along with another alternative, a mixture model, will be applied to a data set from a clinical study from the University of Washington, Seattle. This is a large dataset consisting of over forty patients. This data set consists of subjects with resected gliomas imaged with FDG-PET. Whilst analyses have previously been conducted on this data set by Spence *et al.* [71], this is a secondary analysis examining other techniques, such as nonparametric models, which have not previously been applied to these data. Comparisons between all three models will be presented. In each subject an individual analysis is conducted on four regions of the brain. An analysis is also conducted on every single voxel in the field of view. Analyses such as this are frequently seen in the literature as they provide an advancement of the knowledge of medical sciences.

It is also worth noting that the work in this thesis was automated using batch files in Windows XP and 7 to run analyses. This automation was particularly nice as it was able to analyse all datasets contained in a folder and does not stop until the analyses are completed. It has been suggested that this could be modified to become an R package. This is presently being considered.

1.5 Discussion

There have been great advances with PET since the 1950's. The work invested in getting the technology to where it is today has not stopped and in the coming years we can expect more advances in engineering and science to improve PET scanners. Examples of improvements include scanners that can now conduct CT

scans or MR at the same time as a PET scan as well as time of flight PET scanners and PET scanners with better resolution.

The teams of scientists working on advancing PET are diverse. Although this thesis is a statistical perspective on PET, there are engineers working to improve scanning technology and resolutions, physicists and chemists working on improving and making new radiotracers, physicians working with patients are formulating hypotheses to be tested with PET along with those who try to infer information from PET images. Even today, a large body of researchers are attempting to improve the reconstruction process.

After reconstruction there is still much work being done in PET research. In the case of static imaging, a lot of research is being conducted on cancer, such as sarcoma, with a view to improving treatment and patient outcomes. This work relies heavily on survival analysis [61]. In terms of dynamic imaging there is a volume of work underway, including developing nonparametric techniques to examine PET data such as that in [14, 28, 50, 60]. There are also many other open problems in PET being investigated such as kinetic mapping, segmentation and recovery of the AIF without arterial sampling [33, 58].

PET has been established and validated as a method for understanding cancer in patients, how a patient's heart is functioning, neurological disorders, as well as other areas that PET is only now branching into. It can be expected that in the future the areas PET will be used for will grow, leading to many new problems for researchers to work on in statistics and other sciences.

In PET the role of statistics cannot be understated. In reconstructing the PET image there is a wide variety of statistical problems which have to be solved. It must be remembered that a PET data set is a very large data set. Once the reconstruction is completed, a large number of techniques are applied depending on what is of interest in the scan. In some cases a longitudinal study is presented with the outcome of patients several years after diagnosis available. In this case one may try and establish if parameters obtained from the scan have prognostic value. Once this is established it may be used in a clinical practice to influence the course of treatment. This is only a small taste of the use of statistics in PET. To explain the position of statistics in medical imaging further, if one applies for a medical imaging research grant from the United States National Institute for

Health, one must include a section on statistics. If this is not included then the proposal will not be considered. This seems a responsible attitude for funding agencies to take as the volumes of data being produced are large and given the sensitive nature of any record of a medical procedure, it seems prudent to have statisticians involved from the start of the project.

Chapter 2

Compartmental Models and Inference

2.1 Outline

In this chapter, models which are usually applied to PET data will be introduced. These models are built on the work of Meier and Zierler [47] and have been previously proposed by Kety and Schmidt [39] and Huang *et al.* [34] and are discussed by Gunn *et al.* [26, 27]. While the AIF is known it remains to have a form to model tissue activity, which is generally done using compartmental models. These compartmental model equations have some assumptions, which may not always be justified and so in this chapter we consider the compartmental models from a stochastic point of view based on Markov chains.

In Section 2.2 we begin by discussing an equation called the basic equation of dynamic PET. This equation contains a term representing the activity in blood and another representing activity in tissue. A compartmental model technique in estimating this tissue term will be presented in Section 2.3. In examining compartmental models, we will review the assumptions inherent to the model and the two most widely used models in PET, the one compartmental Kety-Schmidt model and the two compartmental Sokolov-Huang model and discuss parameters of interest. A Markov chain approach to modelling the PET data is proposed in Section 2.4 and the relationship between the compartmental and Markov models will be examined and presented here. A review of statistical

inference will be presented in Section 2.5 which leads to an illustration of the compartmental model which will be presented in Section 2.6.

It is worth highlighting here that there are alternatives to compartmental models based on biochemistry of the human body or the vascular system [3]. In a sense the work others have done such as the Physiome project (physiome.org) has inspired the efforts in this Chapter.

2.2 The Basic Equation of Dynamic PET

In terms of modelling dynamic PET data, there are a number of assumptions required, perhaps the most important of these are the fact that the data is considered linear and time invariant which allow the techniques presented in this section to be used. Linear means that increasing the dose by an amount increases the resolution by a related amount and that this relationship is linear. Time invariant means that we assume had the dataset been generated at an earlier or a later time, it would not effect the results. Perhaps this can be put simply as if a subject is imaged now or several hours later should not effect the results of the scan.

In the case of the PET scan, $C_T(t)$ is the concentration of atoms in a tissue region i.e. the total tissue activity per unit volume (KBq/cm^3). This is expressed as a convolution involving the AIF ($C_P(t)$) which is activity per milliliter (mL) of blood and the tissue impulse response or residue function ($R(t)$), which is necessarily positive and monotonically decreasing [47],

$$C_T(t) = K \int_0^t R(t-s) C_p(s) ds \tag{2.1}$$

where K is a proportionality constant, interpretable as an overall flow, with units mL per unit time. Physically the residue function describes the fraction of indicator, which remains in the tissue in response to an idealised bolus (Dirac delta function) input concentration at time $t = 0$. Initially the residue must be unity ($R(0) = 1$) and from there it decreases (at least does not increase) as t increases. This has the same properties of a survival curve in statistics. The flow and extraction of tracer is thus estimated using the residue function.

If the fractional vascular volume within the tissue is v_B, is in units mL of blood per unit volume of tissue and measurement of the arterial time-course $C_P(t)$ is made at a site (radial artery or heart) which is remote from the tissue region, a delay term (Δ) is included as well as a term to represent tracer in arterial blood within the ROI, then Equation 2.1 becomes:

$$C_T(t) = v_B C_p(t - \Delta) + K \int_0^t R(t - s) C_p(s - \Delta) ds \tag{2.2}$$

Equation 2.2 is termed the basic equation of dynamic PET radiotracer imaging [28, 60, 47].

We now consider some background to this equation by looking at material derived in Meier and Zierler's seminal paper [47] on indicator dilution theory. In this work the authors provide a review of indicator dilution theory and prove the central volume theorem, which states that

$$\text{Blood Flow} = \frac{\text{Blood Volume}}{\text{Mean Transit Time}}.$$

Although this theorem had been used widely before Meier and Zierler's work was published, a proof had not been published. The authors believed "it may be of some value to present a direct proof of its validity under appropriate conditions".

Consider a vascular system, where a radiotracer has been injected into the subject being studied. This radiotracer is making its way through a region of tissue which includes a vascular system. A simple schematic of what a vascular system might look like is shown in Figure 2.1. We note that this is a simpler

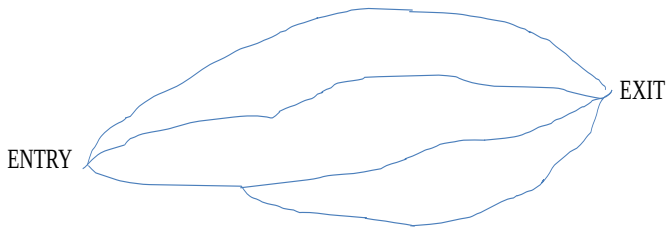


Figure 2.1: *Schematic of vascular system within a region of tissue.*

system than the human body as one can see there is only one entry and exit point in the system. We will assume the system has a constant vascular volume V which flows at a rate of F.

It is worth noting that all fluid particles need not take the same amount of time to get through the system and in fact it is unlikely that they will. As a consequence we will not have a constant travel time but instead will have a distribution of travel times. We will also assume that this distribution is invariant over time to simplify the modelling. Additionally, we require the flow of the radiotracer in the system to represent the total flow of the substance occurring in the body, i.e. the flow of the water tracer should be the same as how normal water circulates around the body. It is also necessary to assume that there is a sink in the system, that is to say that if a subject was injected with ^{15}O water that the total amount of tracer is eventually removed from the system (i.e. tracer ends up in the bladder and is disposed of) and it is also assumed that recirculation does not occur. Recirculation can be thought of simply as the radiotracer doing a second lap of the body while in the blood.

Suppose q units of radiotracer were injected at the start of the vascular system in Figure 2.1. The observed concentration at the exit of this system is a variable and is called $c(t)$. The amount of indicator leaving the system in a time interval of length δt is given by the concentration of fluid at the exit multiplied by the volume of this fluid leaving multiplied by the length of the interval ($c(t)F\delta t$). If all the radiotracer which enters the system must eventually leave the system, then

$$q = \int_0^\infty c(t)F\delta t = F\int_0^\infty c(t)\delta t.$$

However it may be the case in a PET scan that not all the radiotracer will have left the system before the scan concludes so we consider

$$h(t) = \frac{Fc(t)}{q}$$

as the fraction of radiotracer leaving the system per unit time. Naturally the integral of this quantity over an infinite time window will be unity but

$$1 - \int_0^z h(t)dt$$

for real z provides the amount of radiotracer left in the system at time z. Obviously q can be considered as the concentration of tracer entering a region and $c(t)$ the quantity flowing out of it.

2.3 Compartmental Models

As previously mentioned in Section 2.1 PET kinetic analysis is conducted to separate various signals measured by the scanner. In order to isolate the desired component of the signal, a mathematical model relating the dynamics of the tracer molecule in its various states is used. In compartmental modelling, each of these states is represented by a distinct compartment. Compartmental models are a well established technique in conducting kinetic analyses in PET [26, 27, 48, 62, 82], where each compartment is represented by the concentration of the tracer within that compartment as a function of time.

More accurately, a compartmental model is a mathematical way to describe a system where materials or energies are transmitted between different states. For this type of modelling to be feasible, there are a number of assumptions which must be satisfied. Firstly, each compartment must be homogeneous. Once a particle enters a compartment, it instantaneously follows the spatial and temporal distribution of all particles in that compartment. The exchange between compartments is assumed to relate to the volume of the quantity of interest in each compartment (Law of Mass Action). These assumptions are discussed in more detail in the next section. There are more assumptions that are required in the case of imaging that are not always true in other applications of compartmental modelling, which are also presented in the next section.

There are a number of positive and negative aspects associated with using compartmental models. There is a large body of work in the literature on the background to compartmental models [9, 24, 26, 27]. While [24] discusses compartmental models generally, [26, 27] discuss compartmental modelling in the context of PET and Bates and Watts discuss compartmental models in the context of nonlinear regression [9]. Having a wide literature makes it much easier to implement compartmental models and where problems arise, there may be solutions in the literature.

Compartmental models are relatively straightforward to implement. Software such as JSim (www.physiome.org) are designed to make applying compartmental models to data quick and uncomplicated. They can also be implemented in R as is the case in this work [65]. Although writing software to fit these models

can be straightforward, frequently the optimisation involved in fitting the models converges to the wrong minimum or does not converge at all and finding the right solution can be tedious and laborious.

In the rest of this section, a brief discussion of the the assumptions underlying compartmental models and justification for the use of these models is presented.

2.3.1 Assumptions

Homogeneity

In the case of compartmental models, homogeneity means that the distribution of tracer in each compartmental is uniform. This is not an unreasonable assumption to make if a small region of a patient was under study such as a small white matter region of a human brain or a small region of muscle in a limb. One has to wonder if it is appropriate, however, to examine an entire human brain for example and treat this as one compartment. Questionable analyses like these can be seen in the literature and although, to the eye, fits of the model to the data may look appropriate, this may not be the correct model to use and may result in parameters estimated from this model being incorrect. An obvious example of this is the case of a region containing diseased tissue. This will include both healthy and damaged tissue and will not be homogeneous. In fact, if all that was contained in the region was purely the disease, it may still not be homogeneous as aggressive diseases tend to exhibit heterogeneous phenomena. To emphasise this point further we consider a region containing a brain tumour and a neighbourhood of that brain tumour. This could include cancer and healthy tissue, arterial blood, venous blood, necrotic tissue, white matter, grey matter, spinal fluid and possibly bone. This is far from homogeneous.

Instant Mixing Within Compartments

Compartments are assumed to be homogeneous for all time, so when a particle enters a compartment it is assumed that the distribution within the compartment will remain the same. This is not a massive stretch in that if there are a thousand particles in a region and one more arrives it will not significantly effect the distribution of the particles in the compartment. However, in the case of a PET

study, particles may be attracted to one particular region of tissue over another and, if cancer cells are present, this is even more likely. Again as was the case for homogeneity, we highlight that cancer does not behave in a well stirred fashion.

Law of Mass Action

The law of mass action states that the rate at which a reaction occurs is proportional to the product of the concentrations of the participating molecules [74]. This law assumes continuity of the quantity of interest and this means that the compartments can be considered to be in continuum, i.e the number of atoms of tracer is a continuous and differentiable quantity. This allows us to translate the compartmental models to a mathematical formulae later in this chapter. This is at odds with the fact that the tracer is absorbed by the tissue in small amounts.

Assumptions about the tracer

We assume that the tracer being added to the system does not alter the behaviour of that system. This assumption is really stating that the tracer does not change what we seek to measure. In the case of FDG we assume that the FDG being present in the system does not effect the subject's glucose utilisation. This is important as the goal of the scan is to understand a process in the body. However if the tracer changes the process of interest in the body, then we are measuring something different to what we intend.

It is assumed that the tracer behaves in the same way as the substance it seeks to emulate i.e. it behaves similarly to glucose, etc. Again this is obvious as if the radiotracer does not behave in the same way as the substance it emulates, then we are not measuring what we want. In the case of FDG, there are differences between how glucose and FDG behave, however these have been discussed in Chapter 1 and the use of FDG in a clinical setting is well established. We also assume that the labelling procedure does not effect the way the body uses the tracer. Although this is not the case for FDG it has a long half-life which is advantageous in practice as it lasts longer. ^{11}C Glucose would be a more accurate glucose analogue but with a half life of approximately 20 minutes, it is of little use in clinical practice. It has also been highlighted in Chapter 1 of this thesis

that FDG has been validated as a good prognostic indicator for many cancers.

2.3.2 Justification for the Use of Compartmental Models

While the assumptions provided above are somewhat questionable, one should also realise that there is significant justification for the use of compartmental models in practice. Studies carried out *in vitro* have shown that in a test tube setting the models are accurate and this was recently mentioned in the context of nonparametric modelling by O'Sullivan *et al.* [60]. The extension from this to small heterogeneous regions of the body seems reasonable and is the main justification for this in practice.

2.3.3 The One Tissue Compartment (Kety Schmidt) Model

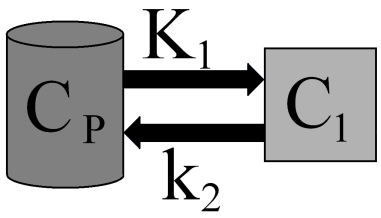


Figure 2.2: *One tissue compartmental model [27].*

The present formulation of the transport of drug through a subject has been described by Kety and Schmidt [39] and Phelps *et al.* [63]. As previously mentioned, this can be written down simply by assuming the rate of change of concentration of drug in tissue $C_1(t)$ is equal to the rate constant associated with going from the plasma into tissue multiplied by the amount of drug in the blood and subtracting from it the concentration in the tissue multiplied by the rate constant associated with moving from tissue to plasma. For the Kety-Schmidt model, shown in Figure 2.2, this gives Equation 2.3.

$$\frac{dC_1}{dt} = K_1 C_P(t) - k_2 C_1(t) \tag{2.3}$$

Here our method of solution is to assume that the concentration in the plasma is known and simply solve this equation for $C_1(t)$. This is sensible as in the case of

a PET study the arterial input is usually measured by arterial sampling. Equation 2.3 is a first order linear ordinary differential equation [10] and can be solved by the technique of an integrating factor [10]. Detailed workings are present in Appendix A at the end of this chapter, where the solution is found to be:

$$C_1(t) = K_1 \int_0^t e^{k_2(t-s)} C_P(s) ds. \tag{2.4}$$

One can see that $C_1(t)$ can be related to the fundamental equation if the residue function $R(t) = e^{-k_2 t}$ giving

$$C_T(t) = v_B C_P(t-s) + K_1 \int_0^t e^{k_2(t-s)} C_P(s) ds.$$

It is worth highlighting that the residue is the key parameter, we seek to estimate in this work and that physiologically it represents the proportion of tracer remaining in tissue, in response to an idealised bolus and physiologically it can be interpreted as the fraction of indicator remaining in tissue.

2.3.4 The Two Tissue Compartment (Sokolov Huang) Model

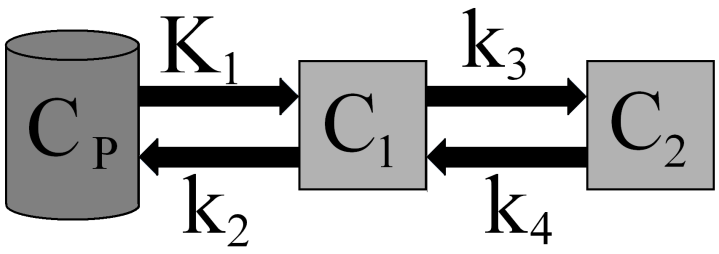


Figure 2.3: *Two tissue compartmental model [27].*

In the case where the tracer undergoes a reaction in tissue, a second tissue compartment is added [63, 70] as shown in Figure 2.3. For example the FDG tracer is transformed into FDG-6-P and back during the imaging procedure. Similar to the Kety Schmidt model, one can write down first order linear continuous deterministic differential equations for the system.

$$\frac{dC_1}{dt} = K_1 C_P(t) - (k_2 + k_3) C_1(t) + k_4 C_2(t) \tag{2.5}$$

$$\frac{dC_2}{dt} = k_3 C_1(t) - K_4 C_2(t). \tag{2.6}$$

Again, we assume that the concentration in the plasma is known and simply solve this equation for $C_1(t)$ and $C_2(t)$. Equations 2.5 and 2.6 are a coupled system of order linear ordinary differential equations and can be solved using Laplace's transformation [10, 34, 82] and linear algebra [6]. The workings in finding a solution to these equations are given in Appendix B at the end of this chapter. One can see that $C_1(t)$ and $C_2(t)$ can be related to the fundamental equation if $R(t)$ is a mixture of exponentials giving

$$C_T(t) = v_B C_P(t-s) + K \int_0^t (\phi_1 e^{\theta_1(t-s)} + \phi_2 e^{\theta_2(t-s)}) C_P(s) ds,$$

where ϕ_i and θ_i depend on the rate constants and the residue function $R(t) = \phi_1 e^{\theta_1(t-s)} + \phi_2 e^{\theta_2(t-s)}$. Full details relating these rate constants and ϕ_i and θ_i along with the method of solution for the two compartmental model using Laplace's transformation can also be found in Appendix B.

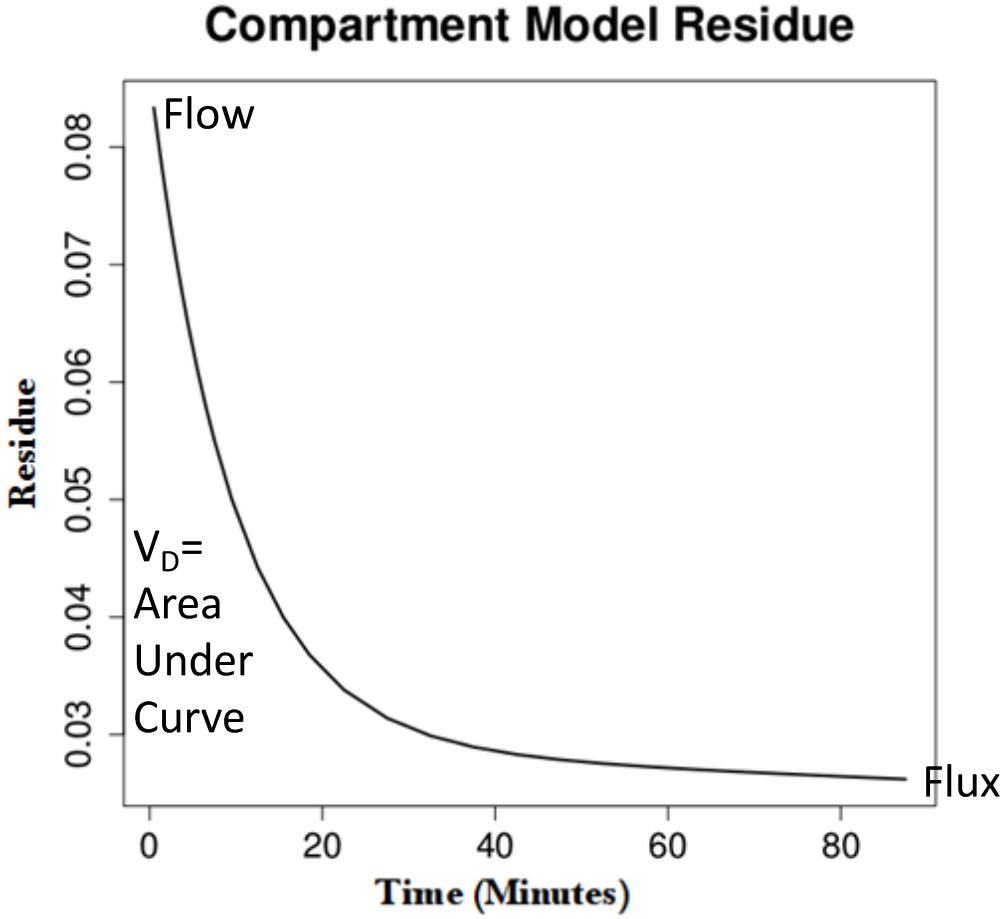


Figure 2.4: *Residue function with key functionals highlighed.*

Parameters of interest which are generated in this work stem from the residue function as is shown in FigureRESILL. Having an estimated residue allows the subsequent parameters to be estimated. We now briefly discuss five parameters. Firstly we discuss extraction, which is the proportion of radiotracer which remains in tissue at the end of the scan. In theory it is hoped that a steady state has been reached in the body and at this point the extraction is constant. In reality this may not be the case as the scan is usually only 90 minutes long. However, it is the best estimate for extraction. It is estimated as the last time point in the residue function.

Flow has already been discussed in the context of defining the basic equation of dynamic PET. The goal in carrying out kinetic analysis is to separate flow from retention. In different scenarios, high and low values of flow can have positive or negative outcomes for patients. The flow parameter can be measured by fitting the compartmental model to data and obtaining the value at the start point of the residue function.

Another parameter frequently measured in PET studies is the metabolite flux, which is the extraction multiplied by flow. At late time points on an FDG scan flux dominates the image.

The volume of distribution (v_D) typically refers to the volume of blood (or plasma) that would be required to account for the amount of drug in the entire body - assuming the drug is distributed uniformly [36]. More simply this can be described as the number of liters of blood required to have the same amount of PET tracer as that contained in the tissue. This can be estimated as the integral of the residue function.

The fractional blood volume (v_B) is the volume of space in an ROI occupied by arterial blood. This is a proportion of the entire ROI so obviously as it represents a proportion, it can only take on values between zero and one. This parameter is of clinical significance when a patient has suffered an injury or illness such as brain injury or stroke, or in cancer care where physicians seek to answer the question of whether a malignant mass has developed its own vascular network, or stroke where blood flow in a key prognostic indicator.

As our goal is to estimate these parameters, it is important to have the right model to fit to the data. Key to this is that assumptions are not violated and

that the model is appropriate for the data. As has already been established in some cases the compartmental models use is questionable at best and so we now consider what happens at the molecular level.

2.4 A Markov Formulation of Compartmental Models

We have already stated that compartmental models are based on a number of assumptions, which some analysts have called into question [14, 28, 50, 60]. These assumptions include that the concentration of tracer in tissue is a differentiable quantity, that there is instant mixing within compartments, that the exchange rate of materials among compartments is related to the densities of these compartments and that the compartments are homogeneous.

An exploration of compartmental models was undertaken with a goal of gaining an insight into how these models can be generated from a stochastic modelling perspective. This flow from the discrete stochastic approach to the continuum hypothesis is an interesting area to study in itself. Although, the work here may not be of direct interest to others for practical use, the understanding of how the compartmental models work that is gained by an analysis such as this is invaluable.

Directly, this is not likely to make an impact on how PET data are modelled as this analysis is a level far far below the resolution, which a PET scanner is capable of, however taking an approach such as this allows us exploration of the models in a new way and provides avenues of future research such as the nonparametric method, which will be presented in Chapter 3.

Compartmental models are widely used in fields such as epidemiology where the variable in question is not a quantity that can be measured in a quantitative state, or in a pharmaceutical plant where containers are undergoing constant agitation. Instead they are used to model different states that the variable can take on. For example, a subject in an epidemiological study can be susceptible to disease (S), infected (I) or recovered (R) from disease and this is how the SIR model in epidemiology has been established. In the case of a PET radiotracer it

can be thought of as being in the circulatory system or in tissue and a similar formulation applies. However, it is not as clear cut as to what state an atom of radiotracer is in, as is the case in the SIR population model as the infected populations are much more homogeneous compared to the human body, and instant mixing may not occur *in vivo*. You are either susceptible or not susceptible, infected or not infected, recovered or not recovered. In the case of a region of tissue, you can be in arterial blood or in venous blood, be absorbed by one of many types of tissue, etc. This is much more complicated.

This following section of this chapter presents an alternative approach to modelling these states using a Markov chain approach. Results related to the Markov model are presented and a limiting case where both formulations are equivalent is presented.

2.4.1 Markov Chains

A Markov chain is a discrete stochastic model, which can take on a number of states. The future state of the process depends only on the present and not on the past. We can express this as

$$P(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x | X_n = x_n).$$

This is why Markov chain models are described as memoryless or time homogeneous. The transitions from state to state are modeled by a matrix of transition probabilities (P) [13]. The $(i,j)^{th}$ entry in P is the probability that an atom presently in state i will be in state j at the next time step. This can be written as a discrete equation

$$\mathbf{x}_{n+1} = P\ \mathbf{x}_n. \tag{2.7}$$

A simple example of a Markov chain is to consider a random walk, where you start at the origin and at every time point you move either one unit left or one unit right with equal probability. In this case where you have walked previously does not effect where you will end up next, all that matters is your present position. This is illustrated in Figure 2.5, where the random process is moving from blood to tissue.

A more complicated example is the exponential distribution. In this case we can write down a probability density function for the distribution as

$$f_x(x|\lambda) = \lambda e^{-\lambda x} \; x > 0$$

$$f_x(x|\lambda) = 0 \;\; \text{otherwise.}$$

The mean and variance can be found as $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$ respectively. This distribution can be found to exhibit the memoryless property as follows: The cumulative density function for the exponential distribution is given by $e^{\lambda t}$. This is the probability that the random variable is greater than some known value called say t. If we try and find the probability of the random variable being greater than $t+\varepsilon$ given that the variable is bigger than ε, we find the same answer, again owing to the memoryless property of the exponential distribution. This example has been highlighted as the memoryless property of the exponential distribution is well known and also because the retention times for particles within a compartmental model are exponential and this distribution is memoryless. The reason the exponential distribution is mentioned here is that the retention time of a particle is exponential. Relating this to the residue function is the basis of further work in this chapter.

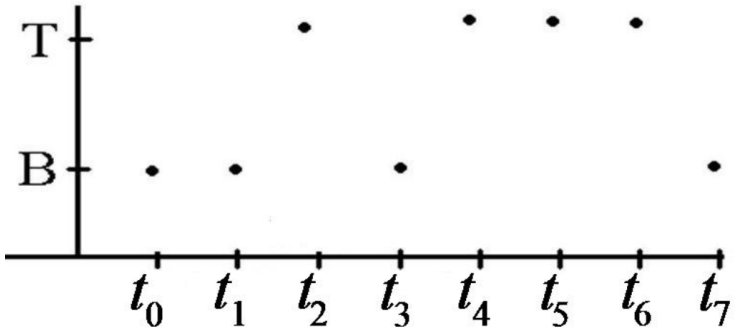


Figure 2.5: *Random process of an atom being in the blood or tissue at various times.*

2.4.2 The Relationship Between Discrete Markov Chain and Continuous Differential Equation Models

Simple Markov Models for Chemical Equations

This section involves some theoretical analysis of the compartmental models from a probabilistic point of view. Work like this has previously been used in chemical equations, but not in the context of PET compartmental models. Anderson *et al.* [5] considers a chemical reaction where A and B mix to form C, the basis of the model is that the probability of a reaction occurring in a time interval Δt should be proportional to the product of the number of atoms of A and B and the length of the time interval Δt.

Let us first assume that the number of atoms in A, B and C are known and called $X_A(t)$, $X_B(t)$ and $X_C(t)$ respectively that the rate constant associated with the reaction of A and B to form C is κ. We can say that

$$P(\text{Reaction occurs in interval } \Delta t | F_t) \approx \kappa X_A(t) X_B(t) \Delta t,$$

where F_t represents the previous known information about the system available at the start of Δt and κ is a rate constant for the reaction. One can see the relationship between this formulation and the mass action formulation used to generate the differential equations for the compartmental models.

We now seek to turn this into a mathematical model. We can say that every time a reaction occurs (called $C(t)$) then the number of atoms of A and B decreases by one and the number of atoms of C increases by one. Hence

$$\mathbf{X}(t) = \mathbf{X}(0) + C(t)(-1, -1, 1)^T,$$

where $\mathbf{X} = (\mathbf{X_A}, \mathbf{X_B}, \mathbf{X_C})^{\mathbf{T}}$, the number of atoms of A, B and C present in the system and we assume that the time intervals Δt are sufficiently small that only one reaction occurs in any interval and that initially $C(t) = 0$. A form for $C(t)$ can be derived from the assumption that the number of reactions are Poisson. In this case

$$C(t) = Y(\int_0^t \kappa X_A(s) X_B(s) ds),$$

where Y is a Poisson process and κ is the rate constant associated with the

reaction. The number of atoms of A and B which are available to react are denoted by X_A and X_B respectively.

Hence we can write

$$\mathbf{X}(t) = \mathbf{X}(0) + Y(\int_0^t \kappa X_A(s)X_B(s)ds)(-1,-1,1)^T.$$

If Y and $\mathbf{X}(0)$ are known this equation uniquely determines $\mathbf{X}(t)$. This can be made clear by considering the first interval of length Δt. If we let $\psi = (-1,-1,1)^T$, then $X(t) = X(0) + \psi$ will determine the value of $X(t)$ at the end of the first interval. In the second interval we will have $X(t) = X(0) + 2\psi$ and so on.

We now seek to justify why the probability of a reaction occurring in a time interval Δt is $\kappa X_A(s)X_B(s)\Delta t$. We first note that the probability of a reaction in an interval Δt is equivalent to $R(t+\Delta t) > R(t)$.

$$P(C(t+\Delta t) > C(t)|F_t) = 1 - P(C(t+\Delta t) = C(t)|F_t)$$
$$= 1 - P(Y(\int_0^t \lambda(X(s))ds + \lambda(X(t))\Delta t) = Y(\int_0^t \lambda(X(s))ds)|F_t))$$
$$= 1 - e^{-\lambda(X(t))\Delta t} \approx \lambda(X(t))\Delta t.$$

While this simple illustration from Anderson's paper [5] is useful, it is more useful to look at a general case as we are interested in compartmental modelling.

General Markov Models for Chemical Equations

Anderson [5] generalises the above chemical model to the case where there are M reactions with N species $S_1 ... S_N$. In this, he writes the chemical reactions as

$$\sum_{i=1}^{N} \nu_{ik} S_i \to \sum_{i=1}^{N} \nu'_{ik} S_i,$$

where ν_{ik} and ν'_{ik} are nonnegative integers. As previously, $X(t)$ represents the number of atoms of each species in the model and ν_k is the vector whose i^{th} entry is ν_{ik}, being the number of molecules of the i^{th} species consumed in the k^{th} reaction. Similarly, ν'_k and ν'_{ik} represent molecules created in the reaction.

Using the same notation as in the case of a single chemical reaction, we can write

$$P(R_k(t+\Delta t) > R_k(t)|F_t) = \lambda_k(X(t))\Delta t + o(\Delta t) \text{ as } \Delta t \to 0.$$

Similar to the simple case the reactions will follow a Poisson process and so $R_k(t) = Y_k(\int_0^t \lambda_k(X(s))ds)$ and noting $\psi_k = (\nu'_k - \nu_k)$, we write

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_k Y_k(\int_0^t \lambda_k(X(s))ds)\psi_k$$

We have already discussed the fact that the rates of the reactions are Poisson but now it seems appropriate to discuss them generally as we have written down the form of a compartmental model equation. Anderson [5] states that the law of mass action says that the rate at which a reaction occurs should be proportional to the number of distinct subsets of the molecules present that can form the inputs for the reaction. That is

$$\lambda_k = \kappa_k \prod_{i=1} \binom{x_i}{\nu_{ik}} 1_{x_i \geq v_{ik}}.$$

This result does not directly effect compartmental models. It is stated here to aid the understanding of the stochastic approach. What this result really means is that if you have a binary reaction such as

$$2S_1 \xrightarrow{\kappa_1} S_2,$$

then in this case the rate of the reaction occurring is:

$$\lambda_k = \kappa_1 x_1(x_1 - 1),$$

which differs from the deterministic continuum case of $\kappa_1 x_1^2$.

The Stochastic Equation for Compartmental Models

Using the work in the previous section on a general stochastic equation for chemical reactions we consider the compartmental model analogous to chemical equations. If we consider the simple one tissue compartmental model in Figure 2.2, we arrive at the following stochastic equation:

$$\mathbf{X}(t) = \mathbf{X}(0) + Y_1(\int_0^t K_1 X_P(s)ds)(-1,1)^T + Y_2(\int_0^t k_2 X_1(s)ds)(1,-1)^T.$$

For the two compartmental model the equation is

$$\mathbf{X}(t) = \mathbf{X}(0) + Y_1(\int_0^t K_1 X_P(s)ds)(-1,1,0)^T + Y_2(\int_0^t k_2 X_1(s)ds)(1,-1,0)^T$$
$$+Y_3(\int_0^t k_3 X_1(s)ds)(0,-1,1)^T + Y_4(\int_0^t k_4 X_2(s)ds)(0,1,-1)^T.$$

Relating Stochastic and Deterministic Equations

Again a main reference here is Anderson [5], particularly for the theorems and lemmas. As compartmental models are generally first order linear models, work on first order chemical reactions is generalisable to compartmental models. Firstly we define what we mean by first order. A first order reaction is one where the rate parameters are linear and the reactions are unary (only involve one substance reacting or changing state). This will be the case for PET compartmental models.

Theorem 2.1 If Y is a unit Poisson process, then for each $u_0 > 0$,

$$lim_{n\to\infty} sup_{u\leq u_0} |\frac{Y(nu)}{n} u| = 0 a.s.$$

Proof. For fixed u, by the independent increments assumption, the result is just the ordinary law of large numbers. The uniformity follows by monotonicity. The classical central limit theorem implies

$$lim_{n\to\infty} P|\frac{\frac{Y(nu)}{n} nu}{\sqrt{n}} \leq x| = \int_0^\infty \frac{1}{\sqrt{2pi}} e^{\frac{-y^2}{2}} dy = PW(u)x$$

where W is a standard Brownian motion. This result suggests that for large n

$$\frac{Y(nu)nu}{\sqrt{n} \approx W(u)}$$

and

$$\frac{Y(nu)}{n} \approx u + \frac{1}{\sqrt{n}} W(u)$$

where the approximation is uniform on bounded time intervals. One way to make this approximation precise is through the following lemma.

Lemma 2.1 A unit Poisson process Y and a standard Brownian motion W can be con- structured so that

$$\Gamma = sup_t \geq 0 \frac{|Y(t)tW(t)|}{log(2t)} < \infty a.s.$$

and there exists $c > 0$ such that $E[e^{c\Gamma}] < \infty$. For a proof of this the reader is referred to Anderson and Kurtz [5]. Here we simply note that:

$$|\frac{Y(nt)nt}{\sqrt{n}} - \frac{1}{\sqrt{n}} W(nt)| \leq \frac{log(nt2)\Gamma}{\sqrt{n}}$$

and $\frac{1}{\sqrt{n}} W(nt)$ is Brownian motion.

Lemma 2.2 [5] Assume that $C = (C_1, ..., C_m)$ is a system of counting processes with no common jumps and λ_l is the F_t-intensity for C_l . Then there exist independent unit Poisson processes $Y_1, ..., Y_m$ (perhaps on an enlarged sample space) such that

$$C_l(t) = Y_l(\int_0^t \lambda_l(s)ds)$$

This lemma suggests the following alternative approach to relating the intensity of a counting process to the corresponding counting process. Again, given nonnegative, non-anticipating functions $_l$, the intuitive problem is to find counting processes C_l such that

$$PC_l(t + \Delta t) > C_l(t)|F_{tl}(t, C)\Delta t,$$

which we now translate into the following martingale problem. In the following definition $J_m[0, \infty)$ denotes the set of mdimensional Cadlag counting paths.

Definition 2.1 Let $_l$, $l = 1, ..., m$, be nonnegative, non-anticipating functions defined on $J_m[0,)$. Then a family of counting processes $C = (R1, ..., Cm)$ is a solution of the martingale problem for $(1, ..., m)$ if the R_l have no simultaneous jumps and there exists a filtration Ft such that C is adapted to Ft and for each l and k,

$$C_l(t \geq \tau_k) - \int_0^{t>\tau_k} \gamma_l(s, C)ds$$

Lemma 2.2 implies that every solution of the martingale problem can be written as a solution of the stochastic equation. Consequently, the stochastic equation and the martingale problem are equivalent ways of specifying the system of counting processes that corresponds to the l .

This gives rise to the useful property that

$$E[\lambda_k(X(s))] = \lambda_k E[X(s)],$$

for all k. There we can write

$$E(\mathbf{X}(\mathbf{t})) = E(\mathbf{X}(0)) + \sum_k E(Y_k(\int_0^t \lambda_k(X_k(s))ds))\psi_k$$

$$= E(X(0)) + \sum_k \psi_k \int_0^t \lambda_k(E[X_k(s)])ds.$$

This gives an easy way to find the mean of $X(t)$. By turning the above integral equation into a differential equation, a solution can be found.

We can recall that the deterministic ordinary differential equation associated with

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_k Y_k(\int_0^t \lambda_k(X_k(s))ds)\psi_k$$

is

$$\frac{d\mathbf{x}(t)}{dt} = \sum_k \hat{\lambda}(x_k(t))\psi_k,$$

where $\hat{\lambda}(x) = \hat{\kappa}_k$ as the reactions in this work are unary.

The Classical Scaling

We begin by defining

$$X^V(t) = \frac{X(t)}{V},$$

which defines concentration per unit volume. This leads to a slightly different stochastic equation

$$X^V(t) = X^V(0) + \sum_k \frac{1}{V} Y_k(\int_0^t \lambda_k(X^V(s)ds))\psi_k$$

$$\approx X^V(0) + \sum_k \frac{1}{V} Y_k(V \int_0^t \tilde{\lambda}_k(X(s)ds))\psi_k$$

$$= X^V(0) + \sum_k \frac{1}{V} \tilde{Y}_k(V \int_0^t \tilde{\lambda}_k(X(s)ds))\psi_k + \int_0^t F(X(s))ds,$$

where $\tilde{Y_k}(u) = Y_k(u) - u$ and owing to the process being centered we must add in an extra term, $F(z) = \sum_k \kappa_k \prod_i z_i^{\nu_i j}\psi_k$.

Using the law of large numbers,

$$\frac{\tilde{Y}(Vu)}{V} \approx u,$$

which implies

$$X^V(t) = X^V(0) + \sum_k \int_0^t \kappa_k \prod_i (X_i^V(s)^{\nu_{ij}}\psi_k ds)),$$

$$= X^V(0) + \int_0^t F(X^V(s))ds,$$

which in the limit as $n \to \infty$ gives the classical deterministic mass action result:

$$\frac{dx}{dt} = \sum_k \prod_i X_i(t)^{\nu_{ik}} \psi_k.$$

Owing to the fact that

$$\frac{1}{\sqrt{V}} \tilde{Y}_k(Vu) = \frac{Y_k(Vu) - Nu}{\sqrt{V}}$$

is approximately Brownian motion.

Letting

$$V^N(t) = \sqrt{V}(X^V(t) - X(t))$$

$$= V^N(0) + \sum_k \frac{1}{V} Y_k(V \int_0^t \lambda_k(X^V(s)ds))\psi_k + \int_0^t \sqrt{N} F(X^V(s) - F(X(s)))ds$$

$$= V^N(0) + \sum_k W_k \int_0^t \lambda_k(X(s)ds)\psi_k + \int_0^t \nabla F(X(s)) - V^N(s)ds,$$

as $n \to \infty$ gives $V^N \to V$ where

$$V(t) = V(0) + \sum_k W_k \int_0^t \lambda_k(X(s)ds))\psi_k + \int_0^t \nabla F(X(s)) - V(s)ds,$$

which gives the approximation

$$X^N(t) = X(t) + \frac{1}{\sqrt{V}} V(t).$$

2.5 Statistical Inference

In discussing statistical inference it is natural to briefly review linear modelling before coming to nonlinear regression for compartmental models. This is only a brief outline and for a detailed treatment of these areas the reader is referred to Bates and Watts [9] and Weisberg [80] for a treatment of nonlinear and linear regression respectively.

2.5.1 Linear Regression

Simple Linear Regression

If we consider the simple case again where there is one predictor x and one response y, which are related and this relationship is a straight line, we can write

$$E(Y|X = x) = \beta_0 + \beta_1 x.$$

In reality measured data contains variability and so the level of variability in the data needs to be quantified. If we assume the errors are independently and identically distributed (iid) coming from a zero mean normal distribution with constant variance(σ^2), then we can write

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

where $e_i \overset{iid}{\sim} N(0, \sigma^2)$ and y_i and x_i are the i^{th} values of the response and predictor respectively.

After estimating the parameters of interest and fitting the estimated line to data the difference between each observed and expected value can be computed. These are called the residuals, i.e. $\hat{e}_i = y_i - \hat{y}_i$.

We now return to estimating the parameters of interest. If we consider the data set and the parameters from a likelihood perspective, i.e. the likelihood of the data given the parameters, the probability of the data set is a product of normal variables and is proportional to

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e(-\frac{y_i - \hat{y}_i}{2\sigma})^2.$$

Finding the maximum likelihood estimator reduces to minimising the residual sums of squares.

Having estimates for the parameters is only part of the story it remains to generate an estimate for the variance. An unbiased estimator is attained by dividing the RSS by its degrees of freedom.

Multiple Linear Regression

We now generalise the discussion on simple linear regression to the case where there is more than p predictors. In this case it is more convenient to use matrix

notation. The model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \text{where } e \overset{iid}{\sim} N(0, \Sigma),$$

The estimates are found to be $\hat{\boldsymbol{\beta}} = (\mathbf{X^T X})^{-1}\mathbf{X^T Y}$. As per simple linear regression, the estimate for the variance is the RSS divided by its degrees of freedom. From this estimates for the variance of parameter estimates can be generated.

An alternative formulation of Least Squares

An alternative formulation of least squares can be done using a QR factorisation. We can write $\mathbf{R}$ as $\mathbf{R} = \begin{pmatrix} \mathbf{R_1} \\ \mathbf{0} \end{pmatrix}$, where $\mathbf{R_1}$ has dimensions $P \times P$ and upper triangular. $\mathbf{Q}$ can be written as

$$\mathbf{Q} = [\mathbf{Q_1}|\mathbf{Q_2}],$$

where $\mathbf{Q_1}$ is the first P columns of $\mathbf{Q}$ and $\mathbf{Q_2}$ is the remaining columns of $\mathbf{Q}$. Hence

$$\mathbf{X} = \mathbf{QR} = \mathbf{Q_1 R_1}.$$

Geometrically the columns of $\mathbf{Q}$ form an orthogonal basis for the response. Also the first P columns span the expectation plane. For example the response can be transformed

$$\mathbf{w} = \mathbf{Q^T y},$$

which can be broken into two components $\mathbf{w_1} = \mathbf{Q_1^T y}$ and $\mathbf{w_2} = \mathbf{Q_2^T y}$. The projection of $\mathbf{w}$ onto the expectation plane is then $\begin{pmatrix} \mathbf{w_1} \\ \mathbf{0} \end{pmatrix}$, in the $\mathbf{Q}$ coordinate space and $\mathbf{X}\boldsymbol{\beta} = \mathbf{Q}\begin{pmatrix} \mathbf{w_1} \\ \mathbf{0} \end{pmatrix} = \mathbf{Q_1 w_1}$, in the original coordinate space.

We can also look at hypotheses testing and confidence intervals for parameters by noting:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} = QR\boldsymbol{\beta} + e.$$

An expression for the errors is given by

$$\mathbf{u} = \mathbf{w} - \mathbf{R}\boldsymbol{\beta},$$

where $\mathbf{u} = \mathbf{Q^T e}$, as $\mathbf{eI_n} \sim N(0, \sigma^2)$, and $\mathbf{Q}$ are orthogonal, then $\mathbf{u} \sim N(0, \sigma^2)$. Also as the errors in $\mathbf{e}$ are iid, so $\mathbf{u}$ will also be iid. Again $\mathbf{u}$ can be broken into two components, one with the first P elements called $\mathbf{u_1}$ and the remaining $n-p$ elements are called $\mathbf{u_2}$ As it is assumed that these errors are normal, their sums of squares will be χ^2, scaled appropriately for variance

$$||\mathbf{u_1}||^2 \approx \sigma^2 \chi^2_p \text{ and } ||\mathbf{u_2}||^2 \approx \sigma^2 \chi^2_{n-p}.$$

Using independence, we can assume that the ratio of mean squares $\mathbf{u_1}$ and $\mathbf{u_2}$ follow an F-distribution with p and $n-p$ degrees of freedom respectively. This can be used to determine if a particular value for $\boldsymbol{\beta}$ is appropriate. This will be discussed in detail shortly.

Weights

Up to now in this chapter, we have considered the case where the variance was constant in the data. However this is not always the case and will not be the case for PET imaging. Instead of assuming the variances are constant we will assume that they can be scaled by a factor at each measure to make the variances equal. This type of scaling is called weighting. Simply we write

$$w_i var(e_i) = \sigma^2,$$

where $w_1, w_2, ..., w_n$ are known positive numbers. In matrix terms this can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad Var(\mathbf{e}) = \sigma^2 \mathbf{W^{-1}},$$

where $\mathbf{W}$ is a diagonal matrix with the w_i in the ith row.

We will continue to use $\boldsymbol{\beta}$ to represent the parameters we seek to estimate, however these will now represent estimates generated using weighted least squares instead of the more straightforward approach. The least squares function we now seek to minimise is now given by:

$$WRSS(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$= \sum w_i (y_i - \mathbf{x_i^T}\boldsymbol{\beta})^2,$$

where x_i^T is the transpose of the ith row of the $\mathbf{X}$ matrix. Using a similar argument to the simple constant variance case, the weighted least squares estimates can be written as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^T W X})^{-1}\mathbf{X^T W Y}.$$

Comparing Regression Models

We will now consider the case where a dataset of interest consists of one predictor (y) and one response (x) which is estimated using least squares. We now wish to find the distribution of the RSS in closed form. We begin by noting that if $\mathbf{A}$ is any non random symmetric idempotent matrix and $\mathbf{z}$ is a multivariate standard normal random vector, then $\mathbf{z^T A z}$ has a chi-square distribution with rank of $\mathbf{A}$ degrees of freedom. This can be shown to be true by finding the singular value decomposition of $\mathbf{A = ODO^T}$, where $\mathbf{O}$ and $\mathbf{D}$ are the matrices generated from the matrix decomposition. If $\mathbf{z^T A z = w^T D w}$, where $\mathbf{w = zO}$ then the first p eigenvalues of $\mathbf{A}$ are equal to one and the rest equal to zero, then $\mathbf{w^T D w} = \Sigma_{i=1}^{p} w_i^2$ is χ^2 with p degrees of freedom.

Bringing this back to the case of residual sums of squares,

$$\mathbf{\hat{e}^T \hat{e} = e^T (I - H) e},$$

can be rewritten as

$$\frac{\mathbf{RSS}}{\boldsymbol{\sigma^2}} = \mathbf{z^T (I - H) z},$$

where $\mathbf{z} = \frac{\mathbf{e}}{\sigma}$. As $(\mathbf{I - H})$ has rank $n - p$, $\frac{RSS}{\sigma^2}$ follows a χ^2 distribution with $n - p$ degrees of freedom as the rank of an idempotent matrix is equal to the trace of that matrix [6].

Consider two regression models, model one: $\mathbf{Y_1 = X_1 \boldsymbol{\beta}_1 + e_1}$ and model two: $\mathbf{Y_2 = X_2 \boldsymbol{\beta}_2 + e_2}$, where model one is a simpler model and is nested in model two. These models will both have a hat matrix associated with them $\mathbf{H_1}$ for model one and $\mathbf{H_2}$ for model two. It is not immediately obvious that if the models are nested then $\mathbf{H_1 H_2 = H_2 H_1 = H_1}$. This is perhaps easier if one thinks of this in terms of projections and the spaces defined by the bases associated with each model.

This can be seen by realising that every vector of the form $\mathbf{H_1 y_1}$ can be written in the form $\mathbf{H_2 y_2}$ as model two is contained in model one. Hence

$$\mathbf{H_2 H_1 y_1} = \mathbf{H_2^2 y_2} = \mathbf{H_2 y_2} = \mathbf{H_1 y_1}$$

and so $\mathbf{H_2 H_1} = \mathbf{H_1}$. Similarly, it can be shown that $\mathbf{H_1 H_2} = \mathbf{H_1}$. We now wish to generate a form for a test statistic to see which of two models should be chosen as the better model. From the previous section, it is known that the denominator $(\frac{||\mathbf{e_2}||^2}{\sigma^2})$ follows a χ^2 distribution with $n-p$ degrees of freedom. We must find the form of the numerator

$$\frac{||\hat{\mathbf{y}}_\mathbf{2} - \hat{\mathbf{y}}_\mathbf{1}||^2}{\sigma^2}$$
$$= \frac{(\hat{\mathbf{y}}_\mathbf{2} - \hat{\mathbf{y}}_\mathbf{1})^\mathbf{T}(\hat{\mathbf{y}}_\mathbf{2} - \hat{\mathbf{y}}_\mathbf{1})}{\sigma^2}$$

This is similar to the previous section where we showed the residuals scaled by variance were approximately chi squared, so $= \frac{\mathbf{e^T(H_2 - H_1)e}}{\sigma^2}$ will have a chi squared distribution with rank of $H_2 - H_1$ degrees of freedom, which is $p_2 - p_1$.

Before we simply take the ratio of these two variables we must show that they have a covariance of zero. If this is the case the distribution of two chi squared variables follows an F-distribution as the errors are assumed to be Normal. If not the problem becomes more complicated. We know that the expectation of $\hat{\mathbf{e_2}}$ is zero so we now need to show that the expectation of $\hat{\mathbf{e_2}}(\hat{\mathbf{y_2}} - \hat{\mathbf{y_1}})^\mathbf{T}$ is zero.

$$E(\hat{\mathbf{e_2}}(\hat{\mathbf{y_2}} - \hat{\mathbf{y_1}})^\mathbf{T}) = E[(\mathbf{1} - \mathbf{H_2})\mathbf{e}(\mu + \mathbf{e})(\mathbf{H_2} - \mathbf{H_1})] = \sigma^2(\mathbf{I} - \mathbf{H_2})(\mathbf{H_2} - \mathbf{H_1}) = H_2 - H_1 - H_1 + H_2 H$$

Hence,

$$F = \frac{\frac{||\hat{\mathbf{y}}_\mathbf{2} - \hat{\mathbf{y}}_\mathbf{1}||^\mathbf{2}}{(p_2 - p_1)}}{\frac{||\mathbf{e_2}||^2}{(n - p_2)}}$$

has an F-distribution with $p_2 - p_1$ numerator degrees of freedom and $n - p_2$ denominator degrees of freedom when both models are correct.

When only the larger model is correct then the denominator still has the same distribution but the numerator is approximately $\mu(\mathbf{H_2} - \mathbf{H_1})^\mathbf{2}\mu$ where μ is the true unknown mean of $\mathbf{y}$. This is not zero when the smaller model is incorrect. Hence this F-statistic is a sensible method to test the hypothesis that the smaller model is correct against the alternative the smaller model is incorrect but the larger model is correct as the value of this statistic will be significantly larger than its expected value if the simpler model is not an appropriate fit.

2.5.2 Nonlinear Regression

Nonlinear regression can be thought of as the generalisation of linear regression to the case where the relationship between predictor and response is not linear. While the basic ideas behind nonlinear regression and linear regression are similar there are some complications which arise due to the nonlinearity of the model.

Estimation and Linear Approximations

A nonlinear regression problem can be written as:

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}; \boldsymbol{\theta}) + \mathbf{e}$$

where $\mathbf{f}(\mathbf{X}; \boldsymbol{\theta})$ is the mean function which is nonlinear, $\mathbf{X}$ are the predictor variables as was the case in the linear model. $\boldsymbol{\theta}$ are the parameters we seek to estimate and as per the linear case, $\mathbf{e}$ are the errors and $\mathbf{Y}$ is the response. Frequently, the following notation is used:

$$\eta(\boldsymbol{\theta}) = \mathbf{f}(\mathbf{x}_i; \boldsymbol{\theta}),\ i = 1, ..., N,$$

which allows the model to be written as

$$\mathbf{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \mathbf{e}.$$

As for the linear case we assume that

$$E[\mathbf{e}] = \mathbf{0} \text{ and } Var[\mathbf{e}] = \sigma^2 \mathbf{I_n},$$

which some may question as there may be serial autocorrelation present in these data, but as this is the method, which has been used in medical imaging for some time, we will use it here. It is also generally assumed the errors are independent, and identically distributed and are Normal as was the case in the section on linear regression. The likelihood function will then be the same as in the linear case and we will seek to minimise the RSS.

In the case of PET data, we cannot assume that the errors are independent and identically distributed as the variance in the data is proportional to the mean. Luckily,using weights is also generalisable to nonlinear regression. The details of this have been omitted from this section as although it does not make the analysis

much more complicated it does make the algebra less tidy. In the case of the PET data which are quasi-Poisson, we have a nonlinear model

$$C_T(t) = v_B C_P(t) + K \int_0^t (\phi_1 e^{-\theta_1(t-s)} + \phi_2 e^{-\theta_2(t-s)}) C_P(s) ds + \frac{1}{\sqrt{wts_t}} \varepsilon_t$$

where $C_T(t)$ is the measured value, v_B is blood volume, $C_P(t)$ is the AIF, the wts are added to take account of the variance as the data are Poisson and the relationship between these and the weights used in the regression are obviously related, ε_t is assumed to be Normal and the other values are estimated by the regression to form the residue function. The parameters estimated here are v_B, and the ϕ and θ values, which are a combination of the rate constants.

One may ask, why would one not use a more direct quasi-Poisson approach and so avoid using weights altogether. It is a very valid point, but as this is the method used almost exclusively in the literature, it was decided to proceed using the same method as others in the field.

Therefore fitting a nonlinear model is similar to the linear case and is based on minimising the RSS. The nonlinearity makes finding this minimum more complicated. In the linear case an explicit formula for the least squares estimate can be written down, while in the nonlinear case $\boldsymbol{\eta}_i(\boldsymbol{\theta})$ defines a surface of dimension equal to the number of parameters we seek to estimate. The point on this surface which is closest to $\mathbf{Y}$ is the point we seek to estimate. It is not possible to write down an expression for the estimate in the nonlinear case so iterative methods are used instead.

The basic ideas of finding the estimator are simple. One finds the point on the surface which is closest to the response and then tries to find what are the values of the parameters at that point on the surface. In the nonlinear case finding this minimum and generating the parameters from the point on the surface may not be straightforward as was the case in linear regression. We will now discuss one method of generating estimators using a Taylor approximation. This is called Gauss-Newton's method [9, 54], which is implemented in the *nls()* function in R [65].

This technique begins by taking an initial guess for the parameters we seek to estimate from the user. This guess can be formulated in many ways. In the case

of PET brain imaging one can look at previous studies estimates for parameters and use these as starting values for the study at hand. Once the starting values θ^0 are provided, $\eta(\theta)$ is simplified using Taylors theorem [1, 9]

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \approx \boldsymbol{\eta}(\boldsymbol{\theta}^0) + \mathbf{V}^0(\boldsymbol{\theta} - \boldsymbol{\theta}^0),$$

where $\mathbf{V}^0$ is a matrix with n rows (number of data points) and p columns (number of parameters we seek to estimate) containing

$$\frac{\partial f(\mathbf{x_i}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_p}$$

evaluated at $\boldsymbol{\theta}^0$ [9].

We now define the residuals

$$\hat{\mathbf{e}} = \mathbf{Y} - \boldsymbol{\eta} \approx \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta^0}) + \mathbf{V^0}(\boldsymbol{\theta} - \boldsymbol{\theta^0}) = \mathbf{e^0} - \mathbf{V^0}(\boldsymbol{\theta} - \boldsymbol{\theta^0}).$$

The next step is to calculate the increment in the RSS. Similar to linear regression a matrix decomposition called the QR transformation of $\mathbf{V^0}$ is useful here. As in the case for linear regression a number of entries in the $\mathbf{R}$ matrix will be zero which allows us to consider a smaller version of $\mathbf{Q}$ and $\mathbf{R}$ called $\mathbf{Q_1}$ (which forms an orthogonal basis) and $\mathbf{R_1}$. By projecting the residuals into $\mathbf{Q_1}$ space the subsequent values for $\boldsymbol{\eta}$ can be found

$$\boldsymbol{\eta}(\theta^1) = \mathbf{Q_1 w_1},$$

where $\mathbf{w_1} = \mathbf{Q_1^T e^0}$.

As $R_1(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = \mathbf{w_1}$, the new value for $\boldsymbol{\eta}^1$ will give a fit that will be closer to the response than the previous estimate for a suitable step size. An important question is when should you stop iterating. Some argue once the relative increment over a step is sufficiently small [9]. However this does not guarantee convergence. Another alternative Bates and Watts propose is to examine the improvement in the residual sums of squares. There are many other alternatives including examining gradients or continuing until the expectation and residual surfaces are orthogonal.

The Gauss-Newton approximation is a linearisation of the nonlinear problem. Naturally this leads us to apply linear inference techniques to the problem. Approximately, $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, (X^T W X)^{-1}\sigma^2)$, which allows us to follow the steps in

linear regression. Recall in the linear case of linear regression a confidence region was defined by

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{X}^{\mathbf{T}}\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq P\hat{\sigma}^2 F(P, N - P; \alpha).$$

In the nonlinear case this becomes

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\hat{\mathbf{V}}^{\mathbf{T}}\hat{\mathbf{V}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq P\hat{\sigma}^2 F(P, N - P; \alpha)$$

or

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\hat{\mathbf{R}}_{\mathbf{1}}^{\mathbf{T}}\hat{\mathbf{R}}_{\mathbf{1}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq P\hat{\sigma}^2 F(P, N - P; \alpha).$$

This also allows us to define an ellipse which defines the inference region for the parameters $\{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \sqrt{Ps^2 F(P, N - P, \alpha)}\hat{\mathbf{R}}_{\mathbf{1}}^{-\mathbf{1}}\mathbf{d}$ such that $||\mathbf{d}|| = \mathbf{1}\}$. Again this is familiar from our discussion of linear models. Inference bands for the response are generated by replacing $\mathbf{x_0^T b}$ in the linear model with $f(\mathbf{x_0 b}, \hat{\boldsymbol{\theta}})$, giving

$$f(\mathbf{x_0 b}, \hat{\boldsymbol{\theta}}) \pm \hat{\sigma}||\mathbf{v_0^T}\hat{\mathbf{R}}_{\mathbf{1}}^{-\mathbf{1}}||t(N - P, \alpha/2).$$

The fact that these results are based on the linear approximation means that these are only approximately true in the nonlinear case.

It has been mentioned that in the linear case that the estimates exist in closed form while this is not the case for nonlinear regression. In fact, in the linear case, the sum of squares function itself can be found analytically. This is not generally possible in the nonlinear case.

It is also worth highlighting that although the Gauss-Newton method is emphasised here as it is widely used and is implemented in the *nls()* function in R [65] there are other methods such as Newton-Raphson which could have been used. This however would be less computationally efficient than the Gauss-Newton algorithm, which does not require second derivatives to be calculated.

2.6 An Illustration with FDG Data from a Brain Tumour Study

We now illustrate the steps involved in turning a PET image into a dataset and fit the compartmental model to this dataset. We begin by displaying a dataset in

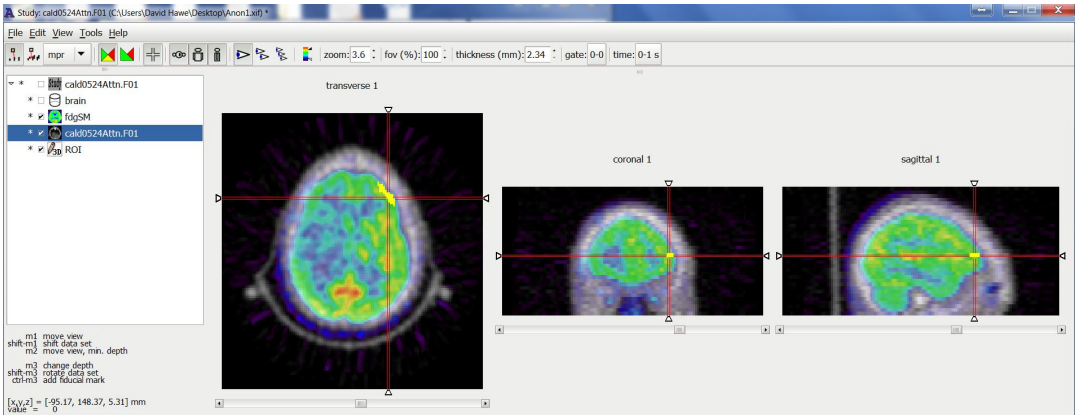


Figure 2.6: *FDG-PET scan generated on a GE Advance scanner shown in AMIDE [42].*

AMIDE [42], where the transverse, coronal and sagittal views are shown in Figure 2.6 with a grey matter region highlighted in yellow. This particular dataset has been reported on in [71]. It is made up of 128 $\times$ 128 voxels for 35 slices over 31 time-bins. The first time-bin is one minute and this is before the patient is injected with the radiotracer. It is followed by four 15 second bins, four 30 second bins, four 1 minute bins, four 3 minute bins and fourteen 5 minute bins. This timing information is important when modelling PET data as it defines the start time, end time and duration of each time-bin. As the data is made up of the number of emissions per time-bin, the number of emissions per bin is divided by the length of the time-bin so that they can be compared with each other. This was discussed in Chapter 1.

An AIF is also available for analysis. This is shown in Figure 2.7. In this case it has been derived by arterial sampling as follows: five measurements 10 seconds apart, two 20 seconds apart, four 30 seconds apart, four 60 seconds apart, two 180 seconds apart and nineteen 300 seconds apart.

AMIDE provides the facility to extract the information in the highlighted ROI to a text file containing the measured values in the region and their coordinates along with an estimate of how accurate the measured values are likely to be. This file can be taken and used in statistical software packages such as R [65] for analysis.

When reading the scan data into R, one must realise that the measured values are in one large column. It is necessary to take the column of measured values

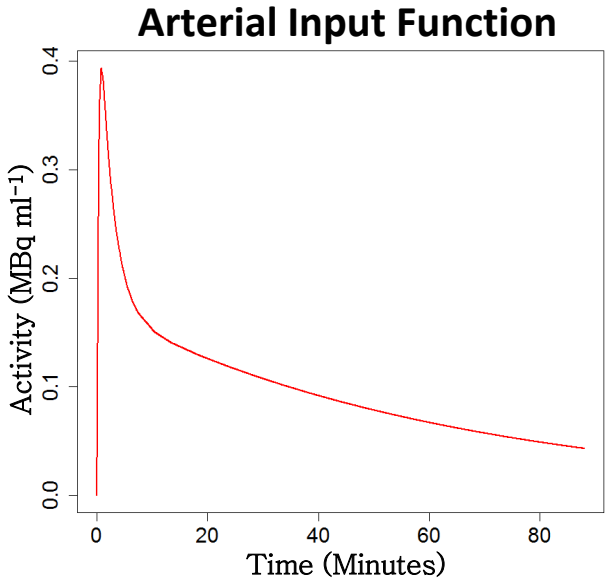


Figure 2.7: *AIF for an FDG PET study.*

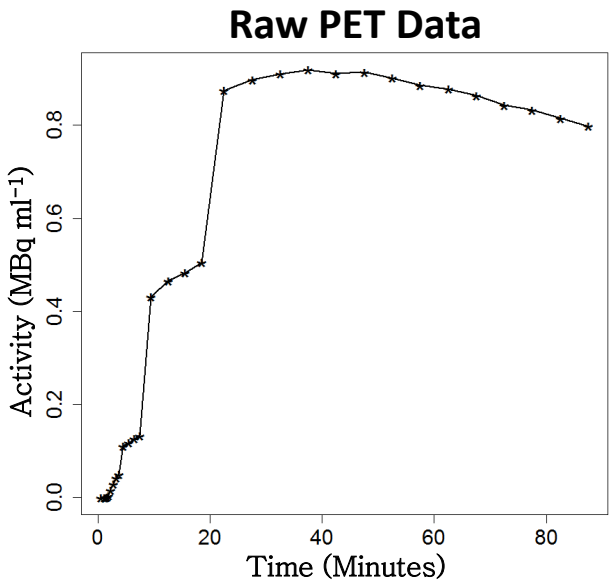


Figure 2.8: *Raw uncorrected FDG-PET counts from PET scan image.*

and transform this into a matrix with the number of columns being the number of time points. This gives a matrix of dimension the number of voxels in the ROI by the number of time points. Generally we average over the ROI to get one value to represent each time point. A picture showing what this looks like is shown in Figure 2.8.

This picture does not seem particularly appealing to the naked eye. It is not

smooth and appears to have steps in the function. Dividing the value at each time-bin by the length of the interval makes this curve smooth. However we are interested in glucose uptake and not in values decreasing as a result of decay. Accordingly, a correction for decay, which is based on a half life of 108 minutes is made. Following this decay correction and a correction for the duration of the time-bin we then arrive at the time activity curve which was introduced in Chapter 1 of this thesis. This is shown in Figure 2.9.

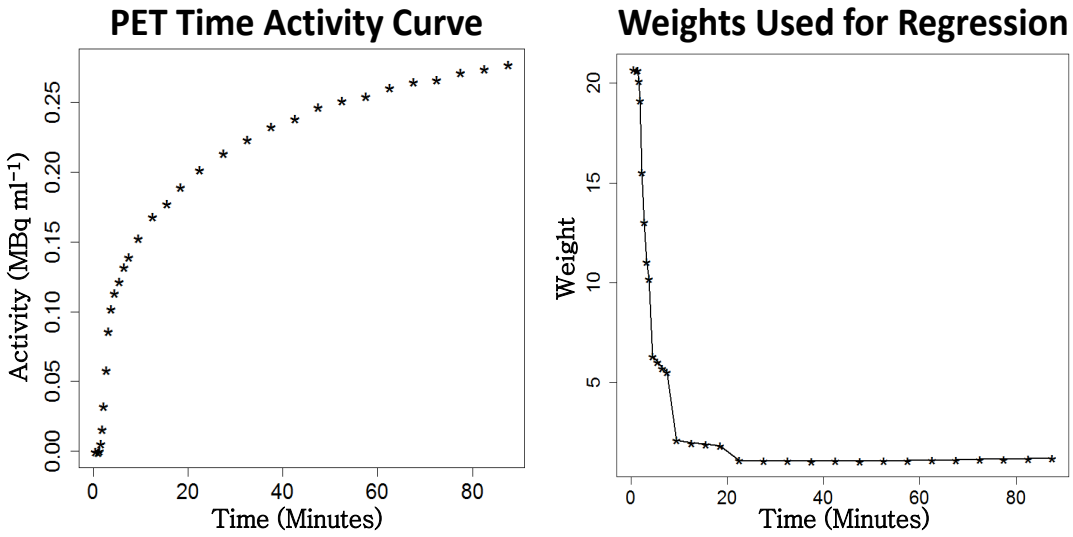


Figure 2.9: *Time activity curve corrected for decay and duration and weights used in nonlinear regression.*

We have previously discussed the assumptions of regression in Section 2.5. One of these was that the variance of the errors was constant. In the case of PET data, the variance of different time bins will not be the same as the data are counts over intervals of varying intervals. In implementing this model, we need to chose weights for the regression. These weights are shown in Figure 2.9. As the data are quasi-Poisson the weights chosen were

$$\frac{1}{z_k + \varepsilon},$$

where z_k is the measured PET data at the k^{th} time bin. The ε, which is one tenth of the average of all measured PET values is added in to prevent dividing by zero errors [28, 45, 60].

The model is fitted using the following R code. Full details of the $modx1$ and $modx2$ functions which represent the AIF and tissue compartments respectively.

```
nlmod <- nls(z ~ cbind(modx1(ts,te,tp,cp,hlife,del),
    modx2(ts,te,tp,cp,hlife,del,k2)),weights = wts,
    start=list(del=.6,k2=.03),
    algorithm = 'plinear',trace = TRUE,
    control=nls.control(maxiter=100,warnOnly=TRUE) )
```

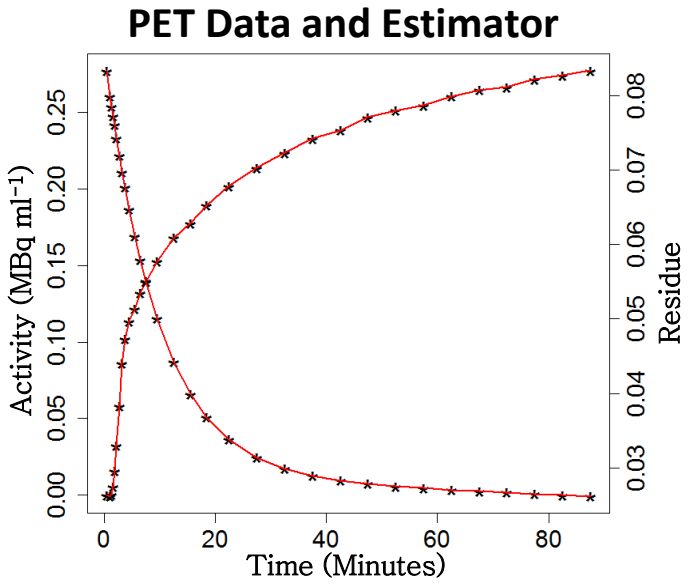


Figure 2.10: *Two compartmental model fit to FDG-PET time activity curve along with estimated residue function. The black stars are the data and the red line is he compartmental model fit. Both lines on the residue function (red line and stars) are the compartmental model estimator.*

Having finished running this code an estimate is generated for the data based on the compartmental model, which is shown in Figure 2.10. This fit gives a number of estimates for parameters of interest along with standard errors and these are presented in Table 2.1. It is worth noting that sometimes it is advantageous to scale parameters to have similar ranges, however to make these results as aligned as possible, this was not done here [68].

The key to the work in this thesis is the residue function which allows the functionals reported above to be calculated. This residue function is also presented in Figure 2.10.

In Weisberg [80], which has been reviewed earlier in this chapter, it was stated that after generating a fit it is not acceptable to simply say that the job is done

Flow (SE)	Flux (SE)	V_D (SE)	V_B (SE)
0.083 (0.007)	0.026 (0.003)	3.03 (0.1009)	0.18 (0.01)

Table 2.1: *Parameter Estimates along with standard errors in parenthesis for flux, flow, volume of distribution (v_D) and blood volume(V_B) for the FDG glioma region.*

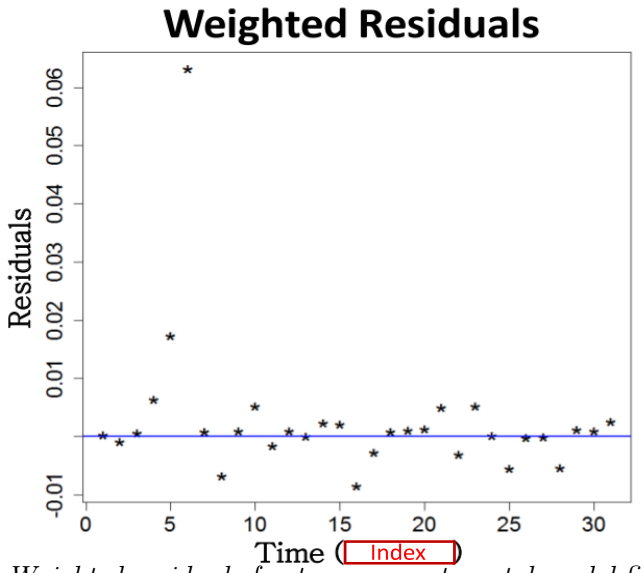


Figure 2.11: *Weighted residuals for two compartmental model fit to FDG-PET data, note the x-axis is an index not time.*

and move on. Analysis of the fit should be undertaken to see how appropriate it is. We begin this by examining the plot of the weighted residuals for the fit. The weighted residuals are

$$\frac{1}{\sqrt{wts_k}}(z_k - \hat{z_k}),$$

which are clearly related to the weights used to carry out the regression in this chapter.

and these are shown in Figure 2.11. This fit appears to have an issue as the values of the estimator appear to have a lot of variability at early time points in the scan with two very strong outliers showing a systematic lack of fit here. This is addressed in Chapter 3, which proposes a nonparametric residue function. It also appears that there may be a pattern in the residuals at later times. This plot is far from a null plot, which would be a positive indicator for the model fitting the

Source	DF	SS	MS	F	P
SSReg	6	0.084745	0.01412425	0.020008	0.0063
Residual	24	16.94159	0.705899		
TSS	30	17.02634	0.0567544		

Table 2.2: *ANOVA table for compartment model fit to PET data.*

data.

Another method in examining fits of models is to consider ANOVA. This has already been discussed in this chapter and here we present the ANOVA table in Table 2.2.

The chi-square test [66] is used to test if a sample of data came from a population with a specific distribution. For the chi-square goodness-of-fit computation, the data are divided into k bins and the test statistic is the RSS. More succinctly, if we consider a χ^2 test to determine the goodness of fit, we are testing the hypothesis that the RSS scaled by the variance follow a χ^2 distribution as is the case if the errors are normal. In the case of this data set, we find that this hypothesis is rejected with a level of significance of 0.01.

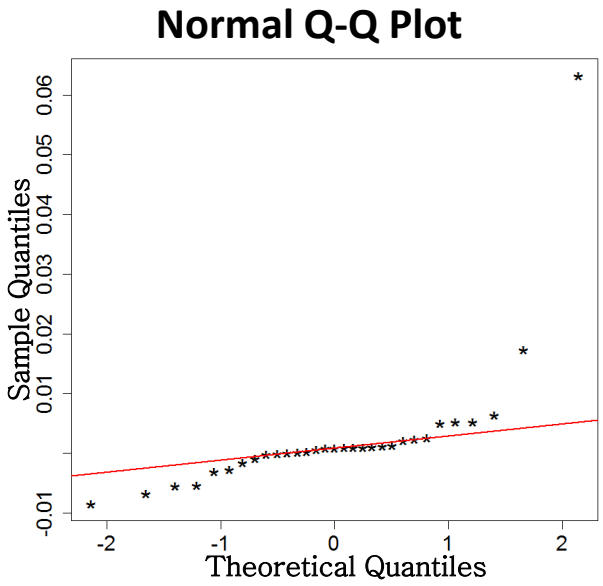


Figure 2.12: *QQ-plot for model residuals.*

Finally, we examine the residuals to see if they appear normal using a QQ-plot

shown in Figure 2.12. This is a plot of the theoretical normal quantiles against the quantiles in the data. If the residuals are approximately normal, then this plot should result on the data being approximately on the line $y = x$, which is the red line in the plot. Clearly, as this does not appear to be the case here the Shapiro-Wilkinson test rejected the hypothesis at the 0.01 level. It is also worth noting that the two outliers are present here again on the QQ plot as on the weighted residuals.

The assumptions about the compartment model have been outlined at the beginning of this chapter and even here in the case of a relatively small region of interest, the assumptions clearly do not hold. The data does not behave as a well stirred compartment. The results lead us to seek an alternative formulation for the residue function, which comes out of compartmental models in this chapter.

Practical Considerations in this Nonlinear Regression

There are numerous difficulties associated with nonlinear regression. This section outlines some of those associated with the work presented here and techniques used to circumvent these difficulties. One case where the assumptions of the distribution of the errors are broken is when the variance is not constant. Testing to see if the variance is constant is possible when the data contains a large collection of points and has a lot of repetition. Even when this is not the case it is possible to make a plot of residuals and see if constant variance seems plausible. It is also possible to use analysis of variance techniques to see if the variance is constant. In the case where the variance is not constant, weights can be used as was the case in linear regression. Taking advantage of the quasi-Poisson nature of PET data allowed this be dealt with.

When the Gauss-Newton algorithm was first introduced, we mentioned that it is necessary to provide an initial guess for the first iteration. This is not trivial. If the initial values are not sufficiently close to the true answer convergence to the wrong minimum is possible and this happened many times in analysing the data contained in this thesis. In the illustration in this chapter fitting PET compartmental models will be demonstrated. For a single study it took hours of trial and

error to generate starting values which converge to the correct minimum. Starting values are found by using values consistent with previous studies and then varying these values until convergence is obtained. It is also worth highlighting that the parameter which appears to have been the main source of issues in doing this analysis is delay.

Aside from these concerns it is appropriate to add constraints to some parameters such as the percentage of blood volume in a region of tissue. Naturally, this should be positive. To accomplish this the *nls()* procedure is implemented with an order restricted function from the $ic.infer$ package in R [25, 65] or with constraints, which the function is designed to use. Using the $ic.infer$ package has also been found to make the model less sensitive to finding starting values. This is not surprising as it reduces the number of parameters in the nonlinear regression part of the estimation by estimating the coefficients of the nonlinear functions separately.

Obtaining convergence is often difficult and when this difficulty occurs one should consider if the mean function is correct, has the data been read in correctly, are there any outliers and are the starting values appropriate? If the answer to all these is correct then getting the values of the parameters and the RSS at each iteration can be useful. In R this is accomplished by including *'trace=TRUE'* in the *nls()* procedure. This allowed an issue with the delay term becoming approximately 10 minutes when blood volume was low to be uncovered and avoided.

Appendix

A Solution of One Compartment Model Ordinary Differential Equation

From Figure 2.2, a coupled set of differential equations can be written down to model deterministically the system of interest assuming the law of mass action [26, 27, 34, 39, 63]. These first order linear ordinary differential equations are presented as Equations A1 and A2. These equations explain how the radiotracer moves between blood and tissue.

$$\frac{dC_1}{dt} = K_1 C_P(t) - k_2 C_1(t) \tag{A1}$$

$$\frac{dC_P}{dt} = k_2 C_1(t) - K_1 C_P(t) \tag{A2}$$

Here, $C_P(t)$ and $C_1(t)$ are the concentrations of tracer in the blood and tissue respectively. K_1 and k_2 are the rate constants. As $C_P(t)$ is known due to arterial sampling or otherwise, the equation regarding $C_P(t)$ can be disregarded and we simply need to solve A1, which is a first order linear ordinary differential equation and can be solved by the technique of an integrating factor [10]. Multiplying A1 by $e^{k_2 t}$ gives

$$e^{k_2 t}\frac{dC_1}{dt} + e^{k_2 t} k_2 C_1(t) = K_1 e^{k_2 t} C_P(t) \tag{A3}$$

$$\frac{d}{dt}(e^{k_2 t} C_1(t)) = K_1 e^{k_2 t} C_P(t) \tag{A4}$$

Integrating both sides gives

$$e^{k_2 t} C_1(t) = \int_0^t K_1 e^{k_2 s} C_P(s) ds \tag{A5}$$

and hence

$$C_1(t) = \int_0^t K_1 e^{-k_2(t-s)} C_P(s) ds. \tag{A6}$$

B Solution of Two Compartment Model Ordinary Differential Equation

In the case where the tracer undergoes a reaction in tissue a second tissue compartment is added to the simpler model. This model is discussed by Sokolov *et al.* and Phelps *et al.* in [63, 70] and a schematic of this model is shown in Figure 2.3. Similar to the Kety Schmidt model, one can write down continuous deterministic differential equations for the system using the law of mass action.

$$\frac{dC_P}{dt} = k_2 C_1(t) - K_1 C_P(t) \tag{B1}$$

$$\frac{dC_1}{dt} = K_1 C_P(t) - (k_2 + k_3) C_1(t) + k_4 C_2(t) \tag{B2}$$

$$\frac{dC_2}{dt} = k_3 C_1(t) - K_4 C_2(t), \tag{B3}$$

Huang *et al.* [34] discuss this model and use a method similar to the one compartmental model, where the concentration in the plasma is assumed to be known and simply solve this equation for C_1 and C_2 given a known C_P. This is sensible as in the case of a PET study the arterial input is usually known. Equations B2 and B3 are a coupled system of order linear ordinary differential equation and can be solved using Laplace's transformation [10].

Decoupling this set of differential equations for the two compartmental model is more complex than simply using an integrating factor to solve a single equation with a single unknown for the one compartmental model. To solve this equation it is necessary to transform the data into Laplace space using the Laplace transformation.

$$SC_1(S) = K_1 C_P(S) - (k_2 + k_3) C_1(S) + k_4 C_2(S),$$

and

$$SC_2(S) = k_3 C_1(S) - K_4 C_2(S).$$

The concentrations $C_1(t)$ and $C_2(t)$ are referred to as $C_1(S)$ and $C_2(S)$ in Laplace space.

We assume that the initial conditions in tissue are zero and solve this set of equations, which leads to:

$$C_1(S) = \frac{K_1}{\alpha_2 - \alpha_1}(\frac{k_4 - \alpha_1}{S + \alpha_1} + \frac{\alpha_2 - k_4}{S + \alpha_2})C_P(S)$$

and

$$C_2(S) = \frac{K_1 k_3}{\alpha_2 - \alpha_1}(\frac{1}{S + \alpha_1} - \frac{1}{S + \alpha_2})C_P(S),$$

where $\alpha_1 = \frac{k_2+k_3+k_4-\sqrt{(k_2+k_3+k_4)^2-4k_2k_4}}{2}$ and $\alpha_2 = \frac{k_2+k_3+k_4+\sqrt{(k_2+k_3+k_4)^2-4k_2k_4}}{2}$. Taking the inverse Laplace transformation gives:

$$C_1(t) = \frac{K_1}{\alpha_2 - \alpha_1}((k_4 - \alpha_1)e^{-\alpha_1 t} + (\alpha_2 - k_4)e^{-\alpha_2 t}) \otimes C_P(t)$$

and

$$C_2(S) = \frac{K_1 k_3}{\alpha_2 - \alpha_1}(e^{-\alpha_1 t} - e^{-\alpha_2 t}) \otimes C_P(t).$$

The concentration in tissue is made up of $C_1(S) + C_2(S)$, which is

$$C_1(S) + C_2(S) = (\frac{K_1(k_4 - \alpha_1 + K_1 k_3)}{\alpha_2 - \alpha_1}e^{-\alpha_1 t} + \frac{K_1(\alpha_2 - k_4) - K_1 k_3}{\alpha_2 - \alpha_1}e^{-\alpha_2 t}) \otimes C_P(t)$$

which is of the form

$$(\phi_1 e^{-\theta_1 t} + \phi_2 e^{-\theta_2 t}) \otimes C_P(t),$$

which is the mixture of exponentials fitted as the residue in the compartmental modelling.

Chapter 3

A Nonparametric Model for the Tissue Residue Function

In the previous chapter, compartmental models were introduced, applied to PET data and a discussion on the assumptions underlying compartmental models was presented. The fact that these assumptions are frequently violated in practice was highlighted, which lead to a Markov chain formulation for blood tissue exchange. The link between the Markov and compartmental models was established in the limiting case. The work in the previous chapter was undertaken in an effort to better understand the compartmental models from a probabilistic approach and it is clear from this work that some of the assumptions behind the compartmental models are not always valid.

This chapter follows on the theme of trying to better understand the intricacies of compartmental models and also looks for alternative modelling techniques. To this end, a novel nonparametric method which does not have the underlying assumptions that compartmental models do will be presented and applied to data from ongoing PET data from the University of Washington, Seattle's PET cancer imaging program. The compartmental model will also be applied to these data and comparisons made between these models.

3.1 Outline and Contribution of this Chapter

In Section 3.2, the methodology associated with the nonparametric model will be discussed. This begins by examining the residue function, which has previously been introduced in Section 2.2 and this is discussed in the context of survival analysis in Section 3.2.1. Following this a review of Kaplan and Meier's work on survival analysis will be discussed and this leads us to propose a piecewise constant form for the tissue residue function. This alternative model which has been proposed in Hawe *et al.* [28] is a large part of the contribution of this thesis.

Details about inference for this type of nonparametric regression will be presented in Section 3.2.2. The parameters of interest generated from this model will briefly be discussed in Section 3.2.3 in the context of how they relate to the compartmental model estimates. Methods of comparing the models will then be presented in Section 3.2.4. A second section on methodology for simulation studies used in understanding these models will appear in Section 3.3 and the results of this study are presented in Section 3.4. These results have an impact on the use of methods from ANOVA being used to compare nonparametric and nonlinear models. An illustration of the model being used will be presented in Section 3.5. This illustration will look at residuals and examine where the model does and does not appear to fit the data and illustrate how to decide which model performs better. We finish this chapter with some discussion and conclusions.

3.2 Nonparametric Model

3.2.1 Residue Functions as Survival Curves

In the last chapter, the tissue residue function was introduced in the context of the solution of compartmental models. It was mentioned that physiologically, the residue can be thought of as the amount of radiotracer remaining in a tissue ROI in response to an idealised bolus, i.e. a Dirac-delta function. Perhaps this can be described more intuitively as being a function which starts at one and decreases as the cumulative distribution of transit times of atoms through the ROI over time is subtracted until there is no tracer left in the ROI or the scan concludes

[60].

If this is the case then the function must be monotonic starting at unity and decreasing with right continuity until it either reaches zero, an equilibrium or the scan concludes. In the case of compartmental models, the residue function is a mixture of exponentials. This obviously satisfies the above criteria. However this is based on the compartmental assumptions, which have already been discussed and are found to be highly questionable. Hence an alternative formulation for the residue is sought which satisfies the criteria of above without having these questionable assumptions. In examining assumptions around what a residue function is, it becomes apparent that these assumptions are the basis for life tables in survival analysis. We now seek to derive an alternative nonparametric approach to model the residue using techniques in survival analysis.

Alternatives to compartmental models have been proposed by Cunningham and Jones [14] and also by Murase [50], who examined spectral analysis, and O'Sullivan *et al.* [60], who introduced a nonparametric estimate for the transit time of tracers through a tissue ROI, and hence the residue, using splines. In all three papers, the nature of the residue being a survival function is preserved.

In 1958 Kaplan and Meier [38] published a nonparametric method for displaying survival analysis data. This is significantly different to alternative methods, which include fitting exponential curves to data. The Kaplan Meier curve is a nonparametric piecewise constant graphical representation of survival. At each time step the proportion of subjects in the study which are still alive is calculated and this value represents a survival rate at that time bin (the proportion used excludes those who have just left the study). This proportion is the number added to the plot to represent the duration of the time-bin. Obviously, the starting value at time zero for this estimator will be unity. This function will hence be always positive and monotonically decreasing. This curve has nice properties for survival analysis in that it is not generally influenced by people withdrawing from a drug trial as it uses the proportion of those remaining in the study to generate the estimator rather than the proportion of those who started, however when a subject withdraws (elects to leave), a vertical mark is made on the plot to highlight a subjects departure from the study.

In the case of PET we are interested in this form as it can represent the

amount of tracer remaining in tissue. Using a form similar to this we will now propose a new PET residue function. While the Kaplan-Meier curve is a graphical representation of survival, there is a clear similarity with the survival problem in the case of a PET study. The method proposed here is a term in a model rather than a graphical representation for survival. However both methods quantify survival or residence in terms of a piecewise constant function. As an aside I highlight that perhaps Meier himself was inspired by survival analysis problems when he wrote his seminal paper along with Zierler [47] on indicator dilution theory already cited earlier in this work, which was the proof of the central volume theorem.

There are a number of advantages to a piecewise constant residue function such as the one proposed here. In the case of compartmental models, there are numerous, sometimes questionable, assumptions behind the modelling. Here all the assumptions that are being made are that the residue is constantly decreasing and is piecewise constant. At each time point we are estimating the tracer residue in tissue. There is no known basis for the form of the residue between time-bins so as an estimator a straight line is used. Other methods such a basis of cubic splines or other smoothed curves could have been used instead but as there is no justification for any type of function, the simplest piecewise constant form seems appropriate. Also, the nonparametric method can be rapidly implemented in R [65] using the ic.infer package [25].

Over the observation window $[0, T_e)$, let $0 = \tau_0 < ... < \tau_J < T_e$ be a collection of $J + 1$ points (defined by the time bins of the PET scan) and let I_j^+ be the indicator for the set $[\tau_j, \infty)$. For $\alpha = (\alpha_1, ...\alpha_J)^T$, a vector with non-negative components satisfying $\sum_{j=1}^{J} \alpha_j \leq 1$, a piecewise constant residue can be defined over $[0, T_e)$ by

$$R_\alpha(t) = 1 - \sum_{j=1}^{J} \alpha_j I_j^+(t). \tag{3.1}$$

The residence density corresponding to this residue is discrete with mass α_j concentrated at τ_j. This somewhat complicated mathematical formulation can be described in a simple way. The residue begins at one and at each time point can drop by any nonnegative amount as long as it does not result in the function being negative.

3.2.2 Statistical Inference for the Nonparametric Residue

Using the methods in [25], implemented in R [65], we will fit the nonparametric model to data. While the methods used here are different to the compartmental model, a lot of the information we previously introduced is still important. We still use the Poisson nature of the data in our modelling. That is to say, variance is approximately proportional to the mean and over homogeneous regions the total activity tends to scale linearly with the size of the region considered. For this reason we will use a weighted least squares approach to the problem. There are a number of reasons for this. One is that we seek a direct comparison to the compartmental model. If we carry out transformations in the data for the nonparametric fit this will not be done for the compartmental model and so we are not comparing like with like. Secondly, the quadratic programming algorithm used for this nonparametric modelling is set up for the weighted least squares problem and as it has been tested and validated it makes sense to use it.

The time-bins for a set of N dynamic scans will be denoted $(t_{sk}, t_{ek}]$ for $k = 1, 2, ..., N$. As we have previously outlined in Chapter 2 there maybe biological and other sources of deviation between the model and the data, the deviation of the data from the model is considered a sum of measurement and biologic error. Thus a reasonable statistical model for the data is

$$z_k = C_T(t_k|\theta) + \hat{\sigma}\frac{1}{\sqrt{w_k}}\varepsilon_k, \tag{3.2}$$

where w_k are the same weights used in the previous chapter, which were the reciprocal of the measured values plus a term to prevent divide by zero errors. In practice the true variance σ is unknown and so $\hat{\sigma}$ is an estimator of the variance. An appropriate estimator for this would be the median absolute deviation of the square root of the weighted residuals. The ε_k's are modelled as a random sample from a standard normal distribution. The Poisson nature of the underlying emissions makes the assumption of independence not unreasonable and when the expected counts are large (a function of injected dose), a Gaussian approximation of the Poisson process is also reasonable. Based on Equation 3.2, inference for parameters is based on the WRSS

$$WRSS(\theta) = \sum_{k=1}^{N} w_k(\theta)[z_k - C_T(t_k|\theta)]^2.$$

For the nonparametric method, we use an R function written by Gromping [25] called order restricted linear regression $orlm()$, which allows a linear model to be written with constraints. In this work we can stop the α_i values from being negative, making sure that the sum of all α's is less than or equal to unity and that $\alpha_1 - \sum_{i=2}^{N} \alpha_i > 0$. This is formulated as an estimation problem for v_B and the α_i values be the parameters we seek to estimate in the regression. Apart from Δ, the unknowns in the model enter linearly and so conditional on weights, the estimation process is as follows:

Initially delay is estimated separately by a line search, then with $\beta = (v_B, K, K\alpha_1, K\alpha_2, ..., K\alpha_J)$ so $\theta = (\Delta, \beta)$

$$C_T(t_k|\Delta, \beta) = X_{k1}\beta_1 + X_{k2}\beta_2 + ... + X_{k(J+2)}\beta_{J+2}$$

where $X_{k1} = C_p(t_k - \Delta)$, $X_{k2} = \int_0^{t_k} C_p(s-\Delta)ds$ and $X_{k(j+2)} = \int_0^{\max(t_k-\tau_j,0)} C_p(s-\Delta)ds$ for $j = 1, 2, ..., J$. The components of β are non-negative and satisfy the linear constraint $\sum_{j>2} \beta_j \leq \beta_2$. Thus for any fixed Δ, $orlm()$ can be used to obtain the unique optimal $\boldsymbol{\beta}$ -value. The implementation of this in R [25] is highly efficient. It is important to highlight that the compartmental model residue function estimates four parameters, two linear type parameters and two exponentials. In the case of the nonparametric model the number of parameters, which we can seek to estimate is equal to the number of time-bins. This means that the nonparametric model has greater degrees of freedom.

Nonparametric estimates of flow, flux, v_D and other parameters can be obtained from the nonparametric residue, and sampling variation (standard errors and biases) derived by simulation. These will be discussed further in the section on parameters of interest. In terms of inference standard errors for parameters generated by the nonparametric method are also calculated using the simulations.

Since the acquisition time-bins are contiguous ($t_{sj} \equiv t_{e(j-1)}$ for $j = 2, .., N$), the nonparametric residue can be readily computed. We take $\tau_j = t_{s(j+1)}$ for $j = 1, ..., (N-1)$. Under the sampling model in Equation 3.2, the piecewise constant residue will be shown to be consistent in dose (with parametric convergence) later in this chapter.

3.2.3 Functional Parameters of Interest

As was the case in the compartmental model we seek to estimate parameters of interest. How these are calculated in the nonparametric case is presented here. Most of this is analogous to the compartmental model estimates. v_B is estimated as the coefficient of the AIF in the model. Flow is again estimated as the start of the residue function, however, comparing the flow values from the nonparametric and compartmental model directly may be erroneous as the average value of the residue over the first time-bin may be a more appropriate comparison as the nonparametric model cannot vary over time-bins. v_D can again be estimated as the integral of the residue function. The metabolic flux ($K\varepsilon$) is also calculated by scaling extraction, which is again the last time point of the residue scaled by flow. A similar argument holds here relating to the nature of the shape of the nonparametric estimator of flux and extraction as has already been described for flow.

3.2.4 Model Comparisons

3.2.4.1 Bootstrapping

We have already mentioned ANOVA and the F-statistic used in comparing models in the context or linear regression [80]. Here the statistic of interest is

$$\text{Model Comparison} = \frac{(WRSS_{2C} - WRSS_{NP})/(df_2 - df_{NP})}{WRSS_{NP}/(n - df_{NP})},$$

where $2C$ represents the two compartmental model and NP is the nonparametric model. In linear regression with nested models, this has been validated. In non-linear regression with nested models this result holds approximately. However in this work, where the models are not nested and one of the models is nonparametric, its use may not be appropriate. Accordingly, we look for an alternative way of finding the distribution of this statistic in the case where the null hypotheses that the compartmental model is appropriate is true. Reflecting on the fact that the F-distribution used in ANOVA is the distribution under which the null is true. We seek to manufacture the distribution without using the assumption that the sampling distribution of the extra sums of squares ESS follows an F-distribution.

This is accomplished using the bootstrap [18, 28, 60]. We begin by fitting the compartmental model and nonparametric model to real data. In doing this we can do two things. Firstly we can calculate the F-statistic, called ESS_0, and secondly we can use the compartmental model fit to the data to generate simulations under which the compartmental model is correct. We do this as follows: We begin by simulating data using the compartmental model fit as the basis for this simulation as defined by:

$$z_{simk} = \hat{z}_{cmk} + \frac{1}{\sqrt{w_k}}\hat{\sigma}\varepsilon_k,$$

where z_{simk} is the simulated data, $\hat{z}_{cmk}$ is the compartmental model fit to the real data, w_k are the weights used in the modelling, which are proportional to the reciprocal of the measured data values, $\hat{\sigma}$ is an estimate of the variance. This estimate is the median absolute deviation of the square root of the weighted residuals, as this means the level of noise in the data will be approximately the same in the simulated data as in the real data and ε_k are random numbers drawn from the normal distribution with zero mean and variance of unity. The compartmental model is used to simulate data as it is seen as appropriate to simulate the data from the "smaller model". Although these are not nested models, we seek to test the null hypothesis that the compartmental model is the correct model. Obviously the new model proposed here must be the alternative.

As the null hypothesis is the compartmental model is the correct model, we seek a reference distribution under which this is true. If the data simulated are truely compartmental model data then we have this.

Once this data is simulated, the ESS statistic on each simulated dataset is calculated. This gives a value for ESS in a case where the compartmental model is an appropriate fit to the data. By repeating this a distribution can be generated which shows approximately what the shape the reference distribution should be if the compartmental model is the true model. By counting the number of simulated ESS values which are greater than ESS_0, and dividing by the number of simulations, we generate a p-value testing the hypothesis that the compartmental model is appropriate.

3.2.4.2 Cross Validation

Using the bootstrap as was the case in the previous section has some drawbacks, one of which is if one was analysing a large dataset, models need to be fitted to data thousands of times. Also the ESS statistic is generally used in the case of nested models and so an alternative method of comparing models is added to our analysis, in the hope that both provide similar results, which would validate our conclusions further. Accordingly we consider an alternative. One such alternative is cross validation which will be familiar to those working on linear regression problems.

Cross validation is based on applying a model to a dataset, on which it has not been calibrated to see how it performs. Generally the dataset is partitioned into the training dataset, which the model is calibrated on, and the validation dataset, on which it is tested. In this work we are interested in a slightly different version of cross validation called leave out one cross validation [22]. This method involves using one observation as the validation set and using the remaining observations as the training set. This is repeated on every data point in the set. In the case of the PET data, we exclude one point in the time-course at a time and fit the model to the dataset without using this point. The residual at this point is then calculated and recorded. This is repeated for every point.

Once this is done we have a set of cross validated residuals of length the number of time-bins in the scan to compare the two models, but how to do this is the next question. It may not be the case that the residuals are normal and so we turn to a nonparametric test. The Wilcoxon signed-rank test is a nonparametric statistical hypothesis test used when comparing two related samples or a single sample to assess whether their population mean ranks differ (i.e. it is a paired difference test). This can test whether the two sets of residuals are of the same magnitude.

An alternative way to compare these models is by calculating the cross validated RSS and determining which is higher. The model with the larger cross validated RSS is considered the weaker model in this case. Using cross validation and bootstrapping will be used to compare models will be illustrated later in this chapter.

3.3 Statistical Consistency of Residue Approaches

3.3.1 Background

A simulation study is a numerical technique for conducting experiments. One example of a simulation study has previously been presented is using simulated data to estimate the improvement in a fit statistics when comparing models. This section gives a more general outline of simulation studies carried out to examine the parameter estimates for the compartmental and the nonparametric models and also discusses the improvement in fit statistics further. In this section uses of the bootstrap for generating properties of estimators and testing hypothesis are discussed. A key reference for bootstrapping studies is found in Efron and Tibshirani [18].

Simulation studies are undertaken as it is vital that properties of methods are understood so they can be used with confidence. For linear regression, many results can be written analytically but in more complicated modelling, this is not always the case. While it would be ideal to have an analytical solution to every problem, a simulation study with a large sample may meet most requirements.

Nonparametric methods for analysing PET data have already been proposed in this chapter and in [14, 28, 50, 60]. This section builds on this work with a view to better understand when the nonparametric technique is appropriate and when the compartmental model should be used. To this end, true compartmental model data is simulated at varying levels of noise and dose, and then analysed. Our aim is to determine which model fits the data better in the case where the null hypothesis, that the compartmental model is appropriate, is true.

If it were the case that true compartmental model data was simulated and the nonparametric model were to be found to be more accurate than the compartmental model then it would show that the nonparametric method is more accurate in general, however this would be an unrealistic expectation. In reality, we seek to examine how the models perform in this when the null is true and expect the compartmental model will outperform the nonparametric method here.

The bias of an estimator is the expectation of the difference between this estimator's expected value (usually mean but the median can be used) and the true value of the parameter being estimated [66]. If this value is zero, the estimator is said to be unbiased. The MSE of estimators has already been discussed in the context of linear regression. A consistent estimator has the property that as the number of observations increases, the estimator converges in probability to the true parameter. A sequence X_n of random variables converges in probability towards the random variable X if

$$\forall \varepsilon > 0, \lim_{n \to \infty} \Pr\left(|X_n - X| \geq \varepsilon\right) = 0.$$

Consistency is related to bias in that consistent estimators are convergent and asymptotically unbiased, though individual estimators in a consistent sequence may be biased.

It is also worth highlighting that the variance plus the bias squared defines the mean squared error (MSE). Frequently people refer to the trade off between bias and variance of an estimator. By calculating the average bias and variance, at varying levels of noise, a trend can be evaluated, which should show variance decreasing with noise and may also show a pattern for bias.

It has been shown previously that the values of parameters can vary between the models [28, 59]. This is generally most evident for flow as will be illustrated later in this chapter, by conducting a simulation study showing how each parameter behaves. The parameter's true values can be compared to their estimated values and then both the bias and variance for each parameter can be determined.

An estimator or test statistic has a true sampling distribution under a particular set of conditions such as errors being normal in linear regression. Ideally, we would want to know this true sampling distribution in order to do statistical inference. This is important, as for example, in the case of linear regression, estimates for standard errors, confidence intervals and hypothesis tests depend on the assumption of normality. Deriving the true sampling distribution of statistics is not always trivial. Earlier in this chapter, the ESS statistic, used to determine whether the parametric or nonparametric model was appropriate was described. In the case of a linear model the distribution of this statistic is known but in the case of more complicated models such as a nonparametric model, it remains to

be validated. The simulation studies to examine convergence used here are also used to gain an insight into the distribution of the ESS.

Another advantage of simulation studies is that while testing models, the data generated is known. This means that the focus of assessing the model's performance can be focused on without worrying about the data itself. By this we mean that the data is known to be truely compartmental or nonparametric. It is not the case that real data from a PET scanner is being examined and the true distrubution is not known or needs to be examined.

3.3.2 Simulation

This chapter aims to propose, present and develop a better understanding of how the compartmental model and nonparametric models work and to analyse the models performance. In chapter 2 the fundamental equation of dynamic PET radiotracer imaging was presented, which models the activity in an ROI as

$$C_T(t|\theta) = v_B C_P(t-\Delta) + K\int_0^t R(t-s)C_P(t-\Delta)ds, \tag{3.1}$$

where $C_T(t|\theta)$ is measured as activity per unit volume v_B and $C_P(t)$ is activity per milliliter of blood, K is flow, $R(t)$ is the tissue residue function and θ is the vector of parameters we seek to estimate $(v_B, \delta, K, k_2, k_3, k_4)$, which are the fractional blood volume, delay and rate constants which define the residue. In practice the constants themselves are not estimated but instead we estimate a linear combination of these which are required for generating the residue.

In Chapter the 2 errors in statistical modelling and residuals were presented. Earlier in this chapter a form for simulating data based on the median absolute deviation of the square root of the weighted residuals as an estimator for the variability in the data and the the assumption that these data are quasi-Poisson was presented. This along with along with an ensemble of random numbers drawn from the normal distribution could generate appropriate estimates of the errors in the data. This was readily accomplished using

$$z_k = C_T(t_k|\theta) + \sigma w_k^{-\frac{1}{2}}\varepsilon_k,$$

where z_k is the simulated value at the k^{th} time point, $C_T(t_k|\theta)$ is the estimator or the time activity course (using the compartmental model), sigma is the median

absolute deviation of the square root of the weighted residuals from the compartmental model fit to the data (a robust estimator of the variance in the data), w_k are the weights and $\varepsilon_k \sim N(0,1)$. Due to the Poisson type nature of the data, the weights are inversely proportional to the measured values. This is because the variability is proportional to the mean of the data and hence $\sigma w_k^{-1/2}\varepsilon_k$ is an appropriate estimator of the noise in the data. In the earlier work this was a suitable way to generate data as our goal was to approximate the noise generated by the true data [80] assuming the compartmental model was appropriate.

In this study, the method of simulating data will be slightly different, although based on the same principle. Firstly, this study will involve varying degrees of noise not just noise levels estimated a typical FDG glioma study. Taking a wider variety of noise levels such as those with more noise and less noise than the glioma set will give a fuller picture of how the models behave. Using the fact that the data are Poisson and that the mean and variance are proportional, we write:

$$z_k = \tau C_T(t_k|\theta) + \sqrt{\tau C_T(t_k|\theta)}\varepsilon_k,$$

where τ can be interpreted as the dose of radiotracer. If τ is varied, the level of noise is changed and this leads to a realistic case where higher doses result in a smoother simulated time activity curve. A slightly simpler version of $C_T(t_k|\theta)$ will be used in this chapter to simulate data. It will be assumed that the simulated region of interest does not have any blood volume in large arteries, i.e. $v_B = 0$. This however does not negate the need for an AIF as the tissue term in the model is made up of a convolution of the AIF and the residue function. For this study, an AIF from a subject with a glioma, which is derived from [71] is used.

We now have a form for the simulation but must decide whether to use the compartmental model or the nonparametric model to generate our simulated data. In the last section, we assumed that the widely used compartmental model was the appropriate model and assumed it to be the true model under the null hypothesis. This was appropriate as the compartmental model has been accepted in the literature as the way the modelling should be done and unless evidence against it is found, it should remain as the model of choice. In this case we will also use the compartmental model to simulate the data as we are interested in seeing how both models perform when the data is truly from the compartmental model

as again this can be thought of as the null hypothesis. We will also undertake a similar study where the data is simulated from the nonparametric model. This helps us gain a far greater understanding as to how each model works.

The simulated data is analysed where the data is 'perfect', i.e. where no noise is present. In this case $\tau = \infty$ but in practice the zero noise case is accomplished by letting $z_k = C_T(t_k|\theta)$. After this we introduce noise of varying levels by varying τ and examine how the bias and variance change as noise levels vary. Hence, perfect data is simulated as

$$C_T(t) = K_1 \int_0^t R(t-s)C_P(s)ds,$$

where $R(t)$ is the residue function. Removing blood volume reduces the number of parameters in the equation and makes it simpler to gain an understanding of how the models behave. Also as v_B is of the same form in both models i.e. the only difference between the models is in the residue function and obviously how the optimisation is conducted. We note however that this does not mean both models produce the same blood volume estimate on the same dataset as this is highly unlikely given the very different nature of the models and the likelihood of having random variability in the system.

The values for flow, k_2, k_3 and k_4 used to generate the data are 0.102, 0.013, 0.062 and 0.0068 respectively, which is roughly what would be seen in a grey matter region of a PET brain study in a normal subject [71]. In the case of the nonparametric study, an estimated grey matter residue function is used. This was generated from the glioma dataset. Various values for τ are considered in this work. The values utilised here contain the majority of noise levels seen in practice so as to give a clear picture of how the models behave in different scenarios. In this case a large number of simulations will be generated at each noise level.

3.3.3 Rates of Convergence

In this section we will discuss how to make use of the simulation study in such a way as to gain an insight into how both models are performing. We will begin by looking at perfect data. In this case we know the parameters of interest and the true residue function. We are not adding noise to the system. We can fit both

the compartmental model and the nonparametric model to the data. In doing so we will get estimates for flux, flow and volume from estimating the tissue residue function for both methods as well as the biases in measuring each parameter. Other measures of quantifying the bias can then be generated such as expressing the bias for parameters as a percentage of the true value.

We can also try to establish a measure of the error of the residue function. One such measure is the integrated squared error (ISE). This is defined as:

$$ISE = \int_0^T (R(t) - \hat{R}(t))^2.$$

Similarly ratios of the ISE expressed as a percentage of the area under the true residue but this is of little practical value as both residues are on the same scale.

In the case where there is no noise, if we repeat the same analysis again, we will get identical results. However, in the case where there is noise present in the data, the data changes every time simulations are conducted and so the results will not be identical. Accordingly at each noise level a lot of simulations are generated. This gives a distribution of the values at each noise level for flux, flow, v_D and the ISE. As these data are skewed we choose to use the median instead of the mean as our measure of centrality.

A useful way to estimate convergence would be to calculate the MSE for the parameters of interest at each noise level. In doing this a simple plot of dose against MSE can be generated to see if the parameters are converging to the true value such as those shown in Figure 3.9. It is also of interest to break the error into bias and variance.

The mean squared errors can be plotted at varying levels of noise for the compartmental and nonparametric method and this should show the rate at which the data is converging to the true value. In this case the rate of change of the error with respect to dose (τ) is assumed to be a constant. Hence the plot of mean squared error versus τ should be exponential with the exponent equal to the rate. Estimating the rate is readily accomplished by transforming the data to a log scale and fitting a straight line to the data using the model:

$$E[RMSE|\tau = t] = \beta_0 + \beta_1 t,$$

where the errors are assumed to be from a normal distribution with constant

variance. In estimating β_1 we will generate estimates for the rates at which the errors converge to the true values and also will have standard errors from the regression which can be used when doing inference.

3.4 Analysis

3.4.1 Data

As mentioned previously in Section 2.1, the basis of the work in the second half of this chapter is a simulation study. The method of simulation has also been described. A plot of some simulated time activity curves, generated from both the two compartmental model and nonparametric models, used in this work (without noise) are shown in Figure 3.1. These data were generated taking the values, which are typical values which would be observed in a grey matter region of a normal subject. The residue function chosen for the nonparametric method was one which was generated by fitting the nonparametric model to a grey matter ROI from a glioma study. Blood volume is assumed to be zero but an AIF is still

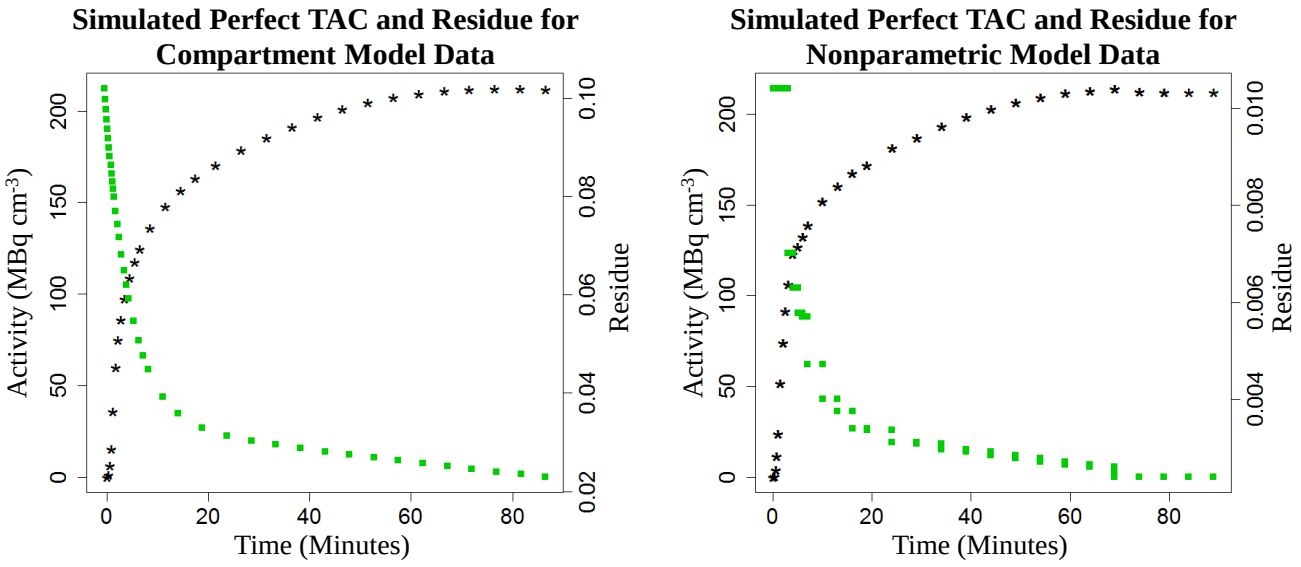


Figure 3.1: *Simulated time activity curves (TAC) for perfect compartmental and nonparametric data.*

required to do modelling. The AIF is also shown in Figure 3.2.

Previously, it was stated that one of the first things which will be examined

	Flow	Flux	v_D
True	0.102	0.023	2.88
2C Estimator	0.102	0.023	2.87
Standard Error	0.004	0.00045	0.04
Percentage Error	10^{-15}	10^{-5}	-0.001
True	0.1004	0.023	2.88
NP Estimator	0.1007	0.023	2.88
Standard Error	0.0308	0.003	0.04
Percentage Error	0.012	0.012	-0.0015

Table 3.1: *True values along with parameter estimates, standard and percentage errors for data simulated from the compartmental model.*

are the outputs from both models fitted to perfect data. Figure 3.3 shows the generated perfect data fitted by both the nonparametric and compartmental models along with the estimated residues from both methods for data generated.

After fitting the models to the data, it is of interest to compare the estimated parameters to the true parameters and attempt to understand if there is a bias in the model. In this case the parameters of interest are flux, flow and v_D, as well as the ISE in estimating the residue function.

A summary of the estimated parameters is presented in Tables 3.1 and 3.2.

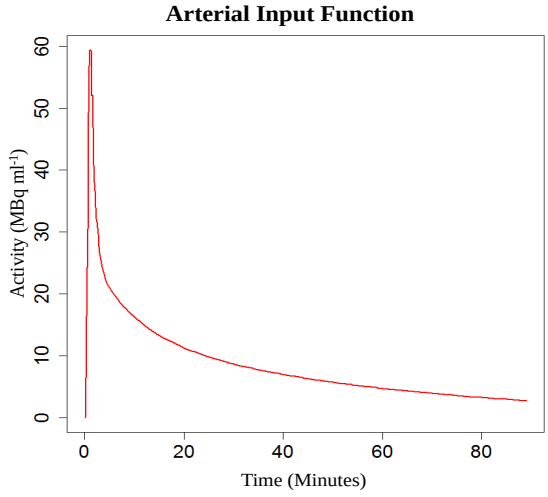


Figure 3.2: *AIF used to generate the simulated data.*

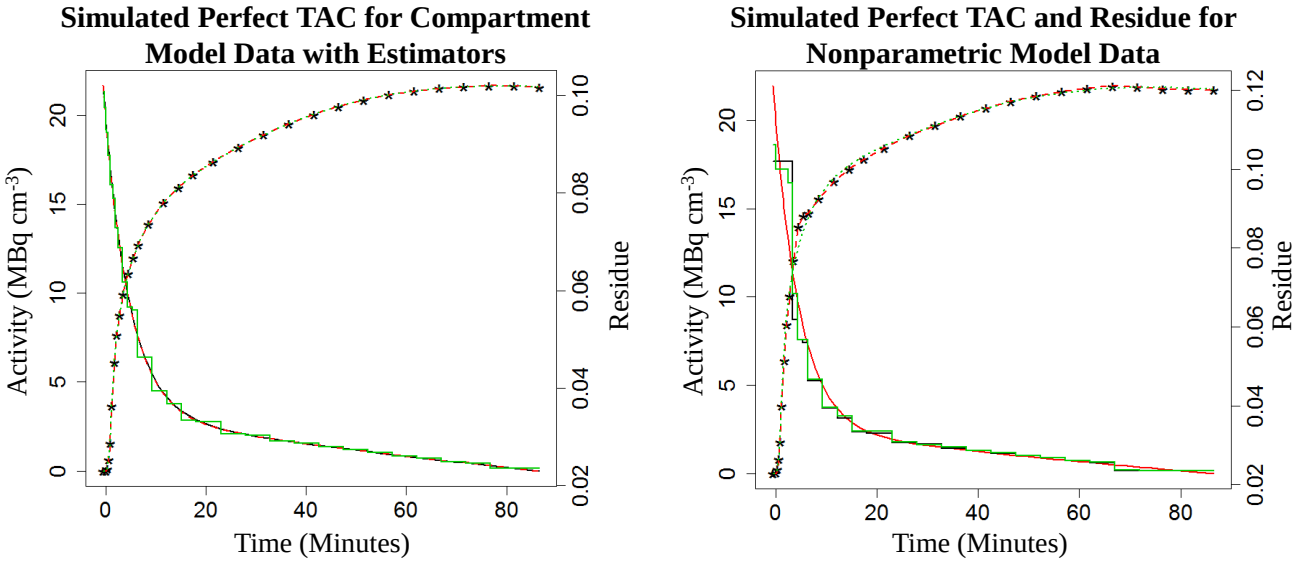


Figure 3.3: *Perfect data (∗) with compartmental (red) and nonparametric (green) estimators. Residue estimators are also shown.*

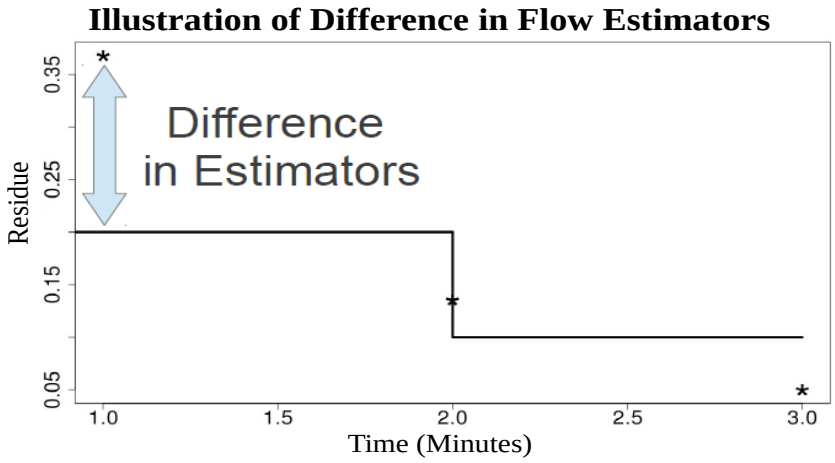


Figure 3.4: *Schematic highlighting how the models estimate the values within time bins in different ways.*

Data was generated from the compartmental model with known true values for flux, flow and v_D are 0.102, 0.023 and 2.88 respectively. As the nonparametric method estimates the value in the first time-bin by a single value, it may be appropriate to compare the nonparametric estimator to the average value of the true residue over the first time-bin. These differences between the estimators are highlighted in Figure 3.4. This leads to the different estimates being compared to different true values. In this case the true values the nonparametric estimator for flux, flow and v_D are 0.1004, 0.023 and 2.88 respectively In all cases when the

	Flow	Flux	VD
True	0.102	0.0235	2.949
2C Estimator	0.1021	0.0229	2.954
Standard Error	0.23	0.016	0.24
Percentage Error	0.186	-2.55	0.0017
True	0.102	0.0235	2.949
NP Estimator	0.1026	0.0235	2.949
Standard Error	0.0001	0.00002	0.006
Percentage Error	0.05	10^{-16}	10^{-16}

Table 3.2: *True values along with parameter estimates, standard and percentage errors for data simulated using nonparametric model.*

simulated data is generated from the compartment model the percentage error is smaller for the compartment model for flux, flow and v_D than with the nonparametric estimator. When simulating on the compartment model the percentage errors for the compartmental estimator are of the order of 10^{-15}, 10^{-5} and -0.001 percent for flux, flow and v_D respectively. For the compartmental model data the percentage errors for the nonparametric estimator are 0.012, 0.012 and -0.0015 percent for flux, flow and v_D respectively.

In the case of the nonparametric data the known true values for flux, flow and v_D of 0.102, 0.0235 and 2.949 respectively. Similar to the compartmental data, the simulated data generated from the nonparametric model has a smaller percentage error for the nonparametric estimators for flux, flow and v_D than with the compartmental model estimator. When simulating on the nonparametric data the percentage errors for the compartmental estimator are 0.186, -0.01255 and 0.0017 percent for flux, flow and v_D respectively. For the nonparametric model data the percentage errors for the nonparametric estimator are 0.05, 10^{-16} and 10^{-16} percent for flux, flow and v_D respectively. It is also worth highlighting that these percentage errors can be considered as percentages representing biases as the data is perfect data with no error term.

Also of interest are the errors in the residue functions for both methods.

In the case of the compartmental model data, the ISE for the nonparametric model estimator is 7.04×10^{-7} for the compartmental model estimator and 1.5×10^{-5} for the nonparametric estimator. In the case of the data simulated on the nonparametric model then the values for the ISE are 1.5×10^{-21} for the nonparametric estimator and 0.0001 for the compartmental model estimator. In both cases the residue is estimated more accurately by the nonparametric model. This is likely due to the extra flexibility in the nonparametric estimator.

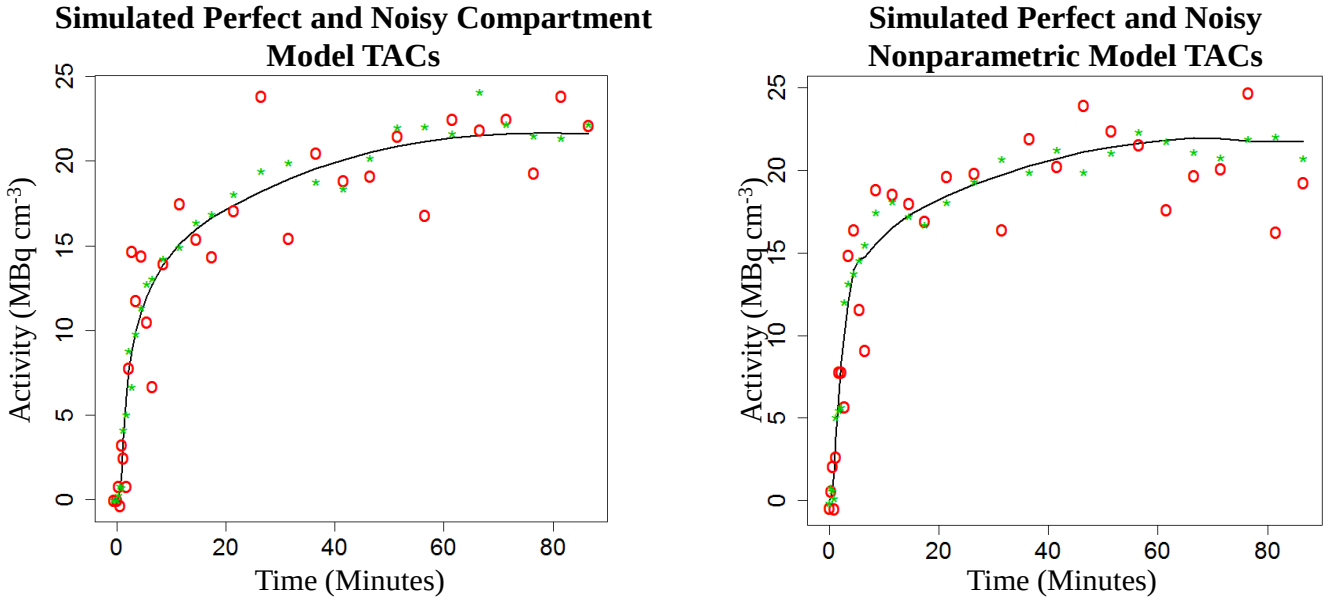


Figure 3.5: *Simulated time activity curves from the compartmental model (left) and from the nonparametric model (right) with high, low and no noise present*

We now consider the case where that data is imperfect and consider simulations with varying degrees of noise. Figure 3.5 shows data generated from both the compartmental and nonparametric models with different levels of noise that will be used in the simulation studies for both datasets. In analysing v_D for example, it was necessary to generate a large number of noise levels to visualise the trends in the data. This means that there is data with quite a lot of noise and quite a low level of noise and our goal here is to see if the estimates get more accurate as noise levels decrease. It was decided for consistancy to use the same large collection of noise levels for flux and flow as well. However taking smaller ranges of noise has little practical significance on the results of the analysis on

flux and flow.

At each noise level 250 simulated time-courses were generated. Each of these were analysed with both models and estimates generated for the parameters of interest. The bias and variance are of particular interest in this context. Box plots of the percentage errors for for flow at varying noise levels for both the compartmental and nonparametric models are shown in Figure 3.6. Similar illustrations are presented for flux in Figure 3.7 and v_D in Figure 3.8.

We now describe each of these figures individually. In the case of flow, there is a larger bias for the nonparametric model on compartmental model data and for the compartmental model on nonparametric data. This is not surprising as one would expect each model to fit its own data better. There is also more variability present in the nonparametric estimates regardless of whether the data are nonparametric or compartmental.

For flux, one can see clearly that there is a larger variance present in the nonparametric estimator regardless of which model the data are generated from. Again this is due to the nature of the nonparametric estimator. At higher noise the nonparametric model has a larger bias than the compartment model for both types of data, however as the level of noise decreases the nonparametric model tends towards having smaller bias in estimating flux for nonparametric data and only slightly more bias than the compartmental model for compartmental data. On nonparametric data the bias of the nonparametric estimator is very low at low noise while the compartment model slightly under estimates the true value.

For volume of distribution, there is more bias in the nonparametric method at very high noise levels but aside from this there is little to distinguish between the two estimators as noise levels decrease for compartmental data. For the nonparametric model the same is true for very high noise but as the noise levels decrease, there is less noise present for the nonparametric model. Both models seem to have a similar level of variability throughout the varying levels of noise if the extremely high noise regions are not considered.

While this interpretation of box plots allows us to qualify the bias and variance in the system, it is perhaps more interesting to study the effect of noise on bias more rigorously. Looking at the rates the different models are converging to the true known values can be done using the root mean squared error (RMSE) as

was discussed earlier in this chapter.

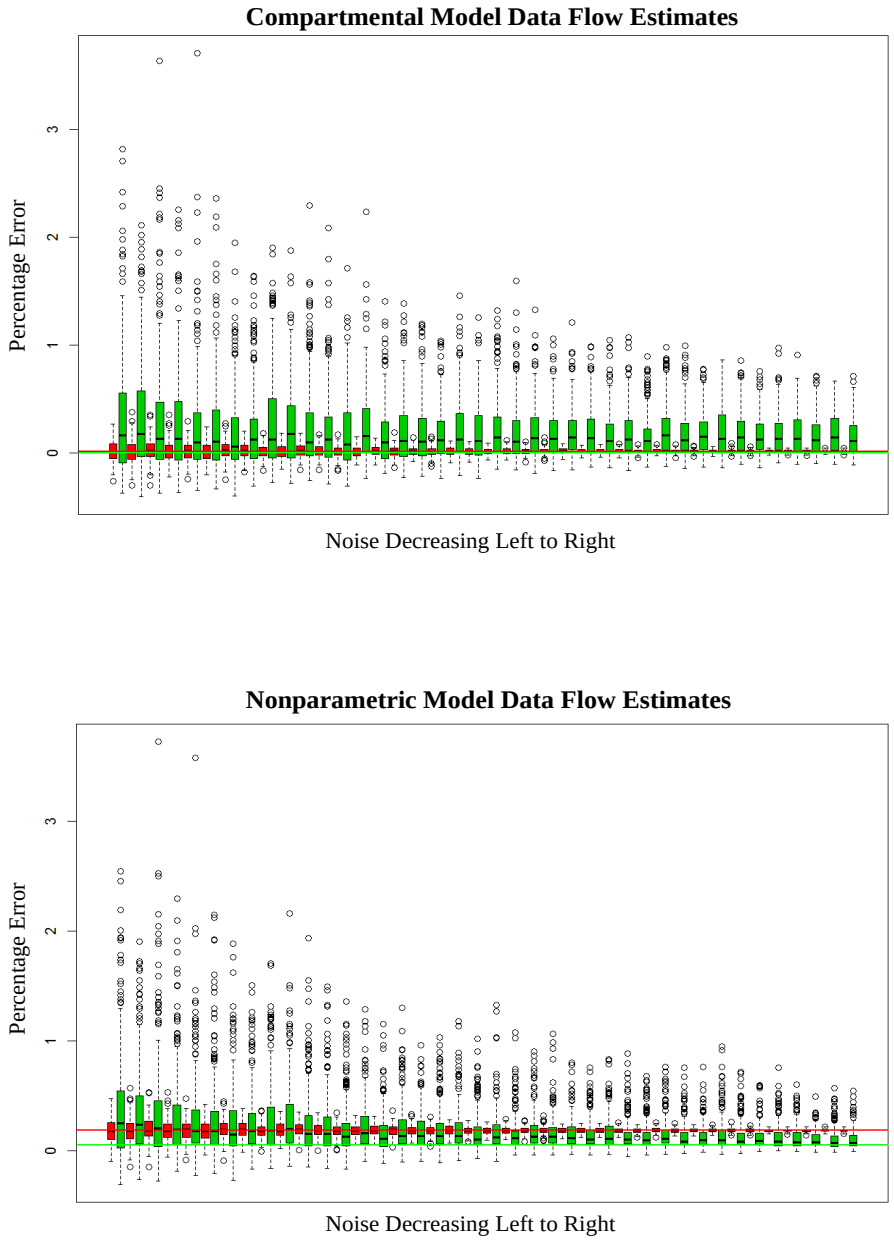


Figure 3.6: *Flow percentage errors for the nonparametric model (green) and the compartmental model (red) on compartmental model data (top) and nonparametric data (bottom). Also shown is the error line on perfect data.*

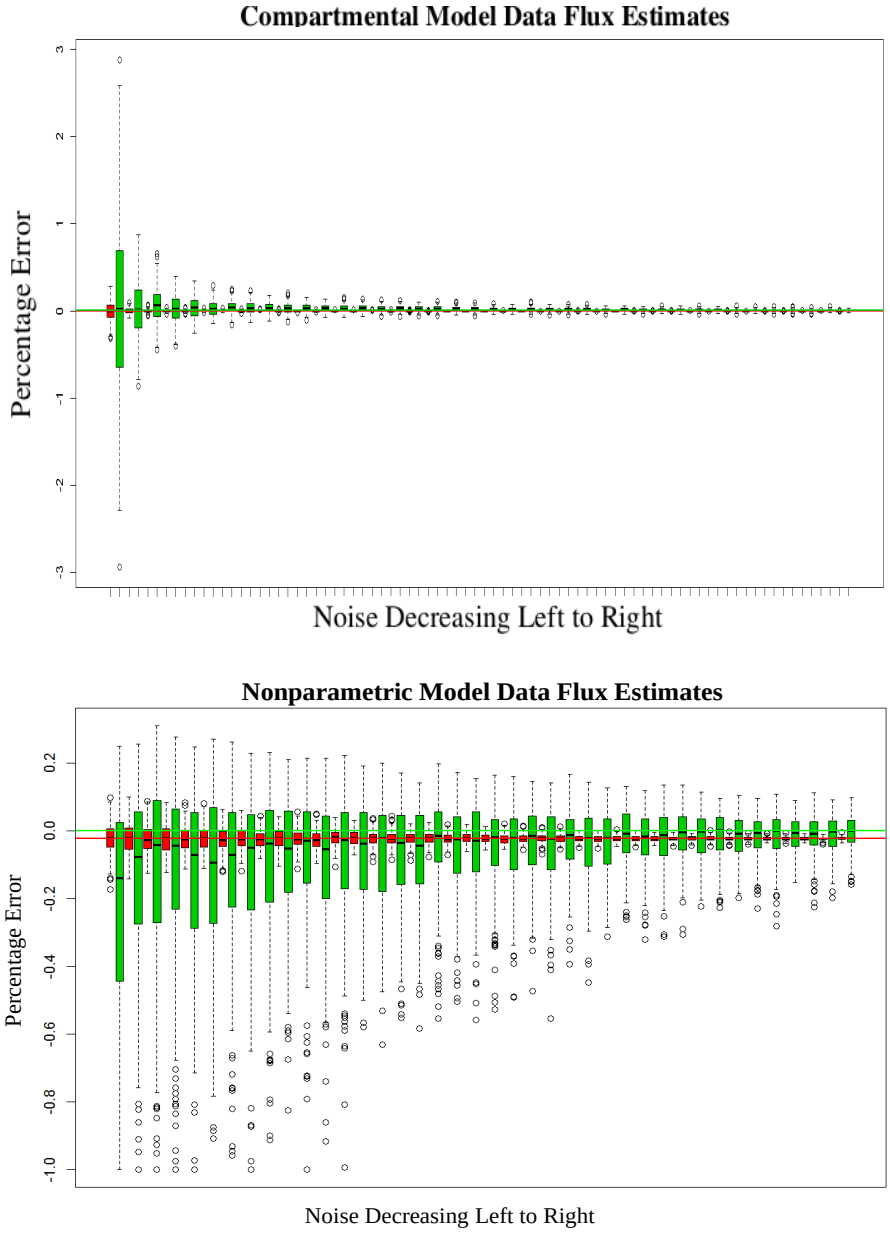


Figure 3.7: *Flux percentage errors for the nonparametric model (green) and the compartmental model (red) on compartmental model data (top) and nonparametric data (bottom). Also shown is the error line on perfect data.*

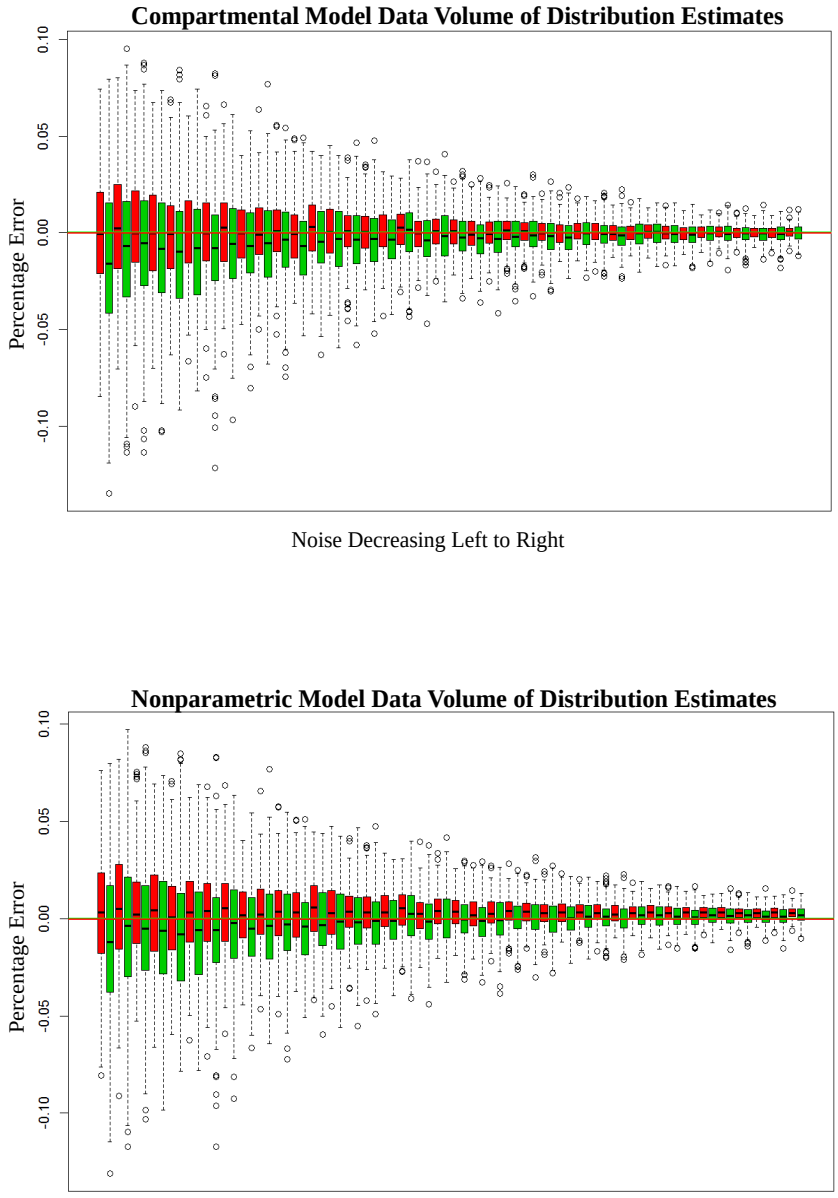


Figure 3.8: *v_D percentage errors for the nonparametric model (green) and the compartmental model (red) on compartmental model data (top) and nonparametric data (bottom). Also shown is the error line on perfect data.*

Compartmental Data						
	Flow		**Flux**		$\boldsymbol{v_D}$	
	2C (SE)	NP (SE)	2C (SE)	NP (SE)	2C (SE)	NP (SE)
Ordinary Regression	-0.96 (0.02)	-0.28 (.03)	-1.01 (0.02)	-0.61 (.02)	-1.00 (0.02)	-0.96 (.01)
Robust Regression	-0.96 (0.01)	-0.26 (.03)	-1.02 (0.02)	-0.61 (.02)	-1.00 (0.01)	-0.96 (.02)
Nonparametric Data						
	Flow		**Flux**		$\boldsymbol{v_D}$	
	2C (SE)	NP (SE)	2C (SE)	NP (SE)	2C (SE)	NP (SE)
Ordinary Regression	-0.001 (0.004)	-0.47 (.02)	-0.07 (0.01)	-0.75 (.02)	-0.86 (0.02)	-0.97 (.02)
Robust Regression	-0.0004 (0.004)	-0.47 (.01)	-0.07 (0.02)	-0.76 (.01)	-0.87 (0.02)	-0.97 (.02)

Table 3.3: *Robust estimates of convergence rates for Flux Flow and V_D for nonparametric method and compartmental model along with ordinary least squares estimates for data generated using the compartmental model (top) and nonparametric model (bottom).*

The RMSE at each noise level is plotted against dose (noise) and this is shown in Figure 3.9 for the three parameters of interest. As we are interested in the rate of convergence, the figures are shown on a log scale. The slope of the lines should give the rate of convergence for the parameters of interest for both nonparametric and compartmental models. The various slopes (standard errors) are given in Table 3.3. Alternatively one could consider a robust linear regression to make sure that any outliers are not skewing the line fitted to the data and this is also shown in the table. In these data little practical difference is found.

In the case of flow the compartmental model converges faster for compartmental data and the nonparametric model converges faster for nonparametric data. In fact the compartmental model does not converge on the nonparametric data. For flux the rates of convergence for the compartment model on compartmental model data are approximately 1, while the nonparametric model converges at a rate of 0.6. In the case of truly nonparametric data these values are 0.1 and 0.7 for the compartmental and nonparametric models respectively. This again shows the compartmental model does not converge to the estimate on perfect nonparametric data. There is very little difference between the models for v_D where regardless of the type of data or the model being fitted, convergence of between 0.8 and 1 is obtained with both models performing slightly better on data generated from that model.

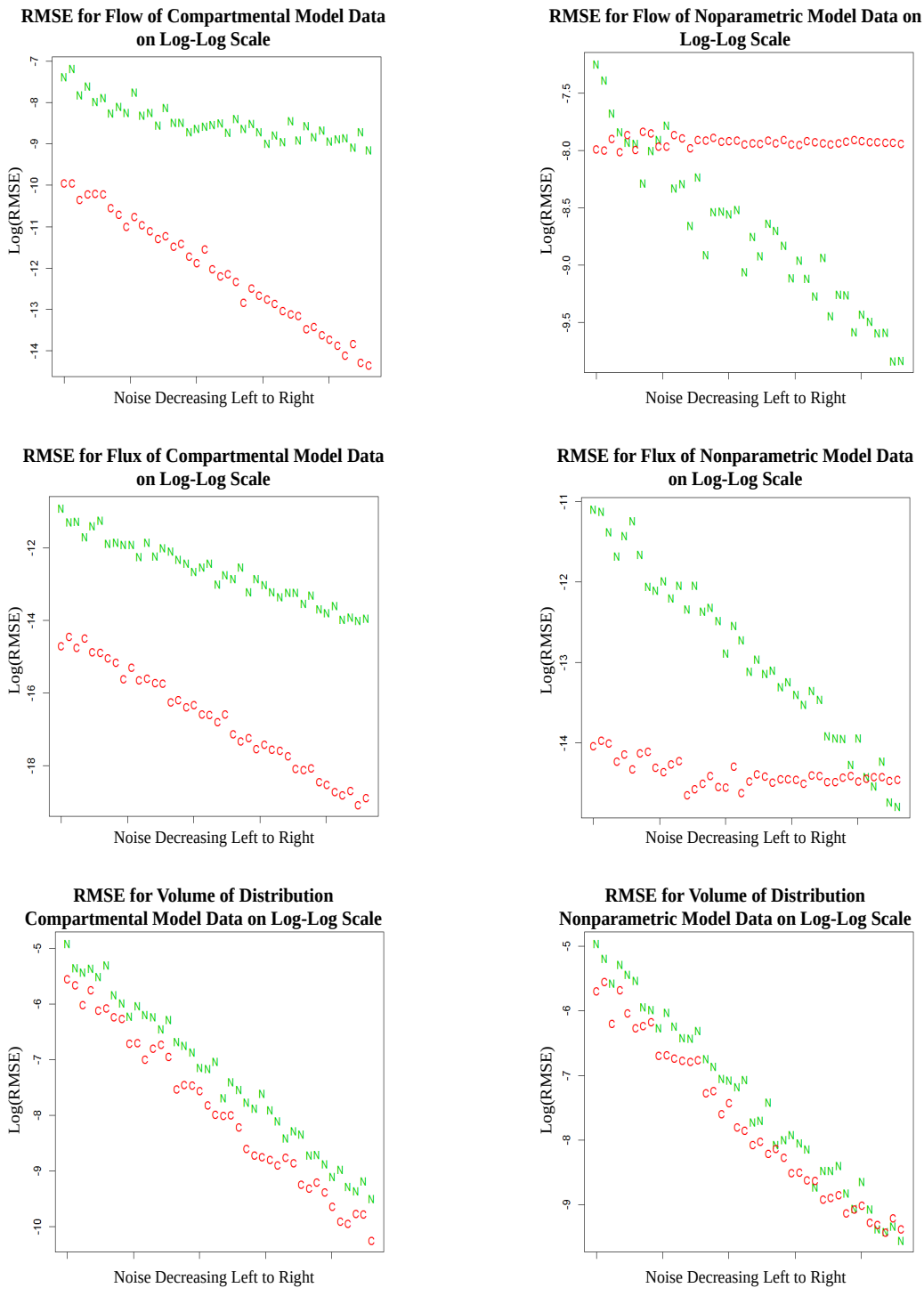


Figure 3.9: *RMSE against dose on a log-log scale using data generated from both models with the nonparametric estimator represented by green 'N' and compartmental estimator represented by a red C with Dose (size of region) on the x-axis.*

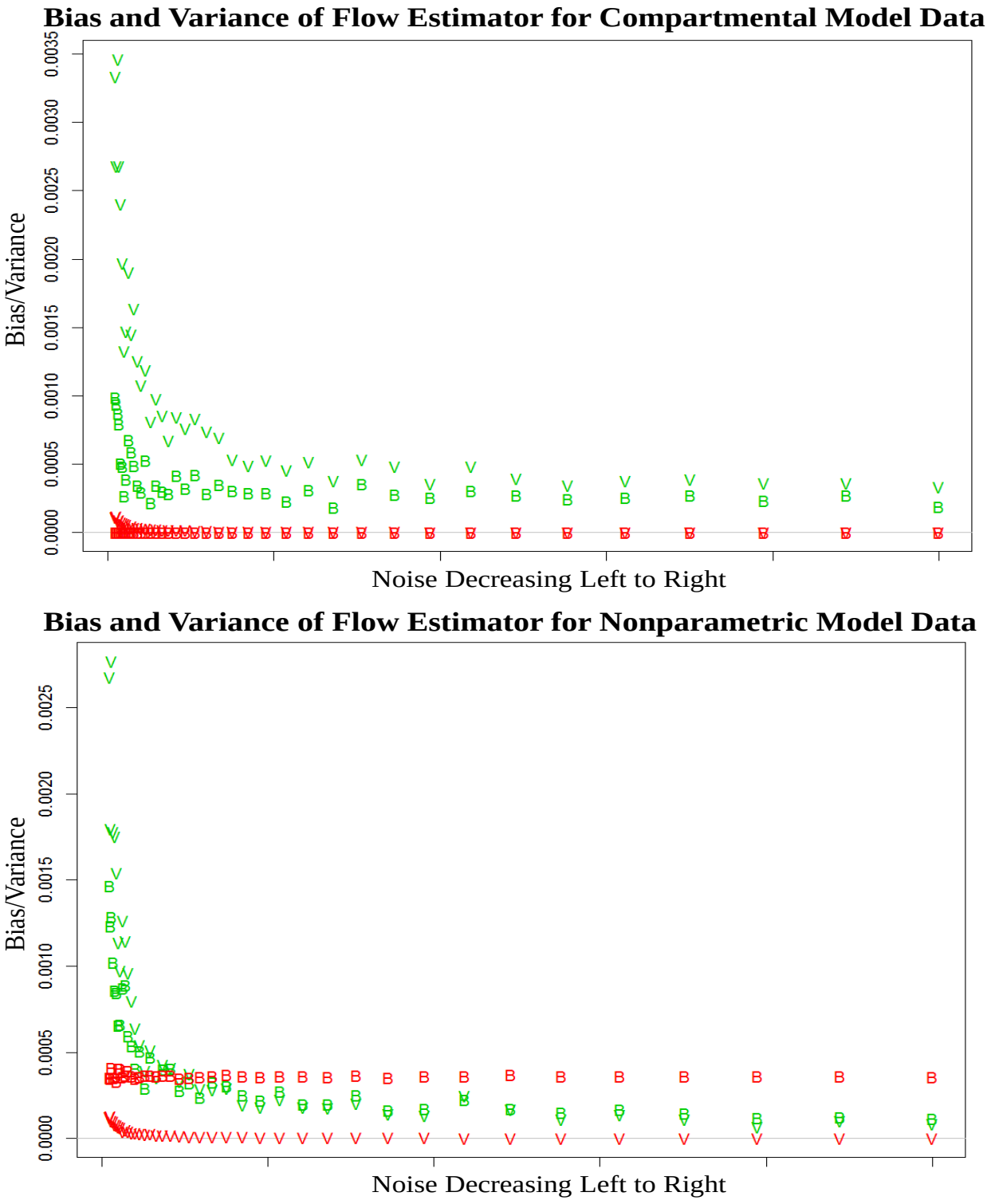


Figure 3.10: *Bias Squared (B) and variance (V) plots for flow for the compartmental model (red) and the nonparametric (green)*

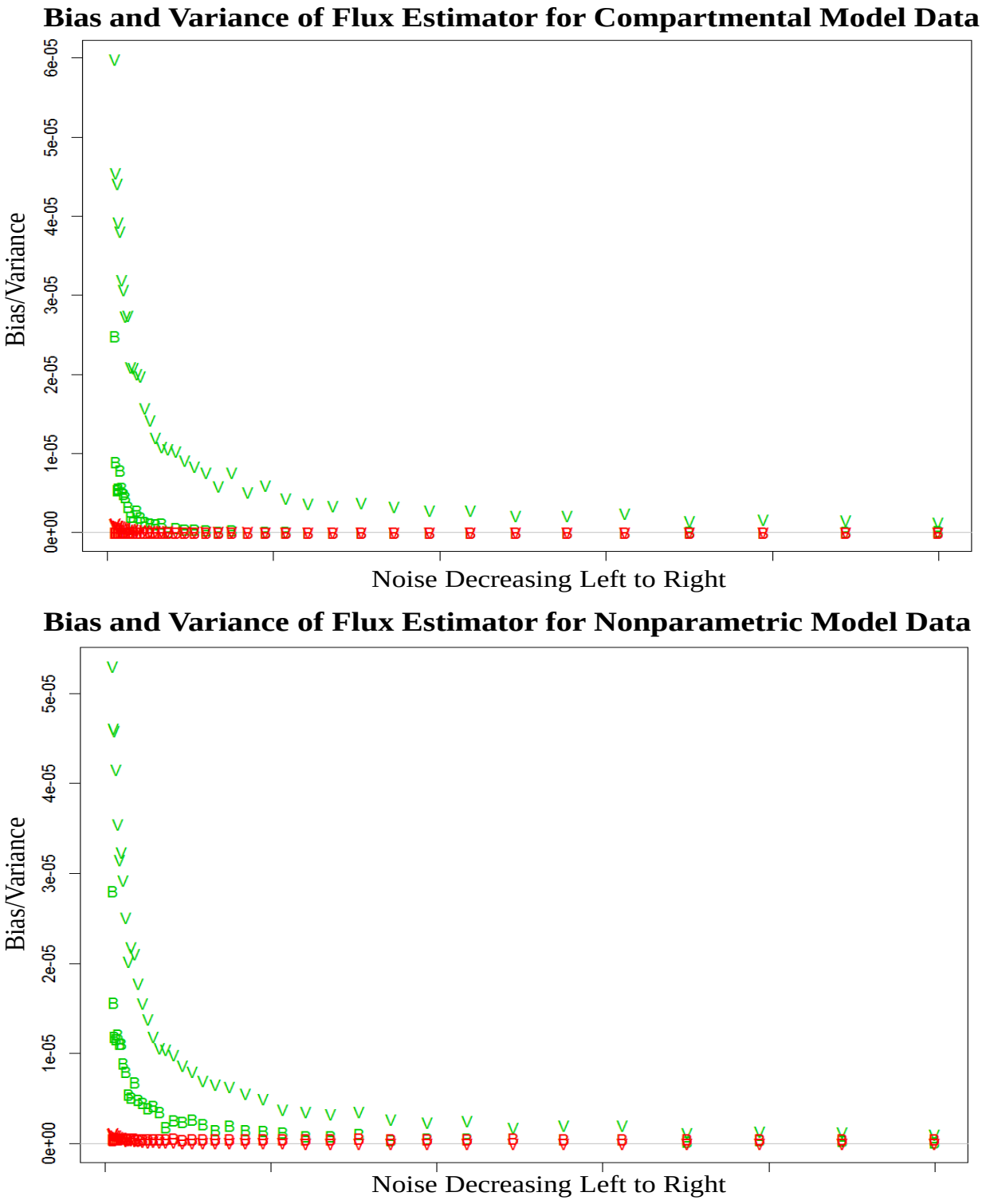


Figure 3.11: *Bias Squared (B) and variance (V) plots for flux for the compartmental model (red) and the nonparametric (green)*

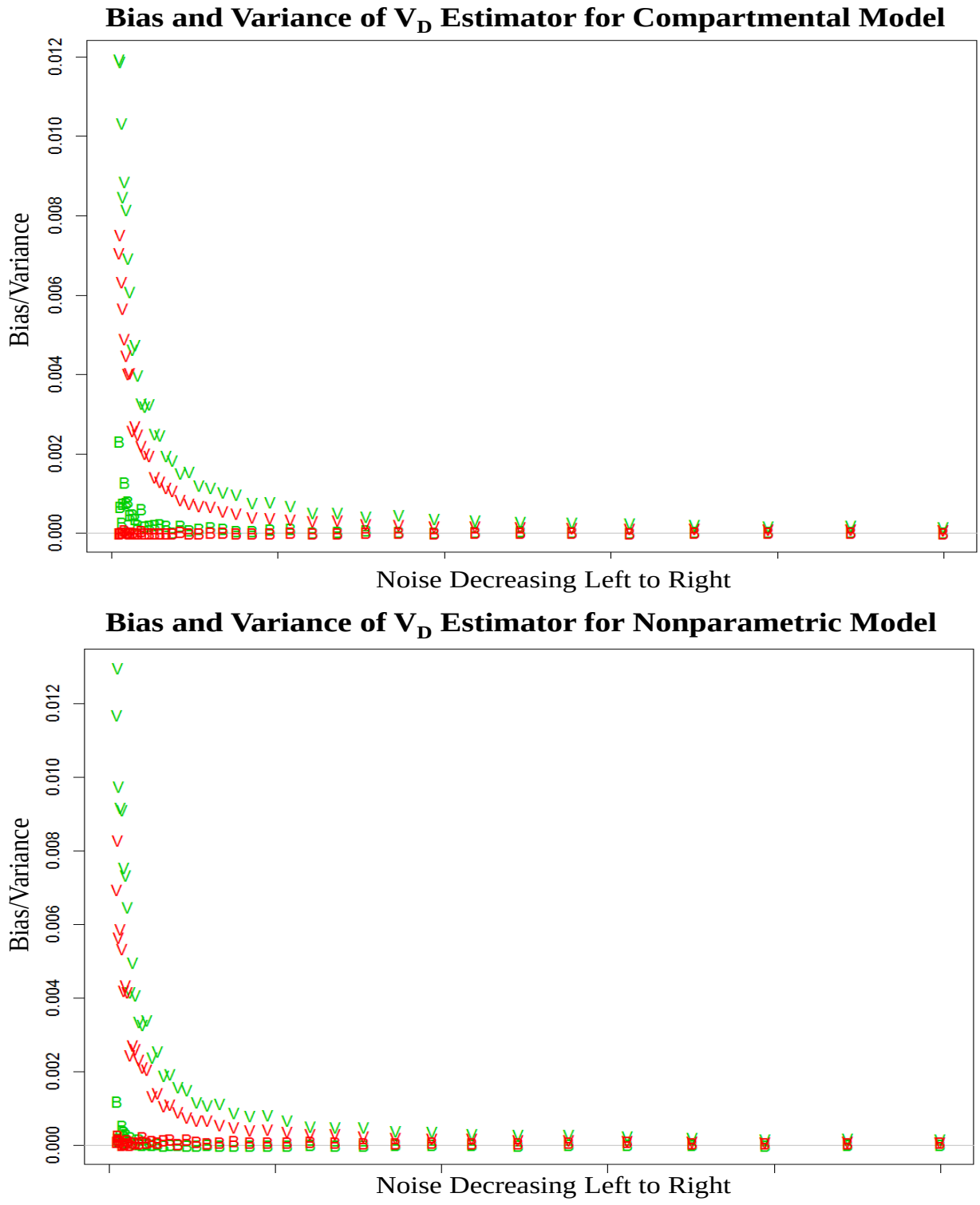
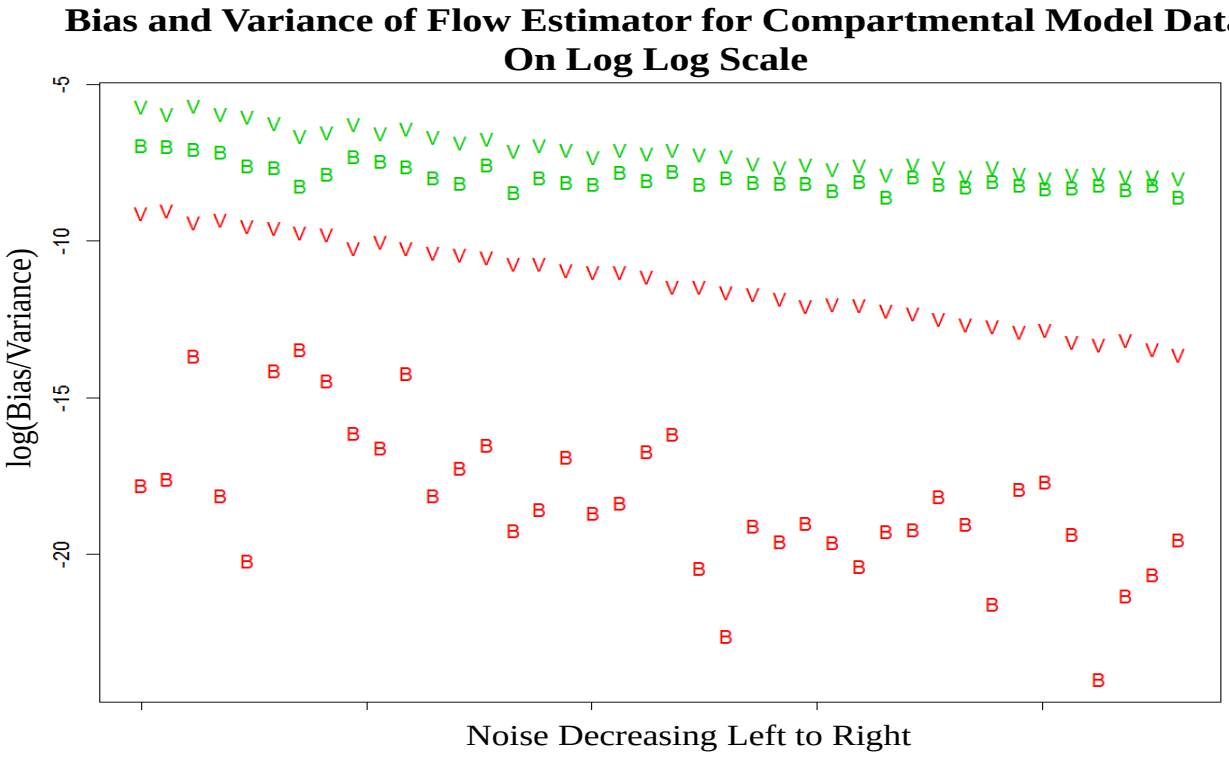


Figure 3.12: *Bias Squared (B) and variance (V) plots for v_D for the compartmental model (red) and the nonparametric (green)*

Aside from the rates of convergence, there are many other attributes of each estimator which can be of interest. One of which is what proportion of the MSE which is bias and what proportion is variance. Bias and variance plots for the three parameters of interest are presented in Figures 3.10, 3.11 and 3.12. While the form of the bias as a function of noise is somewhat erratic, the variance seems to follow an exponential type curve. Transforming this to a log scale is shown in Figures 3.13, 3.14 and 3.15. From this it should be possible to determine the rate at which the variability is decreasing as dose increases. Although a similar analysis is not likely to produce as meaningful a result as in the case of the variance, the rate at which the bias is decreasing will also be examined. The slopes of the lines fitted to these plots are presented in Table 3.4



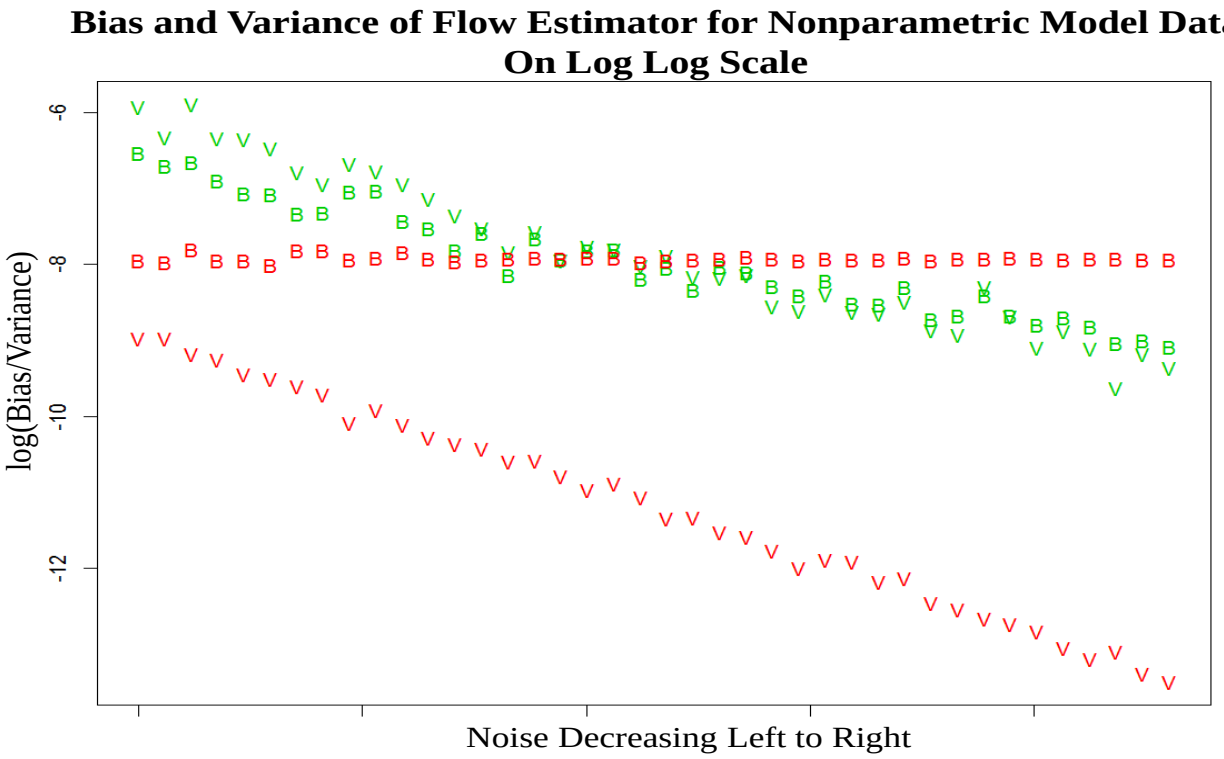
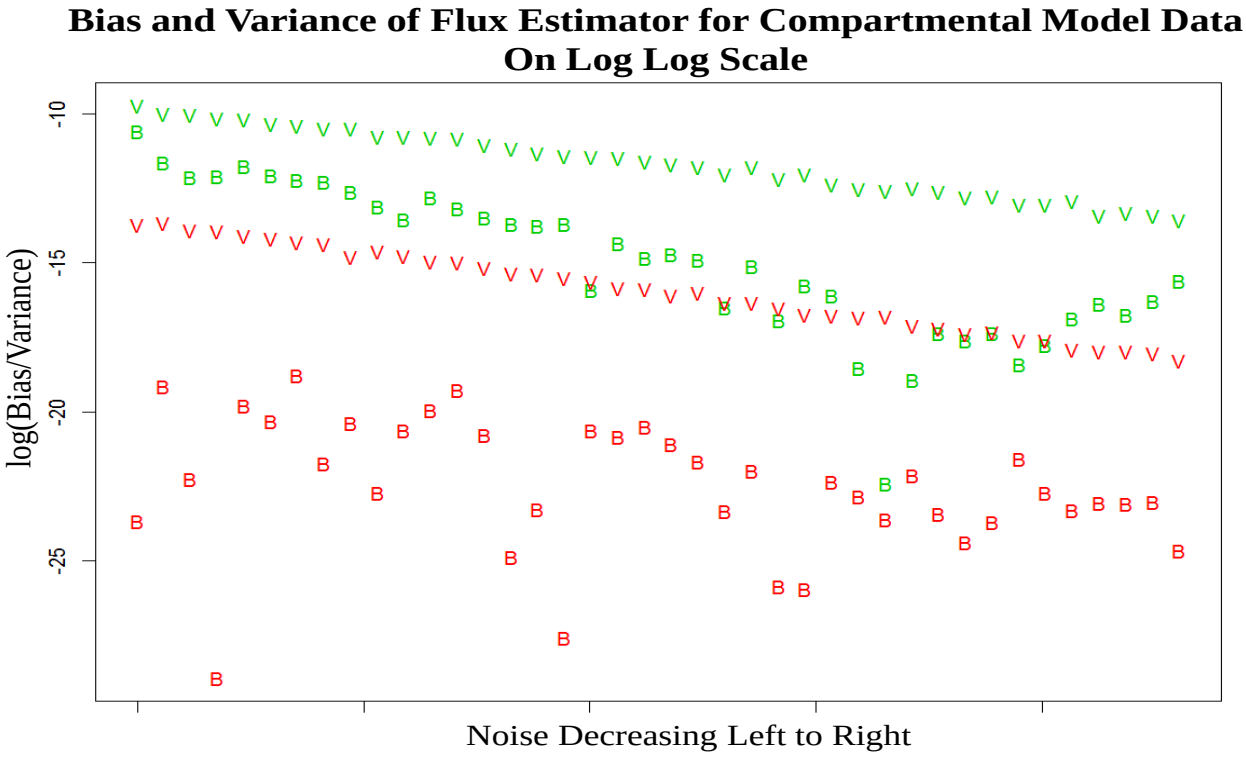


Figure 3.13: *Bias Squared and Variance plots for flow on a log scale for the nonparametric (green) and the compartmental model (red).*



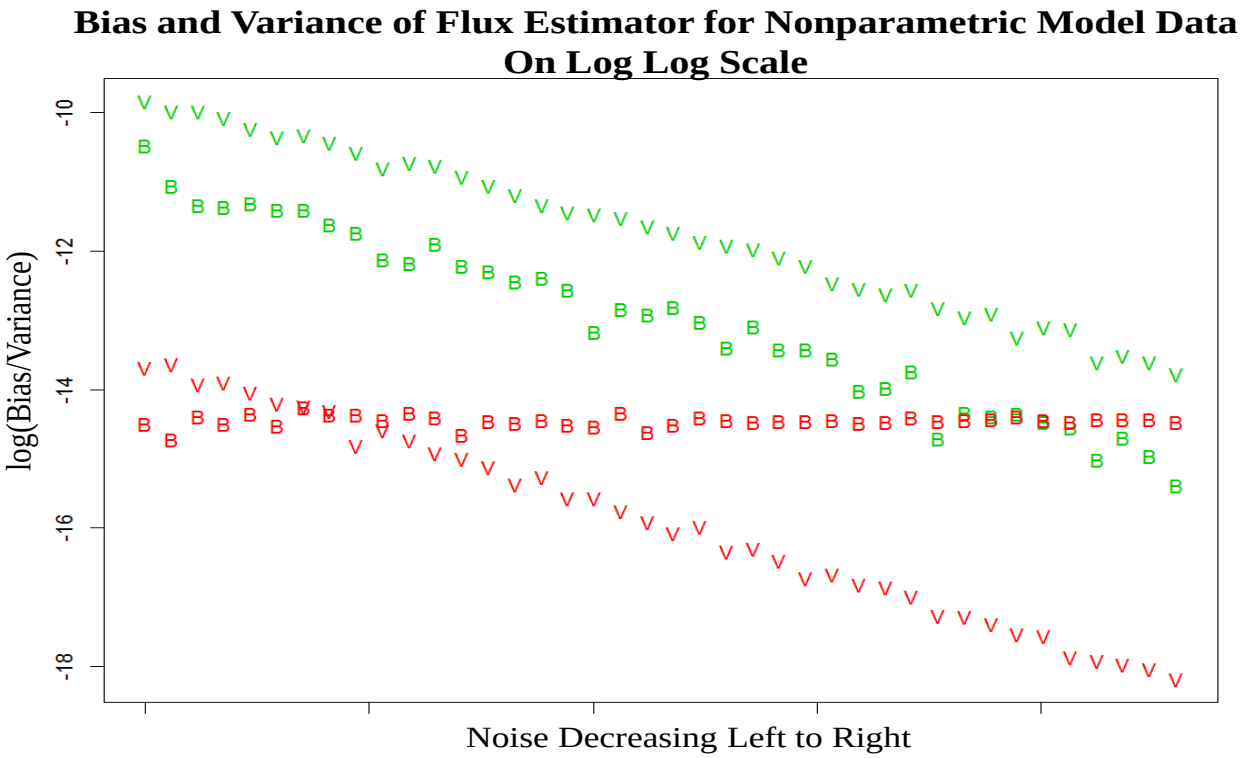
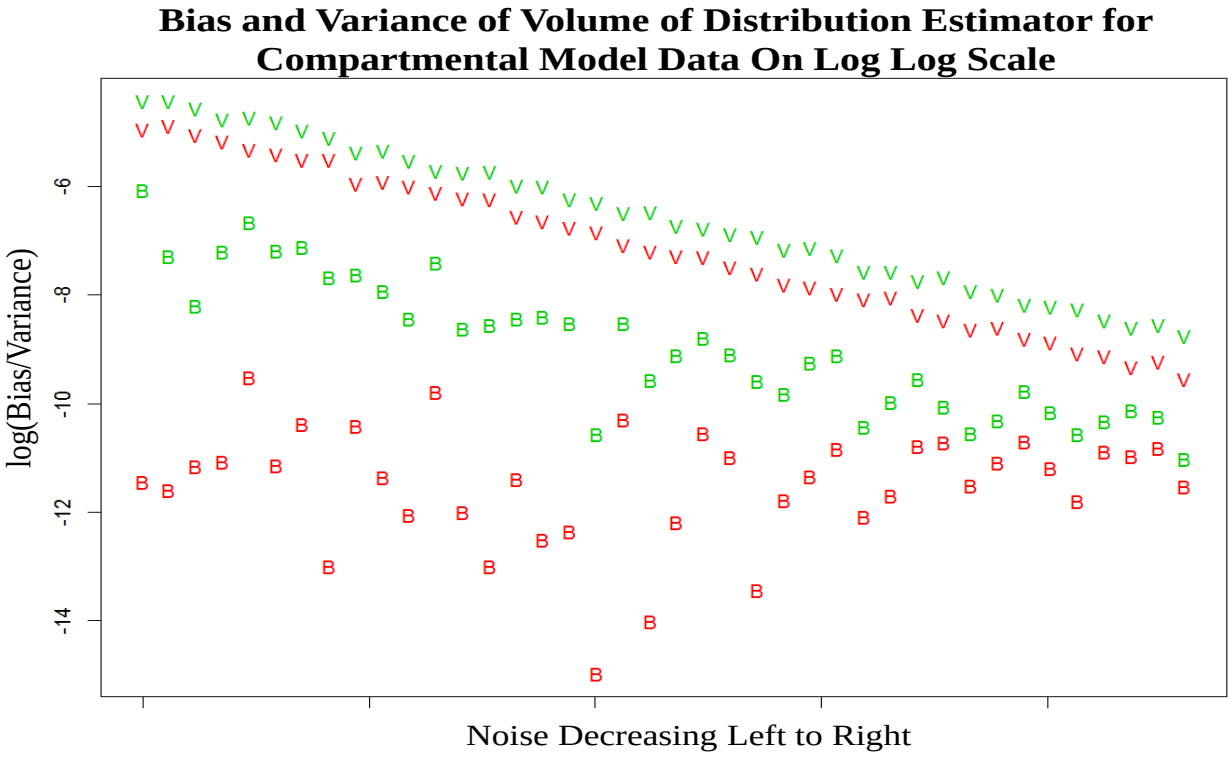


Figure 3.14: *Bias Squared and Variance plots for flux on a log scale for the nonparametric (green) and the compartmental model (red).*



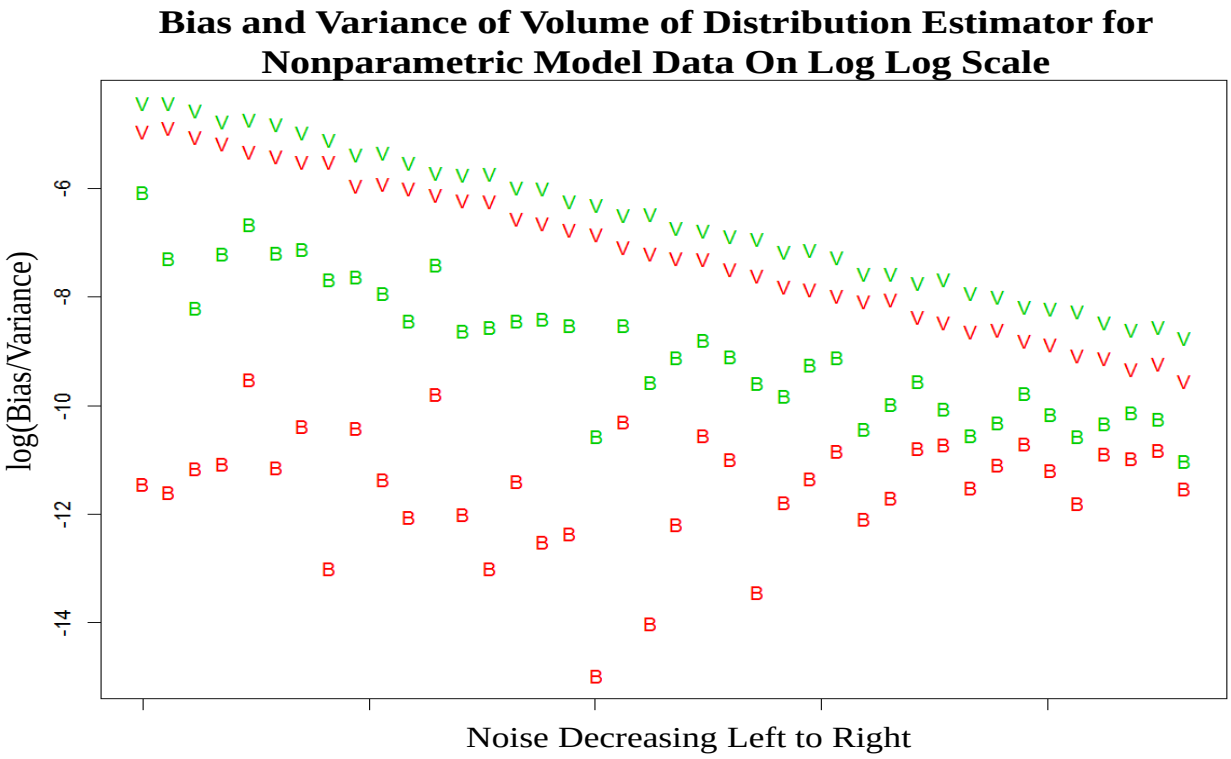


Figure 3.15: *Bias Squared and Variance plots for v_D on a log scale for the nonparametric (green) and the compartmental model (red).*

Compartmental Data Bias	**Flow**		**Flux**		v_D	
	2C (SE)	NP (SE)	2C (SE)	NP (SE)	2C (SE)	NP (SE)
Ordinary Regression	-1.08 (0.2)	-0.23 (0.03)	-0.48 (0.2)	-1.5 (0.15)	-0.01 (0.1)	-0.84 (0.06)
Robust Regression	-1.06 (0.1)	-0.24 (0.03)	-0.67 (0.19)	-1.56 (0.08)	-0.03 (0.1)	-0.85 (0.05)
Nonparametric Data Bias	2C (SE)	NP (SE)	2C (SE)	NP (SE)	2C (SE)	NP (SE)
Ordinary Regression	-0.005 (0.004)	-0.50 (0.02)	-0.004 (0.009)	-0.92 (0.02)	-0.04 (0.05)	-1.04 (0.2)
Robust Regression	-0.001 (0.002)	-0.50 (0.02)	-0.004 (0.009)	-0.91 (0.02)	-0.07 (0.05)	-0.99 (0.2)
Compartmental Data Variance	2C (SE)	NP (SE)	2C (SE)	NP (SE)	2C (SE)	NP (SE)
Ordinary Regression	-0.98 (0.01)	-0.49 (0.03)	-1.007 (0.009)	-0.80 (0.01)	-1.02 (0.008)	-0.96 (0.008)
Robust Regression	-0.98 (0.01)	-0.48 (0.02)	-1.009 (0.009)	-0.80 (0.01)	-1.02 (0.006)	-0.96 (0.008)
Nonparametric Data Variance	2C (SE)	NP (SE)	2C (SE)	NP (SE)	2C (SE)	NP (SE)
Ordinary Regression	-0.97 (0.01)	-0.71 (0.03)	-1.01 (0.009)	-0.85 (0.01)	-1.01 (0.01)	-0.97 (0.01)
Robust Regression	-0.98 (0.01)	-0.72 (0.03)	-1.01 (0.009)	-0.84 (0.01)	-1.01 (0.01)	-0.97 (0.01)

Table 3.4: *Estimates of convergence rates for Flux Flow and V_D for nonparametric method and compartmental model generated using robust and ordinary least squares data simulated from both models.*

As can be seen in Table 3.4, all estimates for variance are negative, which

means that the bias and variance are decreasing for all parameters as noise decreases. This is also the case for some estimates however convergence is close to zero indicating that the bias decreasing at a very slow rate.

When the data generated is from the compartmental model, the bias decreases at a faster rate for compartmental model estimator for flow, however the nonparametric estimator for flux and v_D has the bias decreasing faster. The variance is decreasing faster for all three compartmental estimators for compartmental data.

When the data are generated from the nonparametric model, all three parameters have bias decreasing at a faster rate for the nonparametric model. In the case of the nonparametric data, the compartmental model variance converges much faster for all three parameters, however as the compartmental estimators are very weak and do not come close to convergence in the case of bias, this is not surprising.

3.4.3 Convergence of the Residue Function

Similar to the analysis just presented for the parameters of interest, it is also of interest to see the way the residue function converges to the true residue. A box plot of the integrated squared errors similar to that for the kinetic parameters which have just been discussed is presented in Figure 3.16. This plot is then shown on a log log scale in Figure 3.17. This is very similar to the previous section and again these plots contain the details of when the data was generated from the compartmental model and the nonparametric model.

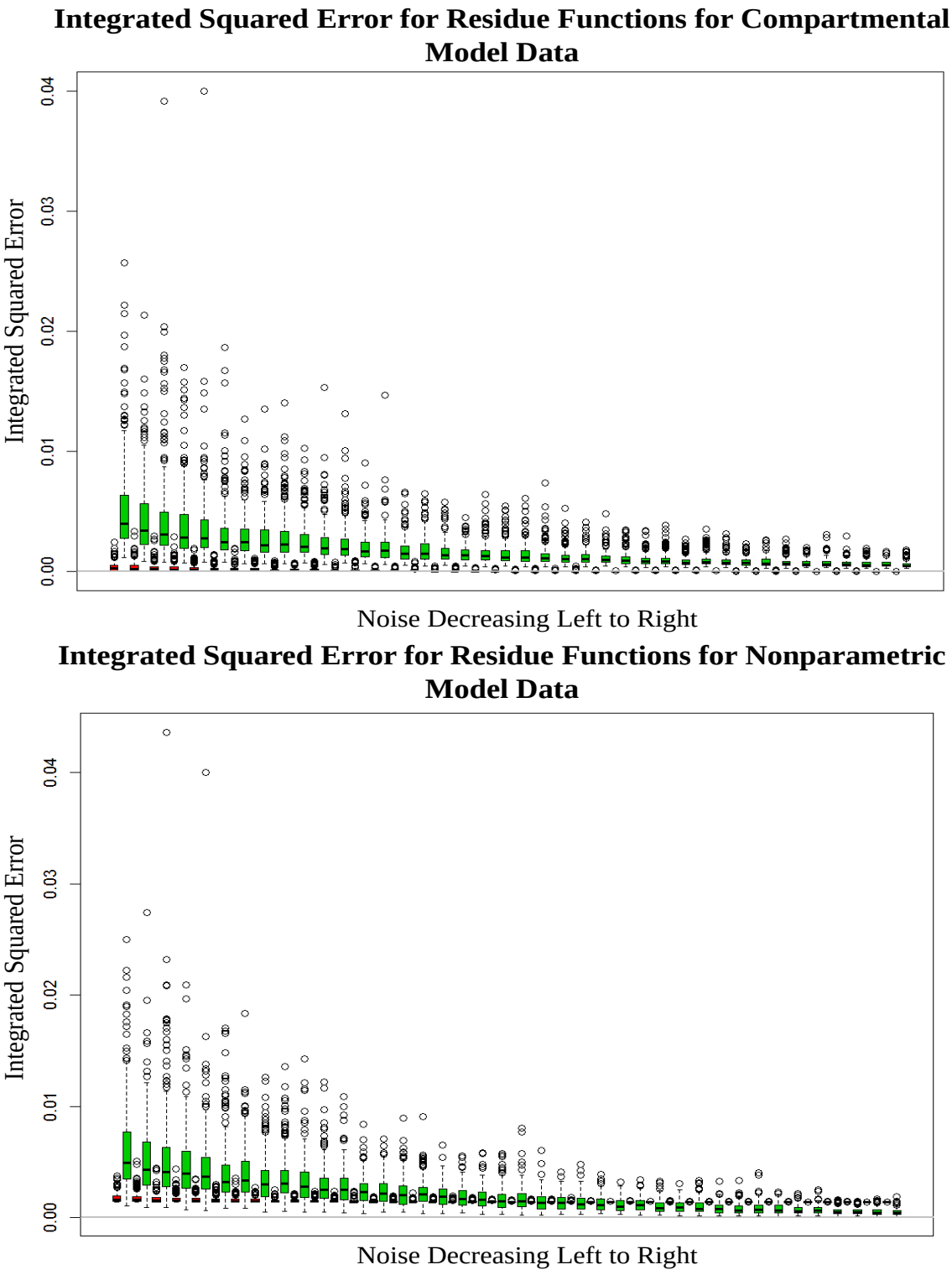


Figure 3.16: *Integrated Squared Errors for residue function with compartmental model (red) and nonparametric model (green).*

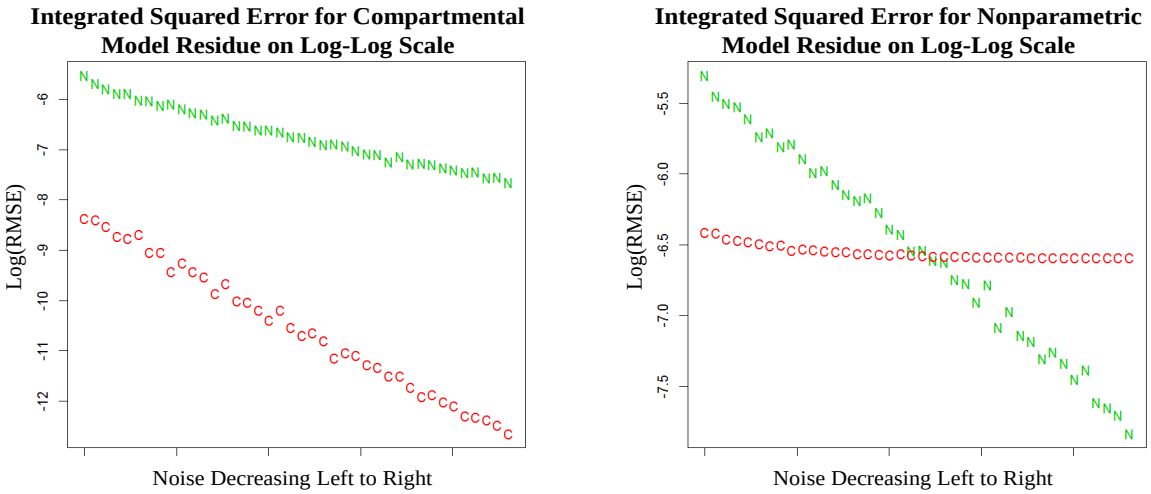


Figure 3.17: *Median Integrated Squared Errors for residue function for compartmental model (red) and nonparametric model (green) for varying noise levels on a log scale.*

Compartmental Model Data	Ordinary Regression	Robust Regression
Compartmental Model	**-0.94**	**-0.94**
Standard Error	0.01	0.01
Nonparametric Model	**-0.429**	**-0.41**
Standard Error	0.006	0.006
Nonparametric Model Data		
Compartmental Model	**-0.03**	**-0.02**
Standard Error	0.0021	0.002
Nonparametric Model	**-0.52**	**-0.52**
Standard Error	0.006	0.006

Table 3.5: *True values along with parameter estimates, standard and percentage errors for data simulated using nonparametric model.*

Similar to the parameters, it is possible to quantify the rate at which the residue is converging to the true residue. If we take the median of the ISE for the parametric and nonparametric methods at each level of noise, we can evaluate whether the models get closer to the true residue as dose increases and if so at what rate. This is shown in Figure 3.17.

The rates at which the convergence occurs is presented in Table 3.5. Perhaps as one would expect the compartmental model does best in the case where the data are compartmental model data and the nonparametric does better on nonparametric data. However the compartmental model is almost completely failing here on the nonparametric data and as it has performed poorly in the analysis of bias and variance of parameter estimates, then it raises questions about its use in practice.

3.4.4 Extra Sums of Squares Statistic

The extra sums of squares statistic or the F-statistic familiar from ANOVA, is used when comparing models, which is made up of residual sums of squares(RSS). Examining this statistic is the next step in comparing the compartmental and nonparametric models. Although, we focus on the ESS statistic, we will begin this section by looking at residuals for both methods and their relationship with a χ^2 distribution. After this we will look at the distribution of the ESS. A plot of the weighted residuals for a simulated region with high noise is shown in Figure 3.19. Box plots of the RSS for both the compartmental and nonparametric models are shown in Figure 3.20 for a large collection of simulated datasets. Here one can see that the compartmental model is behaving as one would expect, but the nonparametric model is improving as noise decreases. This is likely due to the flexibility of the nonparametric estimator.

As the distribution of the RSS scaled by the variance is χ^2 in the case of a linear model, one would expect that this result will hold approximately in the nonlinear case i.e. the compartmental model. The box plots of the weighted residuals seem to suggest that the distribution of the compartmental models is similar at all noise levels and that the nonparametric model has a smaller WRSS for data with less noise. We begin by applying the χ^2 distribution to these data at each individual level of noise and estimate the degrees of freedom of the distribution. An illustration of a χ^2 distribution fitted to one of these datasets for both the nonparametric and compartmental models is shown in Figure 3.21. In this plot, the fits to the data are generated by using maximum likelihood estimators for the degrees of freedom.

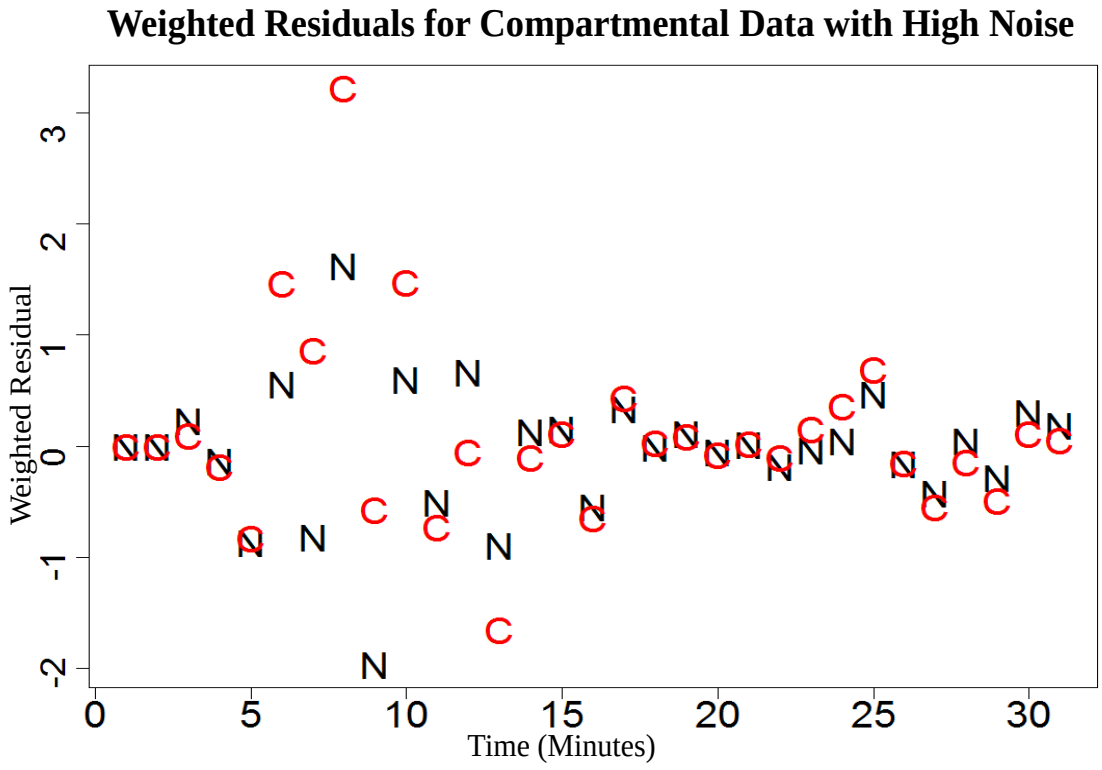
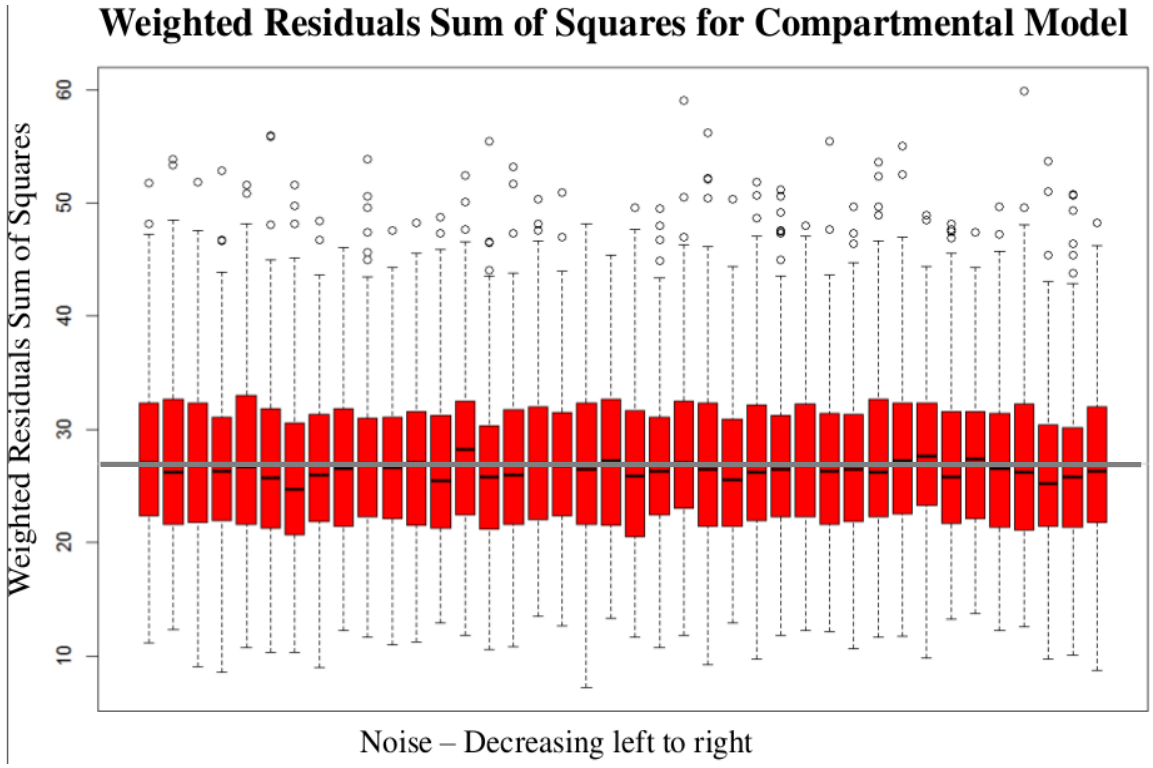


Figure 3.19: *Residuals for both compartmental and nonparametric models.*

These estimated parameters are shown in Figure 3.22. Eyeballing this Figure suggests the parameters may be constant, for the WRSS for the compartmental model. Accordingly we carry out a linear regression analyses to confirm this. The slope of the line fitted to the compartmental model is not statistically different from zero. In the case of the nonparametric model, however, there is clearly a slope present and this is far from being insignificant.

It is also worth highlighting at this point in the the linear model, the degrees of freedom of the WRSS would have degrees of freedom being the difference of the number of data points and the number of parameters we seek to estimate in our model. In the case of the compartment model this is $31 - 4 = 27$ degrees of freedom. For the nonparametric model this result is more complicated and as is shown here, the degrees of freedom are certainly not constant.

Whilst it is intuitive to seek a form for the RSS based on the χ^2 distribution, it must be validated and cannot be assumed that it is working without statistical evidence supporting it. All we have done so far is looked at histograms and boxplots to estimate the distributions in the data. We now consider QQ-plots



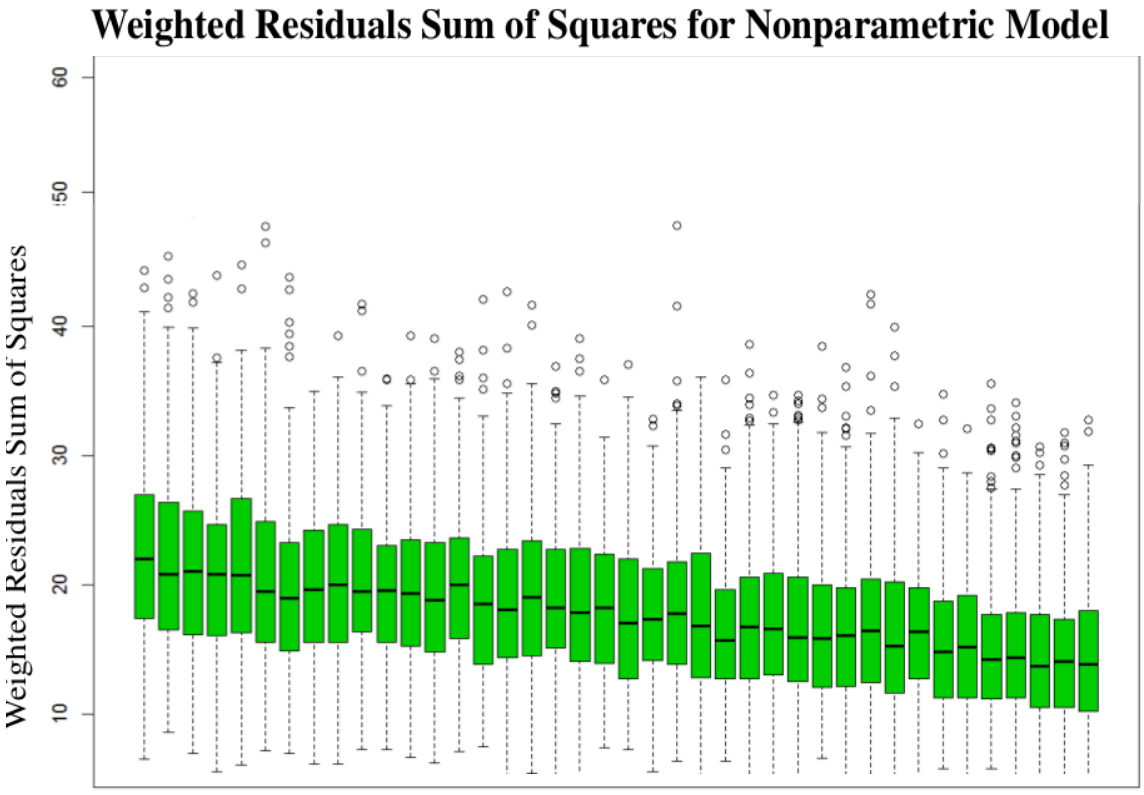


Figure 3.20: *Box plots for residuals sums of squares for both parametric and nonparametric models with various levels of noise.*

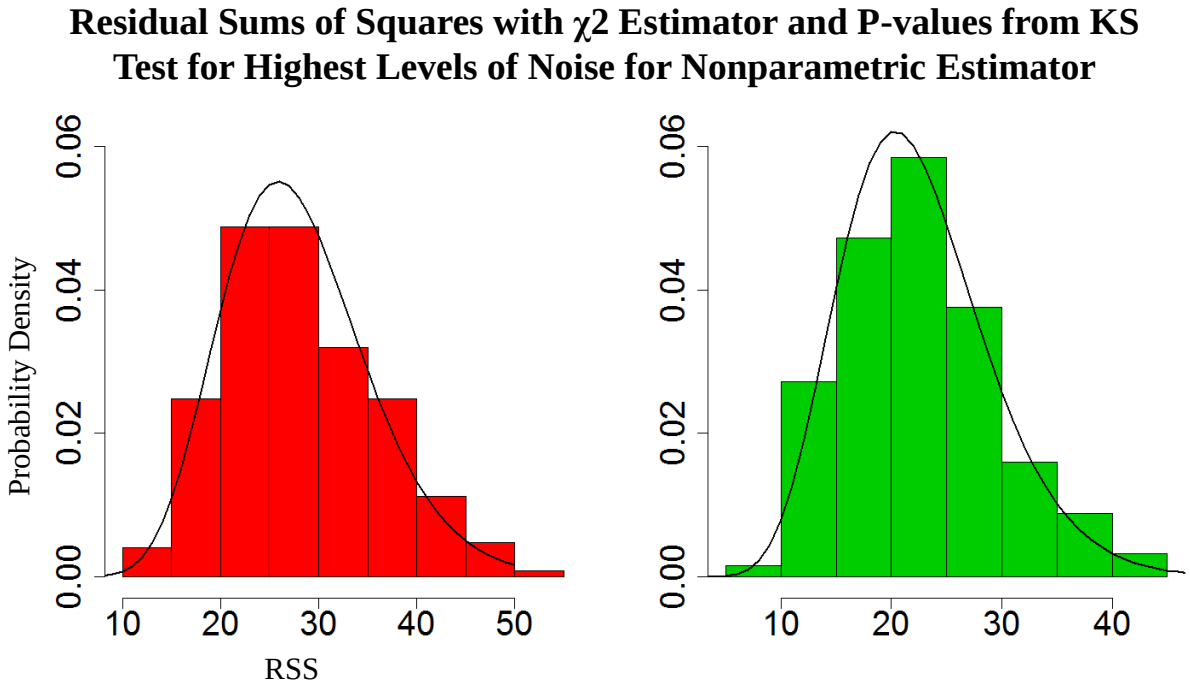


Figure 3.21: *Illustration of the χ^2 distribution fitted to both the compartmental (red) p-value 0.87 and nonparametric (green) p-value 0.73 models.*

comparing the simulated data with the χ^2 distribution can be drawn such as that shown in Figure 3.23.

It is not acceptable however to rely on a graphic such as a QQ-plot for the decision as to whether the model fits or not. Accordingly we look at the Kolmogorov-Smirnov test. This test checks to see if a dataset follows a theoretical distribution by examining the distances between the observed data and theoretical distribution. The p-values from this test are shown in Figure 3.24. If the null hypotheses is true then the data are χ^2 and the distribution of the p-values will be uniform. In this case we see no evidence leading us to reject the null.

As it appears that the data follow a χ^2 distribution, we can now proceed to look at the distribution of the ESS, which follows an F-distribution if the numerator and denominator are not correlated and both χ^2. Similar to what was shown for the ESS we present box plots for RSS for comparing the parametric and nonparametric models with various levels of noise in Figure 3.25.

An illustration of the F distribution fitted to a dataset is presented in Figure 3.26 along with the QQ plot for how the data matches the distribution.

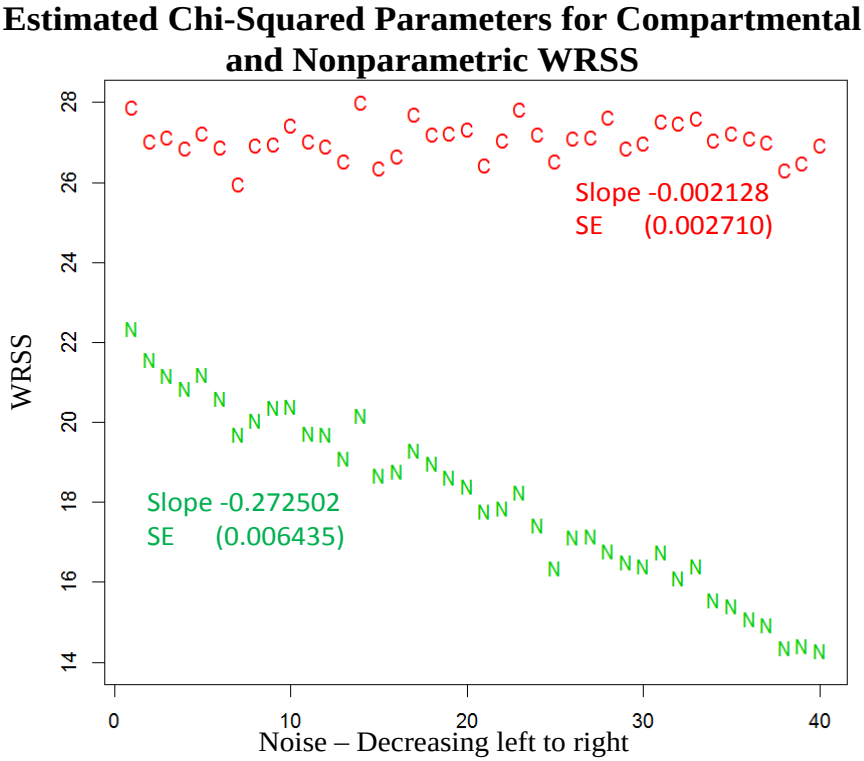


Figure 3.22: *Estimated degrees of freedom for the χ^2 approximation for the WRSS at varying levels of noise along with estimates of the rate of change of parameters with noise.*

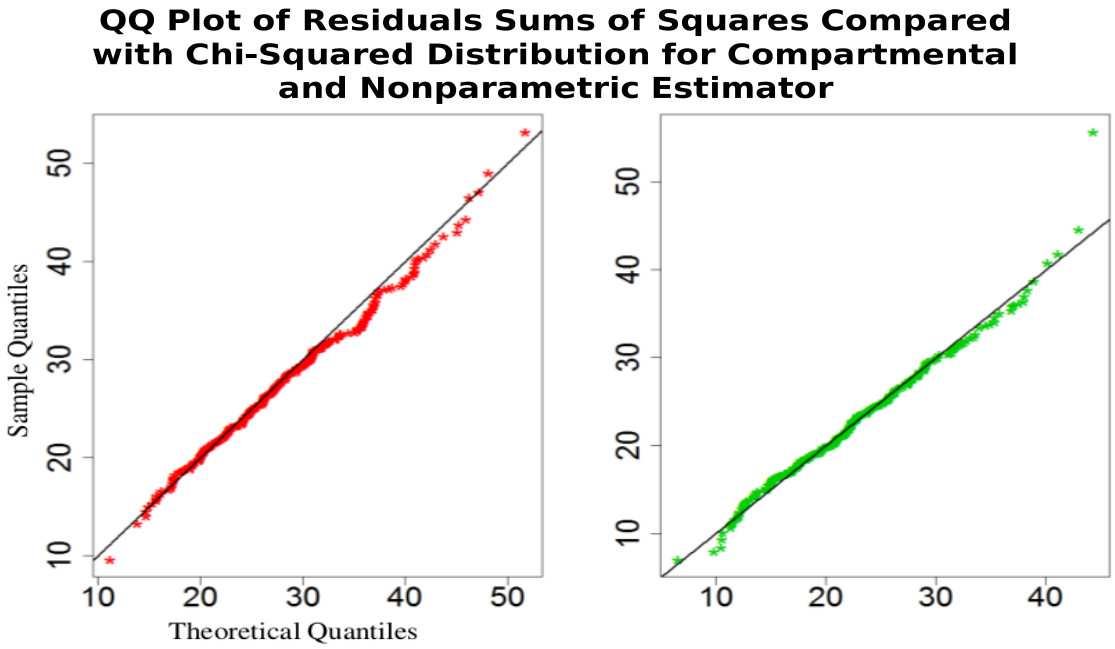


Figure 3.23: *QQ-plots for χ^2 approximation of RSS for compartmental (red) and nonparametric (green) models with 29 and 15 degrees of freedom respectively.*

As the data do not appear to follow the F-distribution, we now consider if there is an appropriate transformation which can change the shape of the data and allow the distribution of the ESS be modelled? We hope to gain an insight into how the models compare by doing this. We consider taking the square weighted residuals, log transformations etc. It is found that if the sample size is significantly reduced (to approximately 20) and hence the power of the Kolmogorov-Smirnov test is reduced then taking the square root of the data does make it possible to fit an F-distribution to the data. However this is far from a result which would allow a statistic like this be used to compare models. Also, as the degrees of freedom for the nonparametric estimator varies with noise and it is unclear what should be the choice for real data, we proceed with the use of cross validation and simulations to compare models.

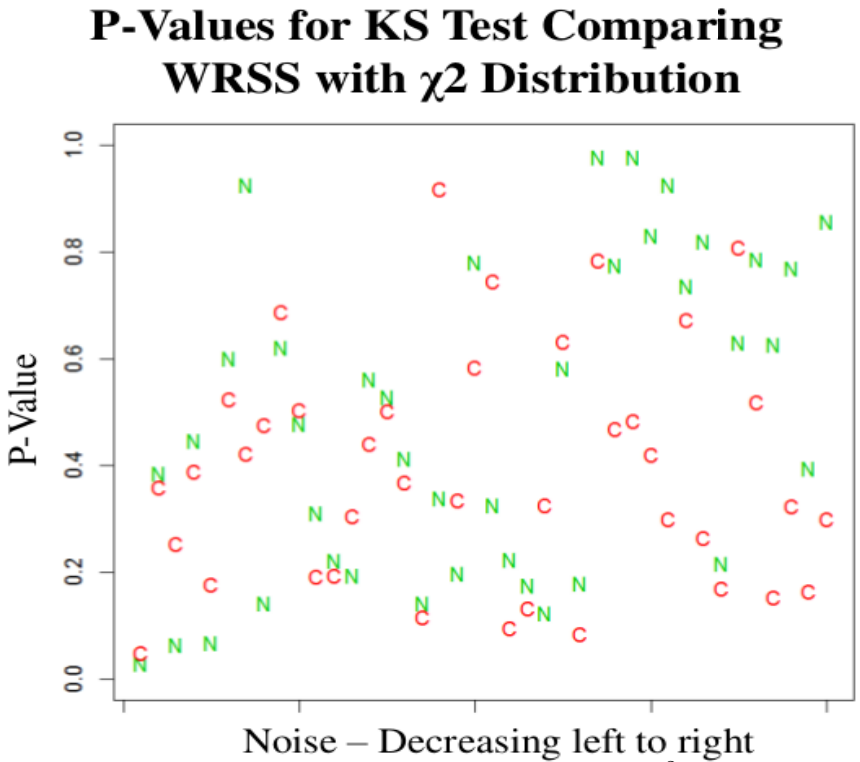


Figure 3.24: *P-values from comparing WRSS to χ^2 distribution.*

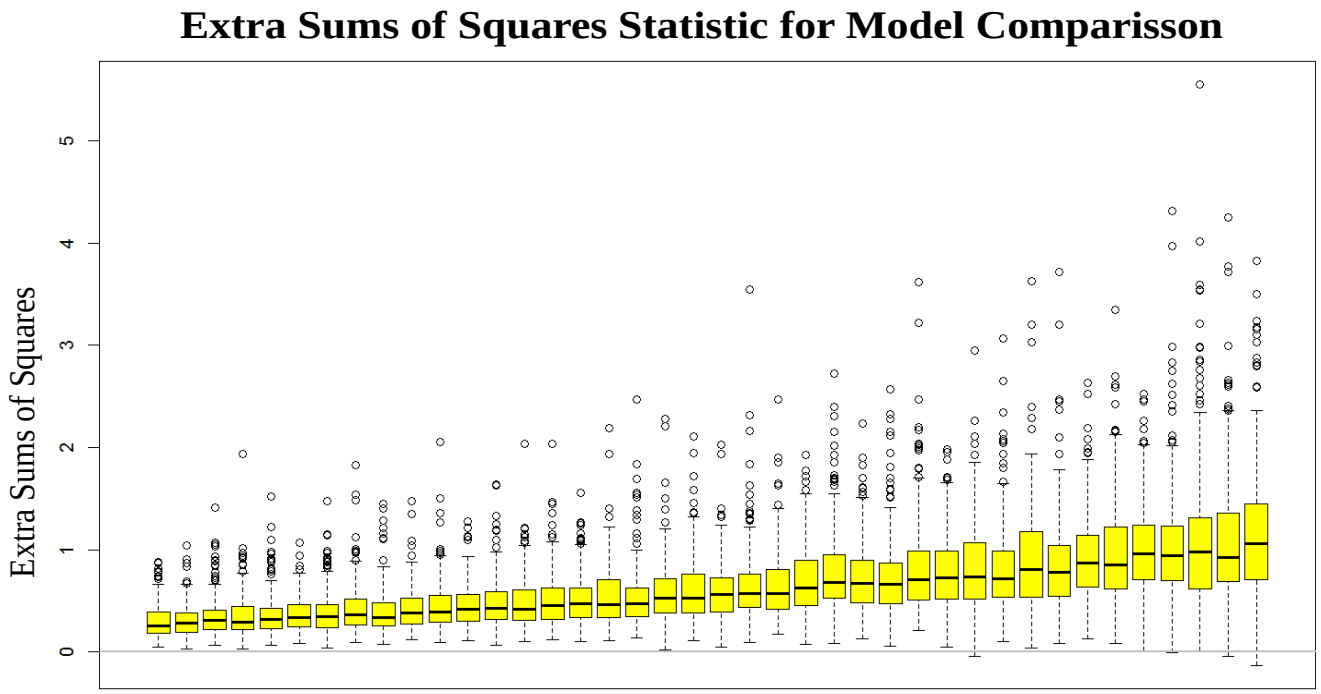


Figure 3.25: *ESS Calculated from for both model.*

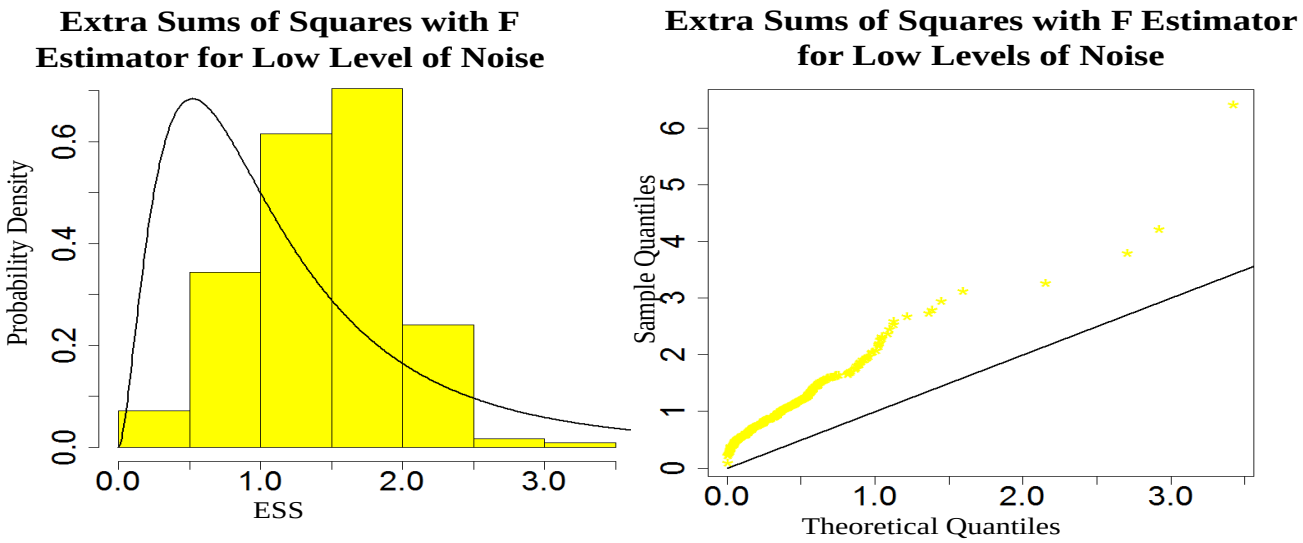


Figure 3.26: *Fit of F distribution to ESS data and associated QQ plot.*

3.5 An Illustration

3.5.1 Data

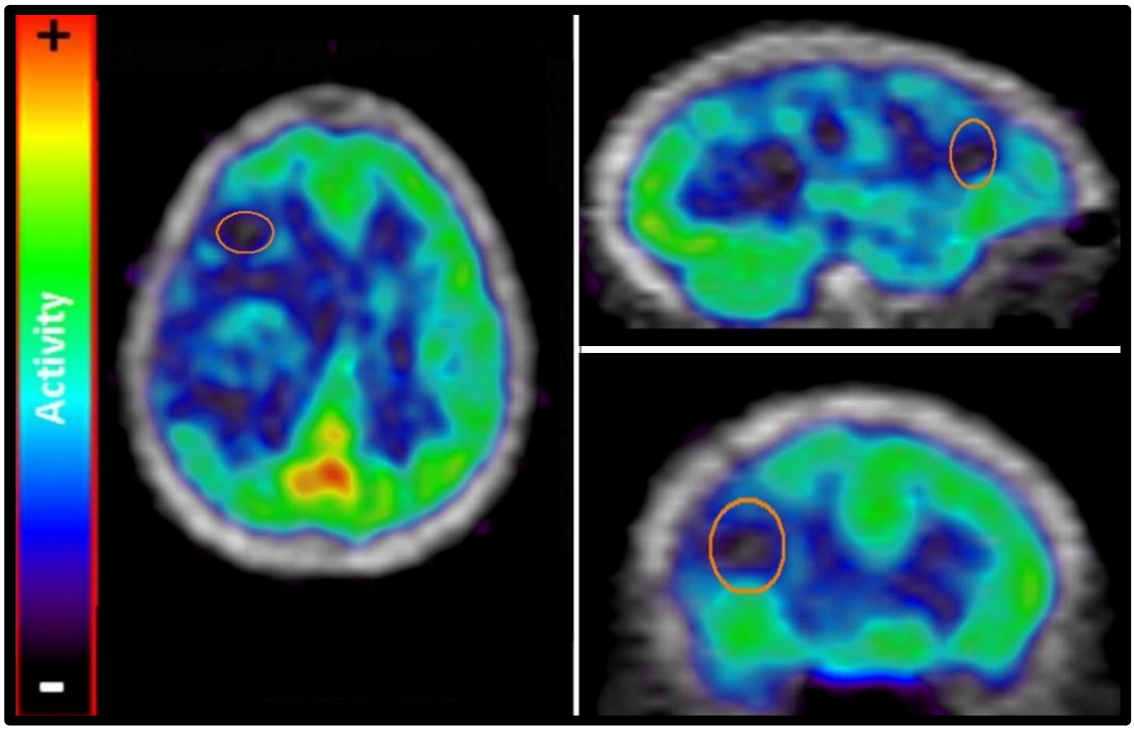


Figure 3.27: *PET scan and CT attenuation image of subject with resected glioma region highlighted.*

As per the previous chapter, we present kinetic analyses for an FDG-PET study conducted on a GE Advance scanner at the University of Washington, Seattle from the glioma dataset [71]. This analysis is also on a whole brain region in a subject with a glioma as was the case in the previous chapter. Again the measured arterial time-course is available.

Figure 3.27 displays tracer uptake in our FDG study. A region of interest for a hypoglycemic tumour region is highlighted but as this analysis is for illustration purposes, we consider the whole brain as our region of interest. This image was generated using AMIDE [42]

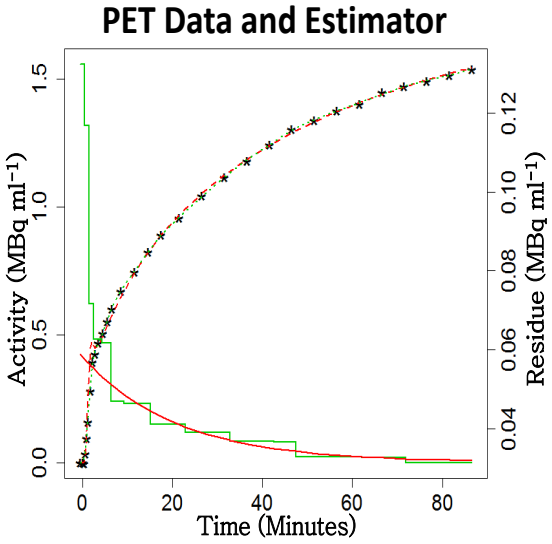


Figure 3.28: *Time activity curve for a whole brain region with nonparametric (green) and compartmental (red) model estimators, with residues.*

3.5.2 Statistical Inference

Figure 3.28 shows the time-course (adjusted for decay). The compartmental model and nonparametric fits are also shown here along with their residues. One can clearly see that the nonparametric residue is larger at the start while the compartmental model residue is restricted to being an exponential curve. The increased flexibility in the nonparametric estimator allows the curve to start at a higher value and finish at roughly the same point. This makes one consider two potential scenarios. Firstly the compartmental model may not have enough flexibility to fit the data. Alternatively the nonparametric method may be overfitting the data.

It is always useful to examine residuals after doing a regression and even PET with complicated models, is no exception. These are shown in Figure 3.29. It appears that both models have variance decreasing over time. The compartmental model has three points between five and ten minutes which appear to be outliers. If these were removed the variance would appear almost constant. On the residual plots the red and green dots are indistinguishable for the first four points and so only the green dots are visible. It is not possible to make these residuals look constant by removing a few points. Boxplots of the weighted residuals are also shown as these provide an insight into the shape of the distribution of the data.

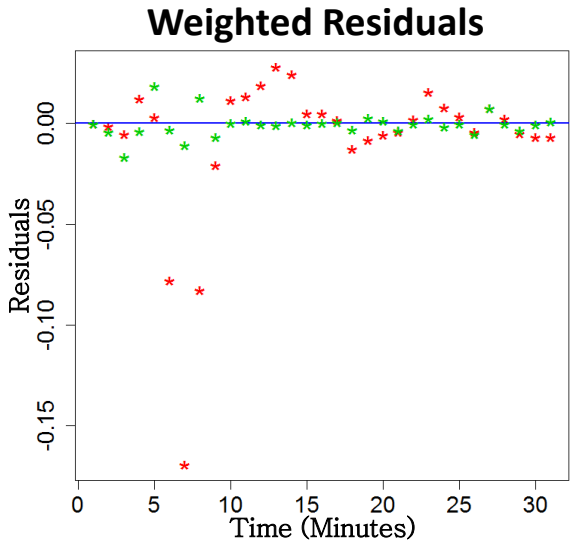


Figure 3.29: *Weighted residuals for the nonparametric (green) and compartmental (red) model estimators.*

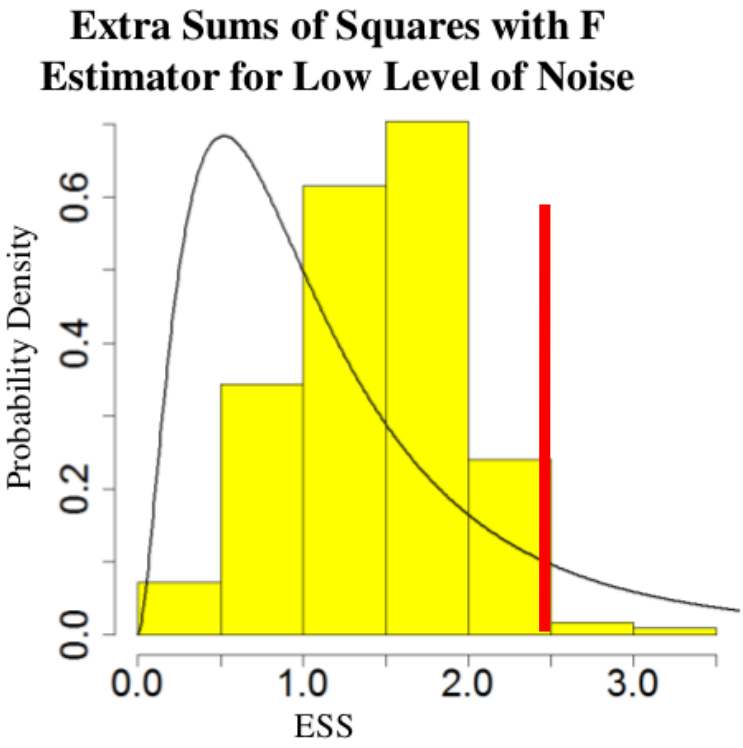


Figure 3.30: *Simulated distribution for F-statistic with critical value in red.*

To simply fit both models and then eyeball them and say that one is better is not sufficient. Two techniques used in comparing models, mentioned previously in this chapter will now be applied to the data.

The illustrative kinetic analysis of the time course data indicates that although the two compartmental model would typically be used for FDG, this model is found to be inadequate. The first involves bootstrapping and has its roots in regression. Here a simulated distribution for the familiar F-statistic in ANOVA is used. Figure 3.30 shows the reference distribution along with the test statistic (red vertical line). Here the test returns a p-value of 0.02.

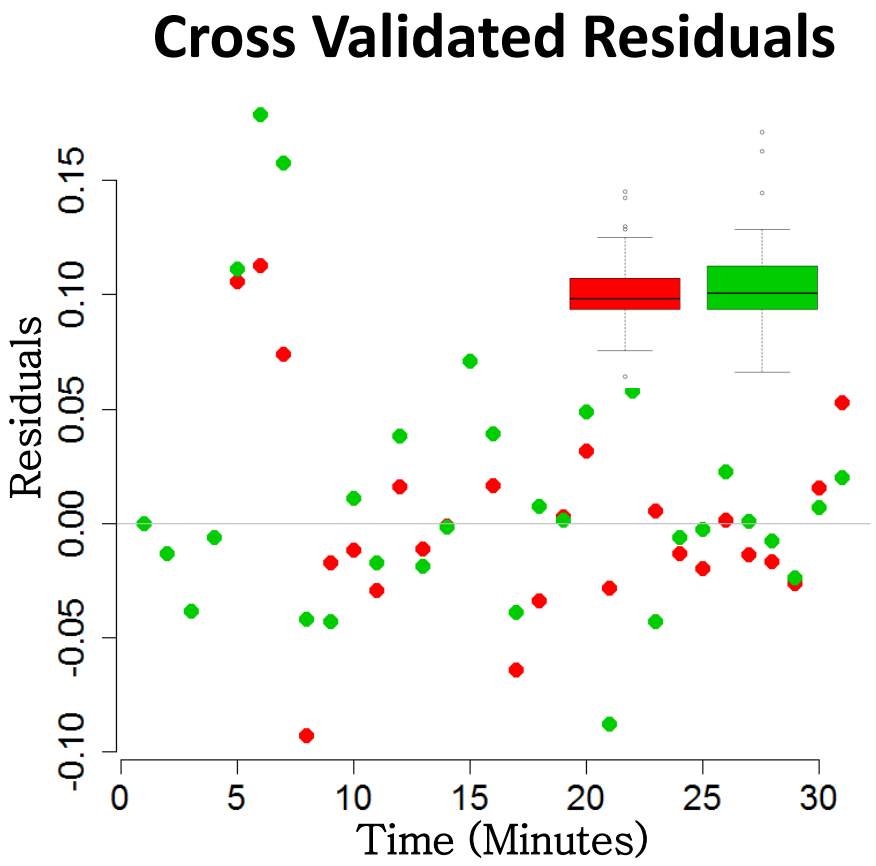


Figure 3.31: *Cross validated residuals for compartmental (red) and non-parametric (green) models.*

As an alternative method of comparing models, we examine leave one out cross validation. A plot of the leave one out cross validated residuals is shown in Figure 3.31. Comparing these residuals using a Wilcox signed-rank sum test gives a p-value of 0.008. If one wishes to compare the magnitude of the cross validated RSS, one finds their sums are 0.0647 and 0.00562 for the two compartmental and nonparametric models respectively. The Wilcox test rejects the hypothesis that the models are equal and the nonparametric model has the smaller residual sums

of squares making it the more appropriate model.

3.5.3 Discussion of Illustration

The results of the illustration are summarised in this section. We find that it is possible to implement the nonparametric method on PET data. In fact there are significant advantages with the nonparametric form. As it is estimated using a constrained linear model, convergence to the minimum value of WRSS is guaranteed. This is not the case for the compartmental model as finding starting values for the estimation procedure which converge to the true value is not trivial. Using the tracer function, which is part of the $nls()$ function in R, it was possible to see where the model was converging to an incorrect minimum. In this case blood volume was negative. The starting value for delay was adjusted and this allowed the model to fit the data. The nonparametric regression is also significantly faster than the compartmental model. The difference in time is difficult to quantify as it is a function of how close the starting values for the compartmental model are to the final estimate.

Both the bootstrapping and cross validation methods find in favour of the nonparametric method over the compartmental model. The nonparametric method estimates flow as approximately three times the compartmental model estimate. The values for flux and v_D are similar.

While the results here provide evidence in favour of the nonparametric model in one dataset and also show that there is a significant difference between estimates for flow for the different methods, this is only for one ROI in one subject. Repeating this work on several datasets gives a more general idea of how the models are performing and this is done in the next chapter.

3.6 Summary

In this chapter the compartmental models and a nonparametric model for modelling the tissue time activity curve were implemented in R [65]. An analysis of the fits of the different models to data was conducted. This analysis is similar to that conducted in linear regression comparing models but cannot rely on the

form of test statistics distribution so bootstrapping and cross validation is used to test the hypothesis that the compartmental model is an appropriate fit against the alternative that the nonparametric model is the true model. In the case of the data presented here, evidence against the compartmental model is found.

While the nonparametric method shows promise, it remained to compare the compartmental models and nonparametric methods in terms of examining convergence consistency variance and bias. A simulation study was conducted to show how the models perform when the data are truly compartmental. Also the statistic used in comparing the models was examined. In terms of comparing models, the flux estimate showed significant bias with the levels of noise seen in practice. Aside from this the nonparametric model converged to the true values. The test statistic was examined and transformations generated which lead to a distribution, which appears to follow the F-distribution. However, the results are not strong enough to merit its use in practice and so a bootstrapping approach will be used. The nonparametric technique applied to the ^{15}O H_2O and FDG data sets has been published in [28].

To summarise here we note that

- Nonparameric model measurably out performs compartment model.
- has difficulty in convergence.
- has fewer outliers in weighted residuals.
- does not have the questionable assumptions of the compartmental model.
- achieved convergence on compartmental model data for all parameters

Chapter 4

The Mixture Model and Analysis of the Complete FDG Glioma Dataset

4.1 Introduction

In this chapter we review a mixture model for the residue. This model has previously been published in [59]. The model is similar in some ways to the models presented previously as it is also used to generate a survival curve to represent the tissue residue function, however it is different in that it generates this curve in an alternative way. This model begins by modelling the time-course of every single voxel in the field of view using the nonparametric method proposed in Chapter 3. Having a massively large set of time courses is not particularly useful for modelling. Hence the number of time-courses is reduced to a smaller set, which capture the variability of the data. These can be used as basis elements in the modelling of ROI time course data.

This involves an analysis of the entire field of view and so maps of kinetic parameters are generated, which will be presented in this chapter. It is also worth highlighting the size of the dataset. Each image is approximately $128 \times 128 \times 35$ voxels over 31 time bins, which is approximately 17,776,640 entries, making up 573,440 time-courses, each of which will be modelled individually as part of the mixture modelling technique.

In Chapter 2 and Chapter 3 a compartmental model and a nonparametric model were introduced respectively and were applied to PET ROI time-course data for illustration purposes. It was highlighted that an analysis of a larger data set containing many scans could allow stronger conclusions to be drawn on the performance of each model. Such an analysis is undertaken in this chapter. This analysis includes the mixture model as well as the compartment and nonparametric models being applied to FDG glioma data and comparisons between all three models are generated, using cross validation, making this a task involving a lot of calculations on a large number of data points.

The data used here contains forty two subjects, with four ROIs per subject. These four regions, which are analysed in each subject are a whole brain region, a grey matter region, a tumour region and a white matter region. These data have previously been analysed using compartmental modelling and this was reported by Spence *et al.* [71]. The work in this chapter is a secondary analysis on the dataset. While the original analysis sought to understand the differences between how FDG and glucose are utilised in cerebral tissue, secondary analyses such as these allow new techniques such as the nonparametric and mixture models to be tested and validated.

While much of this thesis to date has discussed models, statistical inference and other techniques in statistics, a lot of this chapter, particularly its latter half, discusses the imaging aspects of the problem focusing on the data and results of the analysis rather than the techniques, that get us to that point.

4.2 Outline and Contribution of this Chapter

Having briefly introduced this chapter, we now describe the structure and contribution of it. The methodology is divided up into two. The mixture model is presented and reviewed in Section 4.3, which details the nature of the modelling of the whole image on a voxel level as has been proposed by O'Sullivan *et al.* [59]. The difficulties associated with such a large volume of data along with how to implement this mixture method for PET ROI data are also discussed. The method analyses each voxel individually applying the nonparametric method and subsequently applies a backward elimination algorithm to reduce the number of

time-courses which can represent the variability in the data i.e. segment the entire image. These segments are used as basis elements to generate an estimate for the true residue using a least squares approach. Similar to the case of the nonparametric model in Chapter 3 this model will be applied to the data using Gromping's ic.infer package [25] in R [65].

Section 4.4 discusses the data used in this chapter. It should be highlighted that conducting an analysis on a large dataset such as this is a contribution in itself and if one views papers published in the PET literature such as those contained in Mankoff *et al.*, Spence *et al.* and Muzi *et al.* [46, 71, 51, 52], one can see similar analyses to that conducted here. Another example of a similar contribution can be found in O'Sullivan *et al.* [60], where a secondary analysis of an FDG dataset in normal human subjects proposing a nonparametric residue function is presented. This sort of work is also of interest to physicians and other scientists who are interested in the imaging side of the science instead of the purely statistical side and this chapter aims to make a contribution of similar magnitude to the work mentioned above.

An illustration is presented in Section 4.5, which includes comparisons between the three models. This illustration highlights the value in the work of this chapter. The goals of this work include showing that the alternative modelling techniques are feasible in this data set and in PET data generally. This illustration accomplished this by giving a step by step guide on how to do the modelling as well as generating useful results on a single dataset. The comparisons between the models in this section allow us to decide on which model is more appropriate and hence which model should be used to estimate parameters of interest. Such a comparison allows one to see where the new model outperforms the previous model and this helps highlight the contribution that the new models are to the field.

Kinetic maps are a tool which is gaining a lot of interest in medical imaging with PET. They allow key functional parameters to be viewed throughout the entire PET image to gain an insight into the functional properties of the subject being studied. Generating and presenting these maps is one of the contributions of this work. These maps are generated using the nonparametric method as part of the mixture modelling approach. Again these are potentially a very useful tool

to clinicians in practice and in the future they may be used widely to decide a course of treatment.

To put the contribution of this work in context, the concept of the nonparametric model presented in the previous chapter has been published in Hawe *et al.* [28], which analysed one FDG time activity curve, this work shows its application to a large dataset. Aside from comparing the models, this paper highlights the feasibility of a nonparametric model for dynamic PET data. Here we seek to apply the nonparametric and mixture residues to the data alongside the compartmental model to determine which is best, which takes this work further.

The results of the analysis on the entire dataset will be presented in Section 4.6. In a sense this is a generalisation of the illustration but it is worth highlighting that the conclusions which can be drawn from this larger dataset are of much greater value than one simple study, which helps the goals of this work to be realised by allowing stronger conclusions be drawn. Trends in parameters can be found. The relationship between parameters of interest estimates for different models can be computed and any differences between the models can be observed. This analysis contributes strength to the conclusions which are drawn later in this chapter. The chapter finishes with discussion and conclusions in Sections 4.7 and 4.8 respectively.

4.3 Mixture Model

4.3.1 Voxel Level Modelling of Dynamic PET Image Data

A PET scan image as viewed in AMIDE contains coordinates and activity values. At each coordinate in a dynamic scan there is a value of activity for every time bin. These values have been discussed in Chapter 1 but in all the work contained in this thesis and almost all the work on modelling PET data in the literature, the researchers studying the data usually narrow the field of view down to an ROI and fit a model to the average of the data in this ROI. The models are usually compartmental models or alternatives such as spectral analysis or a nonparametric method like that proposed in Chapter 3.

In this section a mixture approach is proposed. It involves considering each

voxel as an individual time-course. As every voxel has a value at each time point, and is a tiny region contained in the field of view, each voxel can be considered as a separate ROI. For this analysis this is done for every voxel. The nonparametric model is fitted to each of these time-courses. The nonparametric model is chosen for many reasons including the fact that convergence is guaranteed and the fact that it is much faster. However, it should be noted that since we are dealing with the smallest regions possible then the compartment model would have been reasonable here as an alternative as the assumptions of homogeneity and instant mixing within compartments would not be violated. At this time it is worth highlighting similar work which has been previously been examined by Hernandez *et al.* [30] who examined the effect of looking at voxel level modelling with a focus on biases and variances of estimators.

A large amount of information can be obtained from an analysis at this level and in the coming sections kinetic maps are discussed. However as we seek to use all these time-courses to model ROI data, we will seek to reduce the number of basis elements for this modelling from the number of voxels in the field of view to a smaller more useful number and this will be discussed first.

4.3.2 Segmentation

The split and merge procedure used to segment the data is described by O'Sullivan [56] and O'Sullivan *et al.* [59]. Voxels are grouped together if their shape and scale are similar. Cluster analysis is used to divide the dataset into a large number, usually about ten thousand of rectangular regions whose scaled time-courses appear homogeneous. The merging procedure combines regions to larger regions with low relative heterogeneity. A constraint is added to ensure regions which are grouped together are contiguous when combining small regions. For the analysis in this chapter, ten segments are used. The choice of the number of segments is discussed in [56] but generally comes down to experimentation to see what works. By this we mean, as the number of basis residues will be reduced further it is necessary to have enough for this procedure. If the number is found to be too low it is increased, but that was not necessary for the FDG dataset.

4.3.3 Glucose Kinetic Maps

Kinetic maps have already been mentioned but in this section they are now discussed in greater detail. An example of kinetic maps from will be shown in the illustration later in this chapter, in Figure 4.5, the estimates for uptake, v_D, flow and flux in a subject with a brain tumour. These images were generated by analysing each voxel in the field of view individually using a nonparametric model and gaining an estimate for the parameters of interest at each voxel. These parameters are then plotted as an image such as that shown here. Similar analyses for breast cancer studies have been presented by O'Sullivan *et al.* [59]

4.3.4 Additive Modelling for the Residue

Initially the number of basis residues was set to be ten or twenty as was described in Section 4.3.2. If, following this segmentation procedure, a set of J basis residues $\bar{R}_j(t),\ j = 1, .., J$ which are derived from the whole image remain, then the residue in a region located at x is appropriately estimated as :

$$R(t, x) = \alpha_1(x)\bar{R}_1(t) + \alpha_2(x)\bar{R}_2(t) + ... + \alpha_J(x)\bar{R}_J(t),$$

for nonnegative α_j's. To simplify the modelling the residue components are normalised to start at unity and end at zero and extraction is included as a separate component. Using this residue and a similar argument to that put forward for the compartmental and nonparametric models, we write

$$C_T(t, x) = v_B C_P(t-\Delta)+\alpha_1(x)\bar{C}_1(t-\Delta(x))+\alpha_2(x)\bar{C}_2(t-\Delta(x))+...+\alpha_J(x)\bar{C}_J(t-\Delta(x)),$$

where

$$\bar{C}_j(t) = \int_0^t \bar{R}_j(t-s)C_P(s)ds \text{ for } j = 1, .., J.$$

That is to say the activity in the tissue is modelled as of a combination of the weighted residue functions convolved with the AIF.

The segmentation generates a mean and sample variance time-course for each segment. Similar to an ROI drawn in AMIDE each segment will be quasi-Poisson. If the mean value in the k^{th} segment for the b^{th} time-bin is denoted by y_{kb}, and

$$E[y_{kb}] = \mu_{kb}$$

then

$$Var[y_{kn}] = \phi_k \mu_{kb},$$

where

$$\mu_{kb} = \int_{t_s}^{t_e} C_k(t) e^{-\lambda t} dt$$

and

$$C_k(t) = \int_0^t R_k(t-s) C_P(s - \Delta_k) ds,$$

where λ is decay, Δ_k is delay and ϕ_k is the proportionality constant, which represents the relationship between the mean and variance of the quasi-Poisson process.

When fitting the linear combination of the segments to data, it is necessary to estimate the scale of each mean time-course of each segment using the $orlm()$ function [25] in R [65] for estimation. Once delay is computed by gridsearch, all parameters can be computed together using weighted least squares with the constraint that all coefficients are positive. Parameters such as flux, flow and volume are then estimated similarly to the compartmental and nonparametric models using the estimated residue for the ROI being studied. An unbiased risk assessment criterion is then used to obtain an overall assessment of the J-component basis [59].

4.3.5 Voxel Level Optimisation of Delay and α Coefficients

Voxel level data is similar to larger regions of data i.e.

$$E(y_{ib}) = \mu_{ib},$$

$$Var(y_{ib}) = \phi_{kb} \mu_{ib}$$

and μ_{ib}, where ϕ_{kb} is similar to ϕ_k but on a voxel level. We can now write as a linear combination of the basis residues

$$\mu_{ib} = \alpha_{i1} \bar{\mu}_{1b}(\Delta_i) + \alpha_{i2} \bar{\mu}_{2b}(\Delta_i) + .. + \alpha_{iJ} \bar{\mu}_{Jb}(\Delta_i) = \bar{(}\boldsymbol{\mu}_b)(\Delta_i)^T \boldsymbol{\alpha}_i$$

where $\alpha_{ij} = \alpha_i(x_j)$, $\bar{C}_j(t) = \int_0^t \bar{R}_j(t-s) C_P(s) ds$ and hence, $\bar{\mu}_{jb} = \int_{ts_b}^{te_b} \bar{C}_j(t - \Delta_i) e^{-\lambda t} dt$. Weighted least squares is used to fit this model to the data, with

weights similar to those used for the two compartmental model and the nonparametric model, however they are based on the reciprocal of the fit at each iteration not the data and at each iteration of the modelling they are updated to the new fit. This is a sensible choice to make as there will be a large amount of noise in the system. This optimisation is implemented using *orlm()*.

4.4 The FDG Glioma Dataset

4.4.1 Gliomas and FDG

Gliomas are a form of tumour, which develops in glial cells in the brain and spinal cord, which are generally quite aggressive and subjects diagnosed with this ailment generally do not have a good prognosis [41]. FDG is used to grade the severity, plan biopsies and to distinguish recurrence from radio-necrosis (death of cells due to radiation) in gliomas [71].

The use of FDG as a glucose analog has previously been discussed in this thesis. FDG has the advantage that it crosses the blood brain barrier similarly to glucose and then accumulates in brain tissue at a rate proportional to glucose. This dataset used here was generated when the relationship between glucose and FDG was being examined. Accordingly all subjects were imaged with ^{11}C glucose and subsequently with FDG. In this work only the FDG data is considered for analysis.

4.4.2 Subjects

Forty-two subjects, all with malignant gliomas were studied by Spence *et al* [71]. These included multicentric (tumours with more than one center) and bilateral (cancer which occurs on both sides of the brain) tumours. The study included both men and women with glioblastoma multiform and anaplastic astrocytoma, which are the two main types of glioma. The type and location of the tumour and time of detection, dictated the times that the subjects were imaged. This means that some of the subjects were imaged between surgery to resect the tumour and treatment. Others were studied after recurrence following treatment. Subjects were subsequently imaged at three month intervals following their initial scan

until death. This means the data includes a number of subjects who were imaged twice or three times.

4.4.3 Equipment and the Imaging Procedure

FDG was generated by bombarding the glucose particles in a medical cyclotron. The purity and activity in the sample was measured and recorded to calculate the dose of radiotracer given to the subjects. The imaging data used in this work was generated on the General Electric Advance whole body PET scanner which generated images made up of 128×128 voxels over thirty-five slices. The scanner contains eighteen rings of detectors with 672 scintillator crystals per ring [71].

Subjects fasted for at least nine hours before the procedure. All subjects underwent CT and MR scans as well as the PET procedure. From these scans, the planes which contained the largest tumour areas were identified. After the patients were secured in the scanner, the scanner was aligned so that the region containing the tumour was contained within the region being imaged. An attenuation image was generated using a source of ^{68}Ge (Germanium) with known activity and rotating it around the subject.

Following this an intravenous line was inserted into the subject to allow the tracer to be injected and a catheter was inserted into a radial artery at the wrist to allow the activity in the arterial blood to be examined and the AIF generated.

Once this was set up the scanner began generating images one minute before injection of the tracer. After the 1 minute scan the ^{11}C glucose was injected over 1 minute or 2 minutes in 10ml or 20ml of saline respectively. Images were then generated as follows: four 20 second bins, four 40 second bins, four 60 second bins, four 180 second bins, fourteen 300 second bins. Arterial blood was sampled at similar intervals to the imaging procedure. The FDG study followed the ^{11}C glucose scan. For FDG the tracer was injected in approximately 10ml of saline over 1 to 2 minutes. The same protocol for the image duration and arterial sampling was used for the FDG image generation as was used for the glucose image generation.

4.5 An Illustration

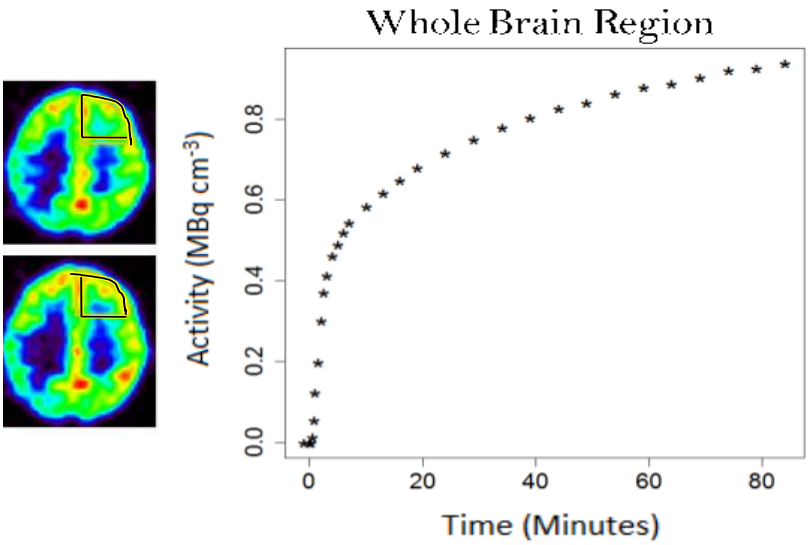


Figure 4.1: *Brain region for analysis in this section highlighted in thick yellow line and thin black line along with time-course from this region.*

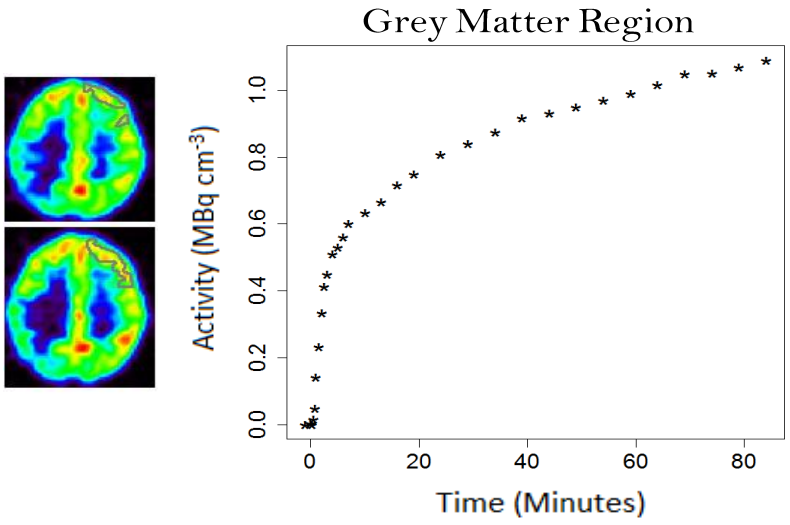


Figure 4.2: *Grey matter region for analysis in this section.*

As was the case in the previous two chapters, an illustration will now be provided which goes through the modelling in detail. We begin by considering the data being analysed. The data in this illustration relates to four ROIs from one subject in the glioma dataset. These four ROIs are a whole brain region, a grey matter region, a white matter region and a region containing the site where

a tumour has been resected from. These regions were used by Spence *et al.* [71] in their paper exploring the relationship between glucose and FDG and were made available for this analysis. All PET images previously displayed in this thesis were generated using AMIDE, but Spence *et al.* used an alternative viewer called Alice to generate the ROIs. The regions shown here have been redrawn in AMIDE as the images appear clearer and it is easier to remove noise from the image in AMIDE than Alice.

These ROIs were drawn using coregistered CT or MR images, which were generated on every subject who was part of the study. The whole brain region is roughly made up of the front right quarter of the brain on the twenty-seventh and twenty-eight slice and is highlighted in yellow in Figure 4.1. The grey matter region is also spread over the twenty-seventh and twenty-eight slices and is highlighted in grey in Figure 4.2. The white matter region is again on the twenty-seventh and twenty-eight slices and is highlighted in white in Figure 4.3.

The tumour region is spread over a larger collection of slices. This is for a number of reasons. Firstly as the whole brain, grey and white matter regions occur throughout the brain a large volume can be obtained on just one or two slices, while the tumour occupies a small volume on each slice. Secondly, as the brain tumour is not part of the anatomy of the brain and is by its nature made up of aggressive tissue, it can spread out in all directions and so is found on many

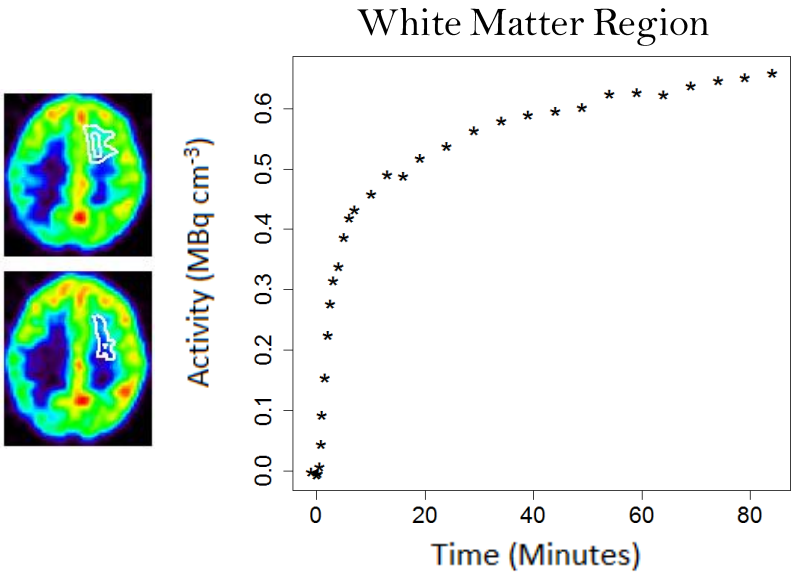


Figure 4.3: *White matter region for analysis in this section.*

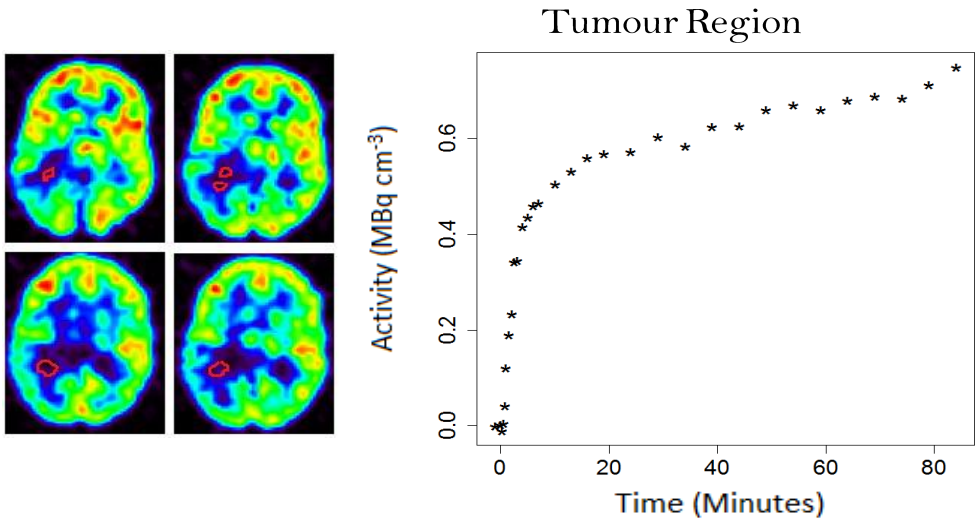


Figure 4.4: *Tumour region for analysis in this section.*

slices. For these reasons the brain tumour region, which is highlighted in red in Figure 4.4 is drawn across four slices from slice seventeen to slice twenty. It is also worth highlighting that there are far more voxels in the whole brain, grey matter and white matter regions than the tumour region, which explains why the tumour time-course has the most noise.

Illustrations of the compartmental and nonparametric models applied to data have been presented in Chapter 2 and Chapter 3 respectively and so this illustration will discuss the mixture model in greater detail than the compartmental and nonparametric models. As the compartmental model only takes approximately one minute to fit to the data once correct starting values are found and the nonparametric is even faster, this is not computationally difficult. As was outlined previously the modelling begins by fitting the nonparametric model to every voxel and then grouping these time courses using cluster analysis according to their shape and behavior, taking into account the proximity of voxels to each other. This is achieved using fortran codes and takes approximately thirty minutes per region of interest to produce ten segments within the image. This work generates a kinetic map of the entire dataset as the nonparametric model is applied to the data at every voxel. Examples of four of these kinetic maps are shown in Figure 4.5.

These kinetic maps show how the flux, flow, volume of distribution and uptake

vary spatially. The site of a tumour is in the cross hairs for each of these kinetic maps and the difference in flux and uptake is visible on these maps. A colour bar is also present on the image which allows the maximum and minimum values in the field of view for each kinetic parameter to be observed.

We are seeking basis elements to use in a similar way to how predictor variables are used in a linear model to apply to our data so the thousands of time-courses needed to generate the kinetic maps need to be reduced. The dendrogram used in the cluster analysis which achieves this along with the nonparametric residues generated by this cluster analysis are shown in Figure 4.6. It is worth noting that at this time it was decided to reduce the number of basis elements to ten, which was chosen as in all data examined, less than ten elements were required for subsequent steps in the process. These ten basis residues are then smoothed as has been described in the previous section. They have the extraction fraction subtracted off and are normalised to have height one. These smoothed residues are shown in Figure 4.7. These residues have been smoothed using cubic B-splines and have had retention removed so that they all end at zero. They are also rescaled so they all begin at unity. This is done using splines and this takes approximately 2 minutes.

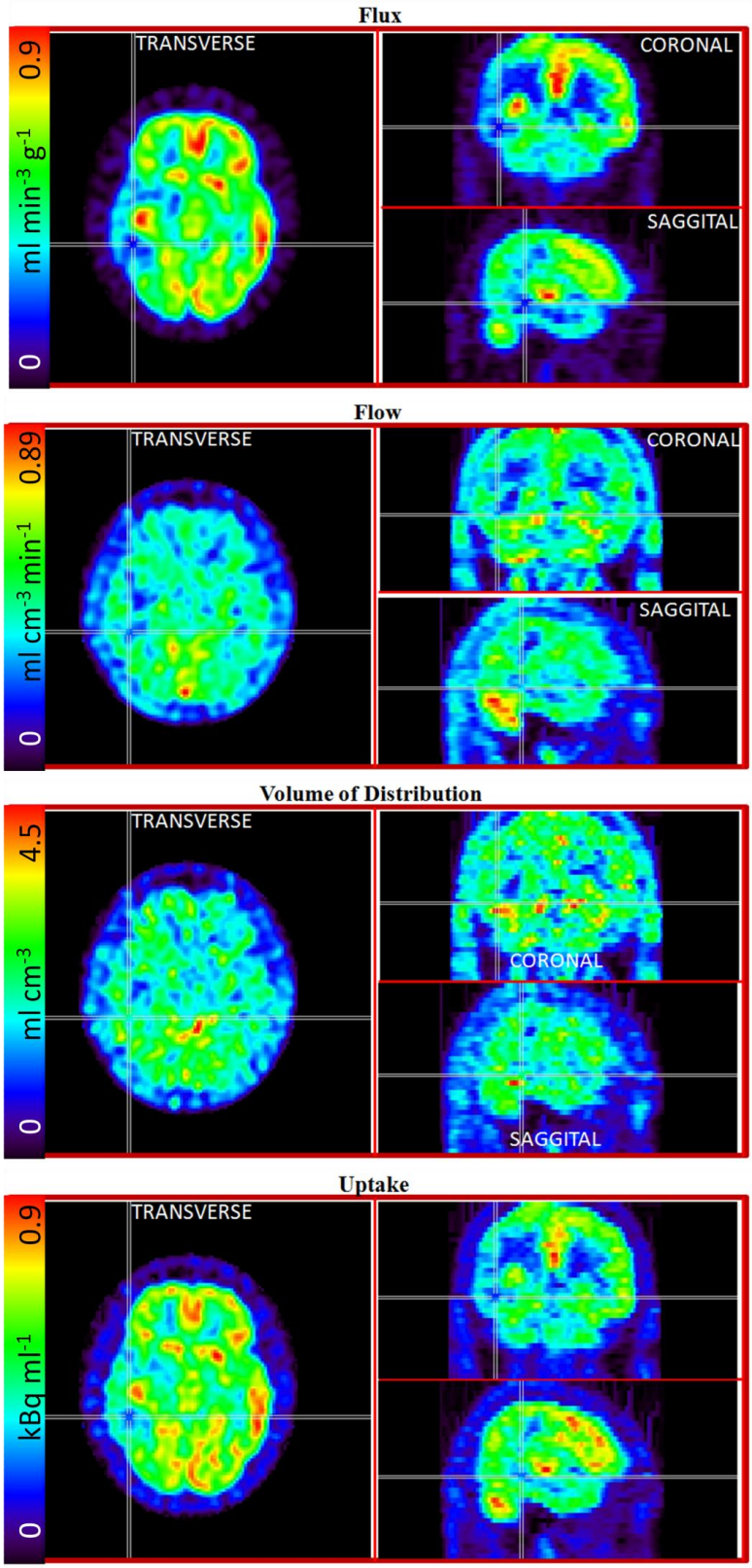


Figure 4.5: *Kinetic maps for flux, flow, volume and uptake from a glioma dataset. The site of the tumour is highlighted by cross-hairs.*

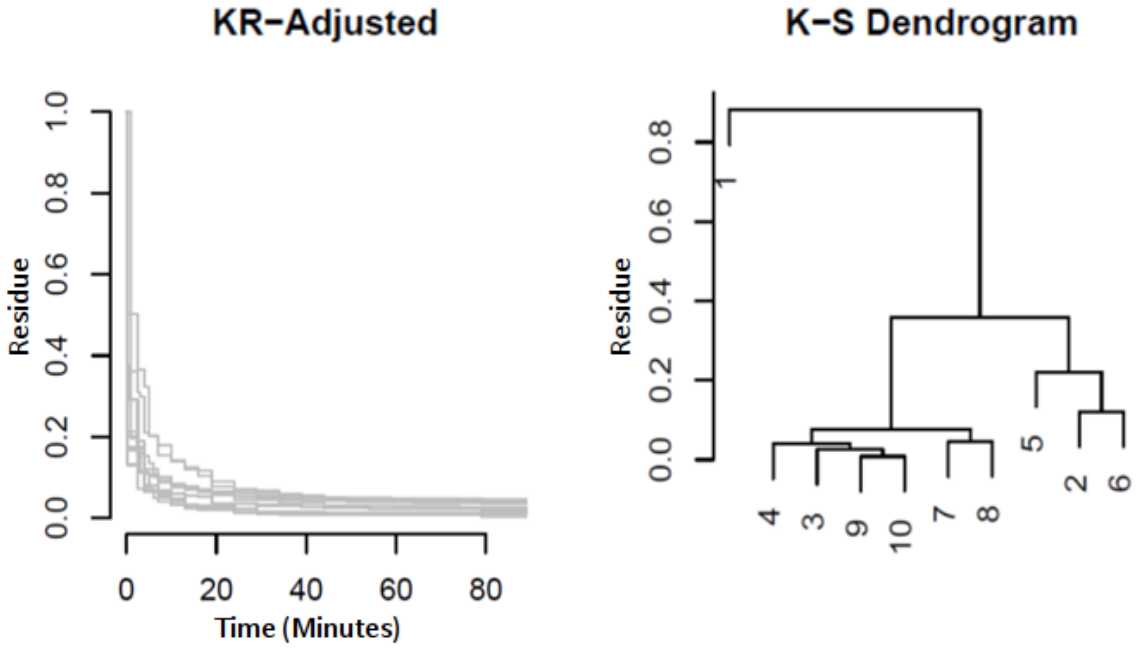


Figure 4.6: *Piecewise constant residues representing segments (left). A dendrogram used in the cluster analysis (right) [59]*

.

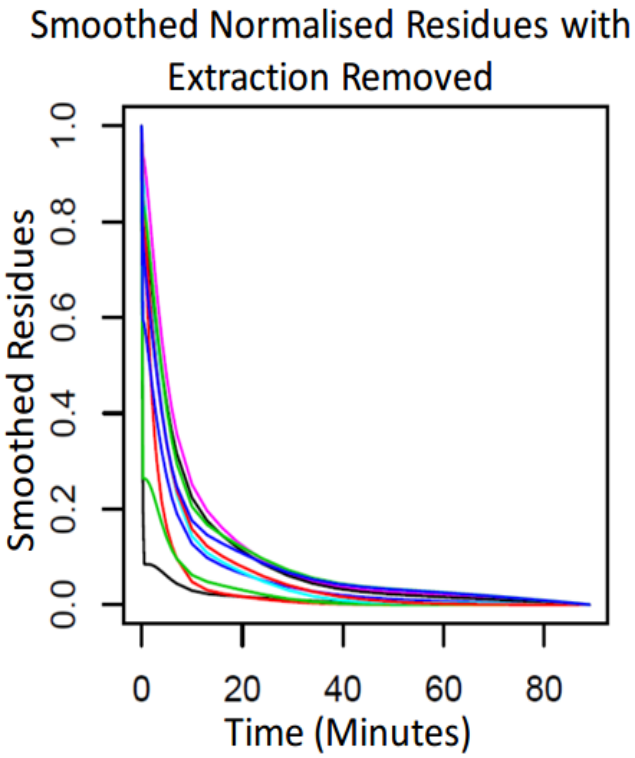


Figure 4.7: *Smooth residues representing segments from Figure 4.6.*

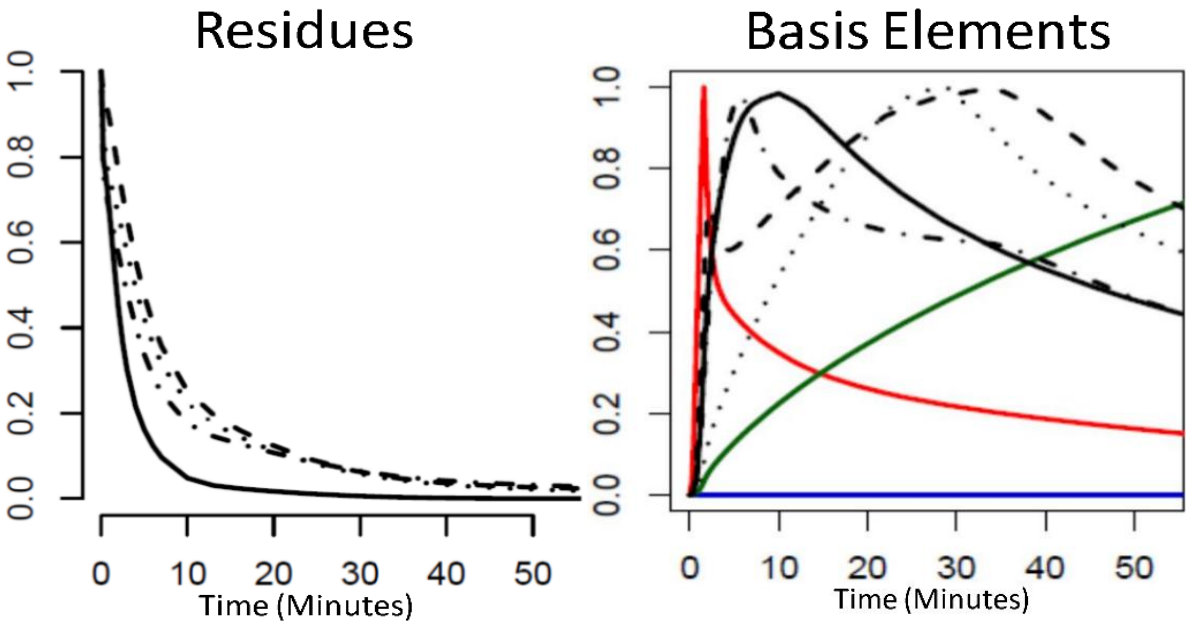


Figure 4.8: *Reduced set of residues and components representing segments in the data along with AIF (red) and extraction (green).*

It can also be seen in Figure 4.7 some of these residue functions seem quite similar and perhaps are explaining the same variability in the data. For this reason the number of components is reduced further to a smaller subset which can still represents the ROI data. This has been described in the last section with regard to the risk values $C(J)$ and is based on a principal component type study which looks at the variability each of the segments explains in the data. In this case the value which minimises this function is four and so there are four residues remaining. These residues are shown in Figure 4.8. This is why we initially chose ten basis residues using the segmentation procedure. As long as the number chosen earlier was bigger than four then the modelling will perform as needed. This takes approximately twenty minutes to run.

It is not the case that these residues are fitted directly to the data. Instead the convolution of these residues with the AIF are the components which are fitted to the data using constrained weighted least squares. These are shown in Figure 4.8.

Before discussing the fitting of these residues to the data, we consider what these represent. The components which the data are broken into are obviously different, but the way they are different is interesting. Huang and O'Sullivan [33]

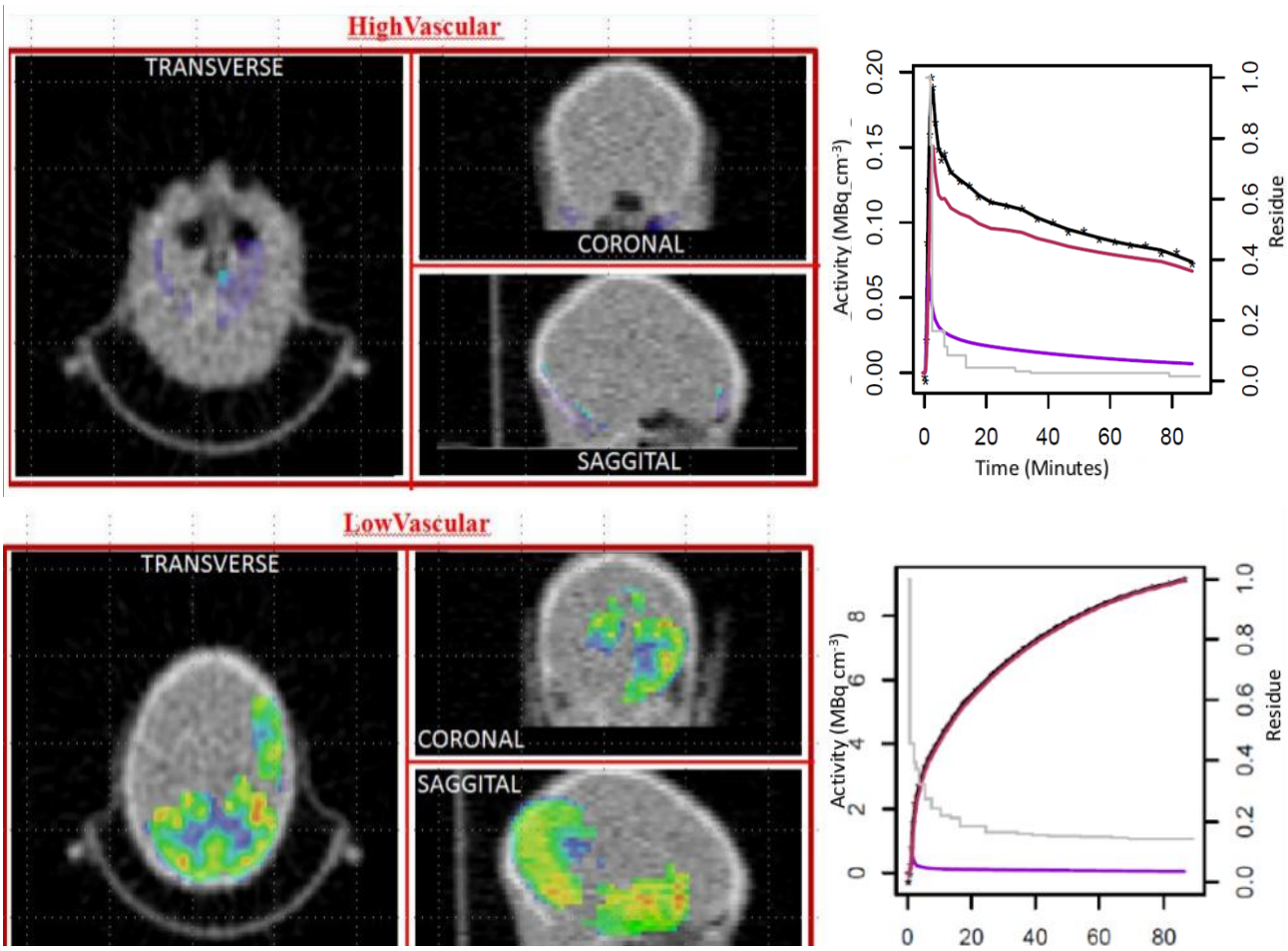


Figure 4.9: *High and low vascular regions used in mixture modelling and time activity curves. The regions are shown here without the surrounding tissue and with the attenuation image for clarity.*

used this segmentation procedure to find areas of tissue which consist of arterial blood and regions with lower levels of vascularity to generate an AIF from the image without blood sampling.

In the case of the FDG dataset, two examples of segments are shown. Figure 4.9 shows a segment with high vascularity and a segment with low vascularity. The time activity curves for these segments are also shown here. One can clearly see an AIF and an FDG time-course like curve with the largely vascular and largely extra vascular regions respectively. Also shown are the residues for these regions and the proportion of tissue and extra vascular space. However in the case of the work in this chapter, we are concerned not with the vascularity of

the regions but how they can be used as basis functions. However knowing this information helps us understand the variability being captured by the different components. We now return to consider the ROI time-course data in the glioma studies.

So far this chapter has discussed the mixture model but now the analysis will include the nonparametric and compartmental models and comparisons between all three models. Although previous illustrations have involved analysing one ROI, this illustration involves four ROIs from one subject and gives a more detailed picture of how the models work. In a sense this can be thought of as a more extensive illustration from the perspective of the imaging side of this work.

This model as well as the compartmental and nonparametric models are fitted to the whole brain, grey matter, tumour and white matter regions in Figure 4.10. In this picture one can see that the three models appear very similar for all four regions.

One can also observe the shapes of the time-courses of different regions. The grey matter regions time-course is increasing at a very fast rate, while the white matter regions maximum is approximately two thirds the maximum of the grey matter region. It is hardly surprising that the whole brain region time-course which contains both grey and white matter has more activity than the white matter region and less than the grey matter region. It is also evident in the fit that the tumour region behaves differently in terms of glucose uptake than the other regions as the data is almost linear after twenty minutes. This is because the tumour region, is actually from a region where tumour has been resected and so the region has previously been hypoglycemic, however the tumour has began to reoccur and this is why there is more activity present than in the white matter region. If the tumour was still hypoglycemic, it would be the case that the tumour's time-course would be decreasing at later times. From these pictures it is difficult to say that one model is fitting the data better than the others.

Even if there appeared to be a difference between the fits, it is not simply enough to look at the fits with one's eye and say one is better than the others. This is open to ambiguity. As a large goal in this work is determining which model is appropriate, then we look at the fits in more detail. In doing so we consider the review of linear regression, presented in Chapter 2 and consider looking at

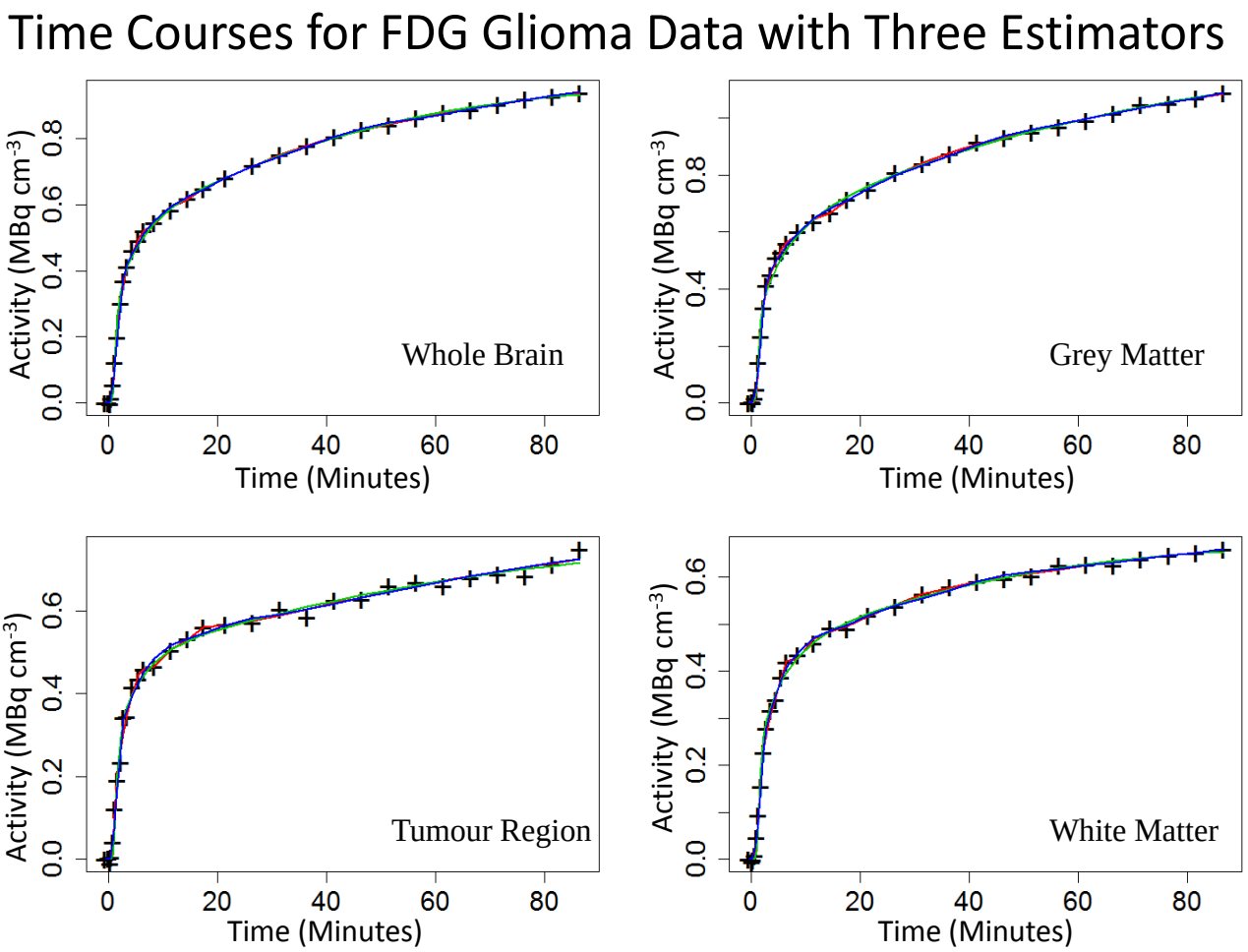


Figure 4.10: *Time activity curve fitted by compartmental model (red), nonparametric model (green) and mixture model (blue).*

residual plots of the fits. It has been mentioned that if the model is appropriate then the residuals will be a random scatter or null plot about the x-axis. The plots of the weighted residuals are shown in Figure 4.11. On this plot the x-axis is not time but simply an index as there would be too many points at early times making it difficult to gain an understanding of the differences between the data and the estimators and to examine if there are any patterns in the residuals. At a first glance these plots do not appear to be a null plot about the x-axis.

At early times, specifically between the fifth and tenth time points, there appears to be some significant increase in the residuals for the compartmental model and to a lesser extent the other models. These points correspond to the maximum value in the AIF and is visible for all four regions. This is not as

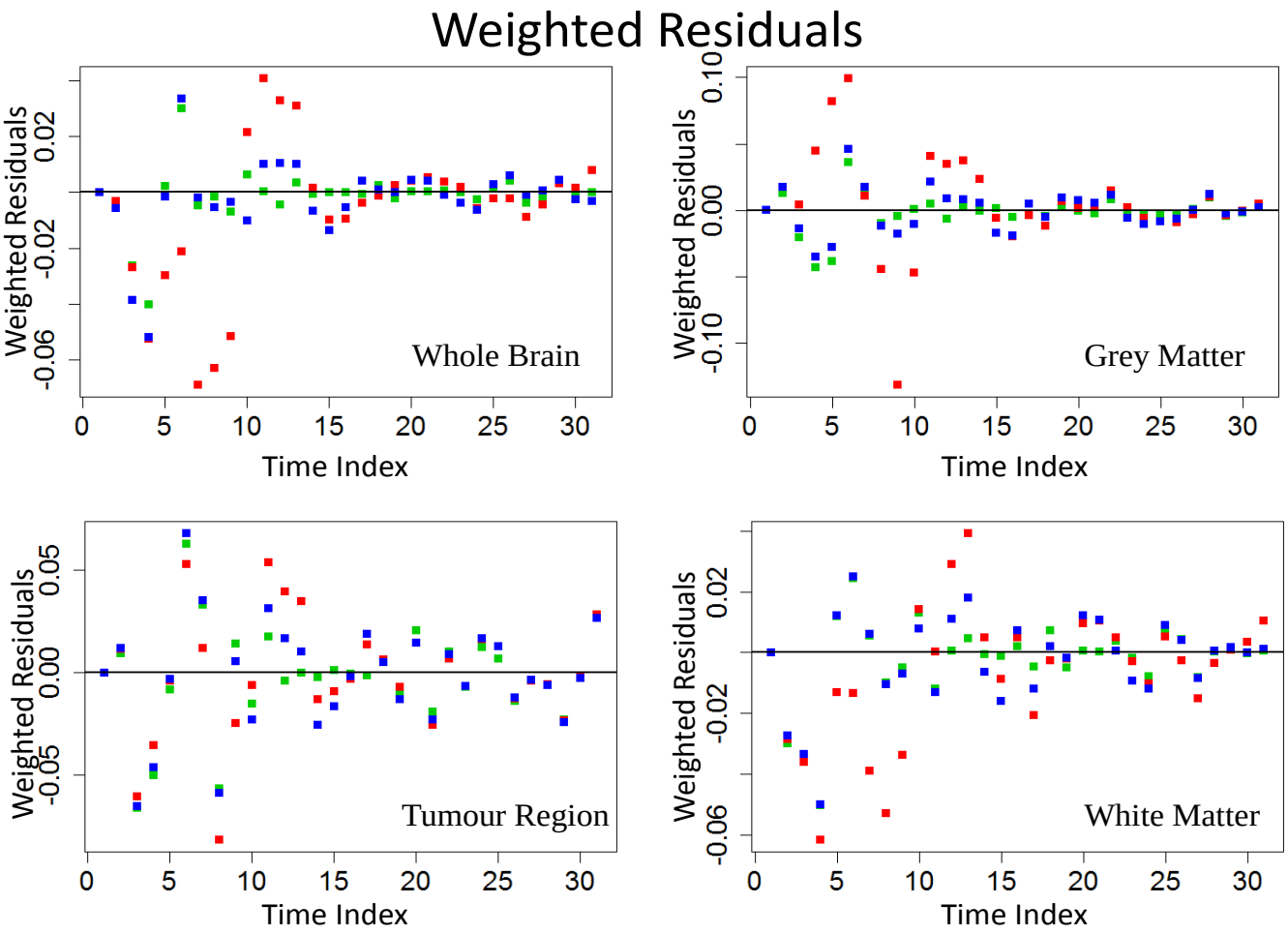


Figure 4.11: *Weighted Residuals for compartmental model (red), non-parametric model (green) and mixture model (blue) fits shown in Figure 4.10. Note these are on different scales but this is due to the shape of the residuals rather than their scale being of interest.*

noticeable for the white matter and tumour regions but is still present.

It is also worth noting that all three models appear to have the variance of the residuals decreasing over time. The fact that the variance is decreasing is not a positive for any model. While questions can be asked about the weights being used in the analysis, it appears that the variance is decreasing at a very slow rate or not at all for the last fifteen time points and that the earlier larger variance, when the AIF is at its maximum makes this appear far more drastic than it really is.

Another point of note is that instead of being a random scatter about the

x-axis, the residuals seem to make a zigzag pattern on the axis particularly for the whole brain and grey matter region. This might indicate some correlation between the residuals. This has been mentioned in the context of linear regression in Chapter 2 and it is worth restating that even if the errors are independent, residuals need not be. However in this case there does appear to be a pattern beyond this in the residuals.

The real question to ask when examining residuals is 'What do we take from this plot'? In answering this there are many questions asked of the compartmental models ability to take the arterial blood component into account. While similar questions are asked of the other models, they are not to the same extent. The fact that a pattern is present in the residuals does ask questions of each of the models. It will be necessary to generate a metric to compare these models. This will be discussed shortly.

Before this we consider this metric, we note that while residuals are one method in comparing the models, it is important not to lose sight of how the models are different and explore this. Examining these differences is key in understanding which model performs best. The three models have different residue functions and we will now look at these individually to try to gain an insight into each of the models.

In the case of the compartmental model, the residue is forced to be a mixture of two exponentials. This will be a smooth curve. The nonparametric is a step function and has more parameters than the compartmental model. Finally the mixture model residue will have flexibility similar to the nonparametric model, but will be smooth unlike the nonparametric model as it is is made up of components which have been smoothed using splines. A plot of all three residue functions for each of the fits in Figure 4.10 is shown in Figure 4.12.

The start, end and area under the curve are of interest as they are the flux, flow and volume values. These three pieces of information provide the vast majority of the detail of the residue function as well as being key parameters of interest to physicians and scientists studying cancer and other aliments. The flux value, being the final value of the residue appears to be very close for all three models. This is encouraging as if all three models are arriving at roughly the same answer then it leads one to conclude that these estimates are appropriate for the data.

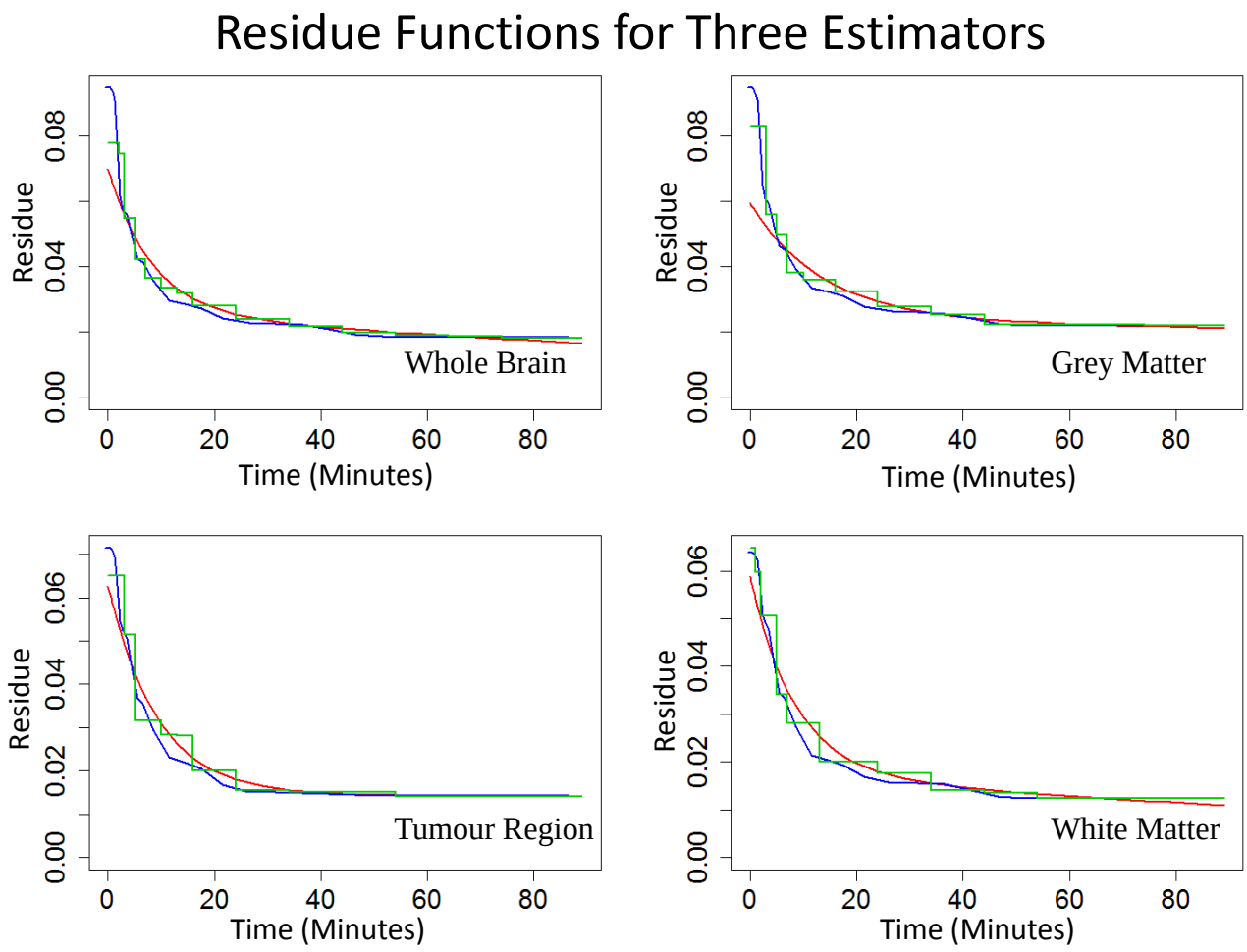


Figure 4.12: *Residue functions for compartmental model (red), nonparametric model (green) and mixture model (blue) fits shown in Figure 4.10. Again these are of different scales as they have been scaled by flow and the shape of these functions is our main concern.*

The area under the curve v_D are also similar for all three models with the early time points being notably different relative to the end points. As the early time bins are short the impact of this on the volume estimate is not as drastic as it would appear from the picture.

This results in very different estimates for flow being generated. This could be due to the fact that the compartmental model does not have sufficient degrees of freedom to capture the variability at early times in the data. By this we mean that the compartmental model may not have the flexibility to generate an appropriate flow value. Spectral analysis is a method which could be used here

which is based on a large number of compartments [14, 50]. Discounting this difference at early times then the residue functions are found to be very similar.

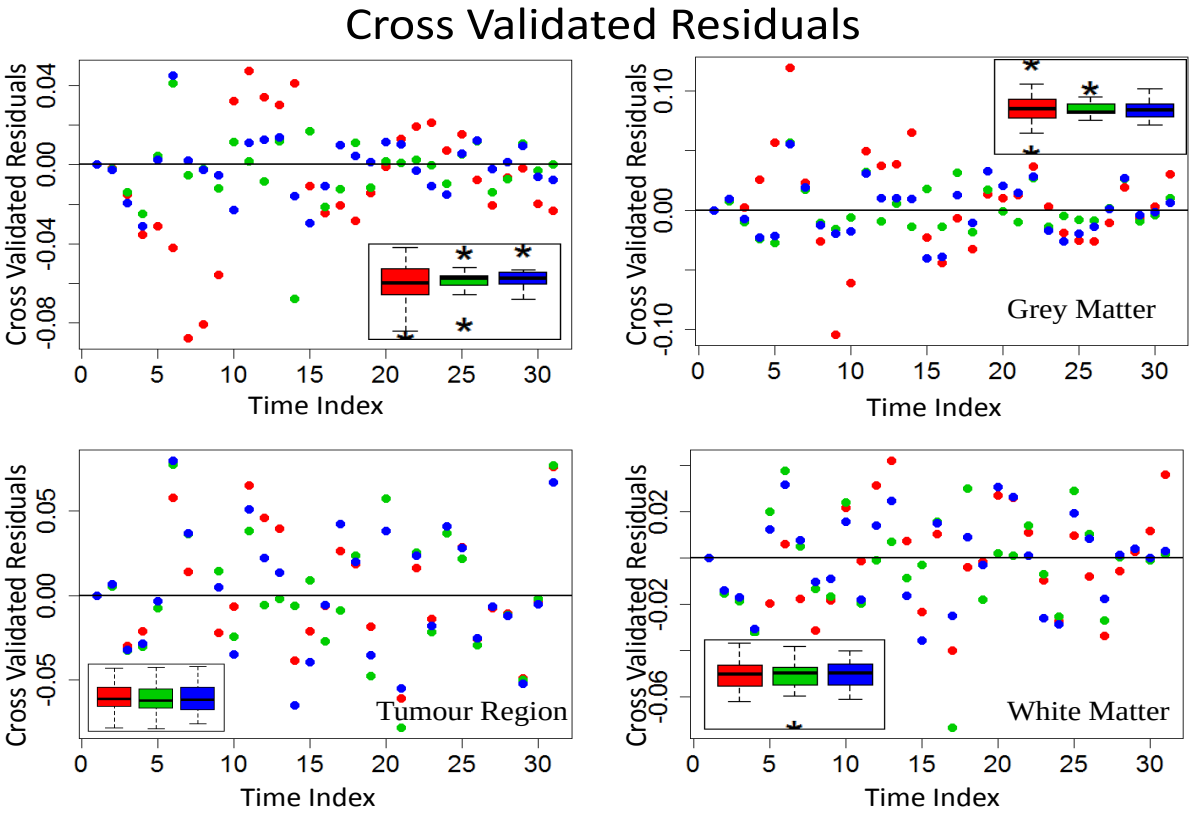


Figure 4.13: *Residue functions for compartmental model (red), nonparametric model (green) and mixture model (blue) fits shown in Figure 4.10.*

To simply stop here after just examining the residuals and residues would not be appropriate, when generating a statistical comparison between the models is clearly useful. The method which will be used here was presented in the last chapter and that is examining cross validated residuals. These cross validated residuals are shown in Figure 4.13.

At early times specifically between the fifth and tenth time point there appears to be some significant increase in the residuals for the compartmental model and again this is likely to be related to the AIF. This is particularly noticeable for the whole brain and grey matter regions but not as noticeable for the white matter and tumour regions. Similar increases in the residuals are present for the other models at these time points except in the case for the grey matter region but they are not as drastic. This is remarkably similar to the residuals on the data.

Model	Brain	Grey	Tumour	White
Compartmental	0.023	0.058	0.039	0.035
Nonparametric	0.008	**0.009**	**0.040**	0.014
Mixture	**0.006**	0.010	0.044	**0.010**

Table 4.1: *Cross validation residual sums of squares for the three models on the four regions. The lowest residual sums of squares for each region is highlighted in bold font.*

If one examines the box plots, the cross validated residuals appear quite similar for the tumour region and white matter region for all three models while the compartmental model has the most variability for the whole brain and grey matter regions.

It is also worth noting that all three models appear to have the variance of the cross validated residuals decreasing over time, aside from the larger variances at earlier time points, when the AIF is at its largest. Again this is very similar to the weighted residuals on the data. It is also the case that it is harder to see which model is performing better from the plot in the case of the white matter region as was the case for the residuals on the data.

The real question to ask when examining residuals is 'What do we do with these plots'? We seek to use these residuals to generate a metric to compare these models. This metric is simply the cross validated residuals sum of squares (CVRSS).

For the compartmental model, the nonparametric model and the mixture model the CVRSS for the whole brain, grey matter region, tumour region and white matter region are presented in Table 4.1. From this table, one sees that the nonparametric model performs best in two out of the four regions being studied. In the other two regions the mixture model does best. The compartmental model does worst in all regions except the tumour region where the mixture model does particularly badly in comparison to the other models.

As it is the case that in two of the regions the whole brain region, the white matter region the mixture model has the lowest CVRSS and in the grey matter and tumour regions, the nonparametric model has the lowest CVRSS, we have

an indication that we may find evidence against the compartmental model when analysing the larger dataset.

w

Another technique which was mentioned as a method in comparing models is the bootstrap and this has been discussed in Chapter 3. Here the compartmental model is assumed to be an appropriate fit to the data as it is the model which is widely used an accepted in the literature and a reference distribution for the improvement in fit statistic is generated by simulations. The test statistic derived on the real data compared to this distribution. An illustration of this was presented in the last chapter, which shows comparisons between the compartmental and nonparametric models. This analysis has been conducted to compare the compartmental and mixture models on a large subset of the data. Using this method p-values generated by this method for the whole brain, grey matter, tumour and white matter regions for this method are 0.11, 0.008, 0.58 and 0.2 respectively, which lead to the similar conclusions as the cross validation results. However as this method takes significant computer time (3 months on a high spec computer.) and it is not expected to produce any different results, it is omitted in the final analysis in this chapter.

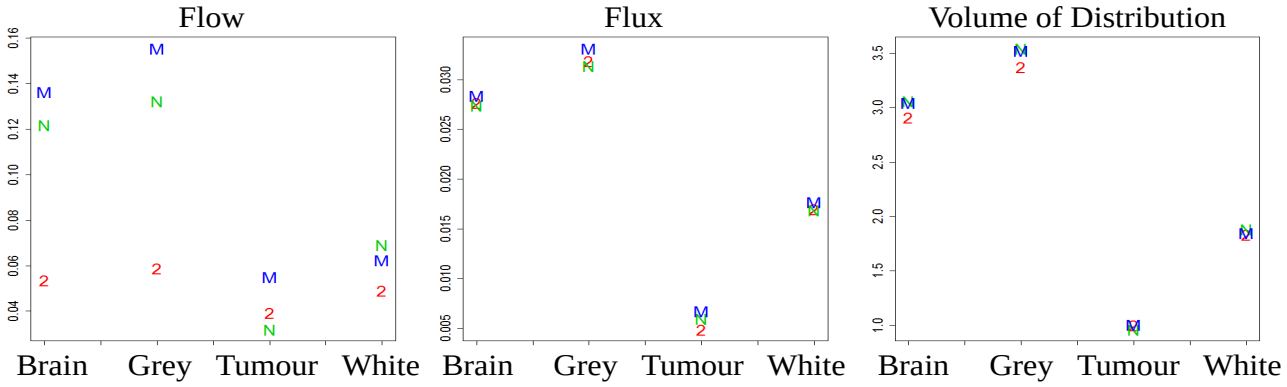


Figure 4.14: *Parameter estimates for flow, flux and volume in whole brain, grey, tumour and white matter regions using the two compartmental (2), nonparametric (N) and mixture models (M).*

A final comparison is drawn between the parameters estimated. This has been left until last as it is important to know which models are performing better before deciding which parameter estimates are appropriate. The difference between

parameters have already been mentioned in the context of studying the residues but knowing which models perform better due to having the cross validation results allows us to have a look at these parameters from a different perspective. The estimates for flow, flux and volume for the four regions being studied are presented in Figure 4.14.

As expected flux and volume values are very similar for all three models. However as the starting value in the residue is quite different for the whole brain and grey matter regions, then the flow estimates are very different. As these regions had the compartmental model having CVRSS an order of magnitude larger than the nonparametric and mixture models, this is not surprising. It would lead one to chose the nonparametric or mixture model estimate as the estimator of flow in this case. We now proceed to report the summary of results from analysing the large glioma dataset with the methods illustrated here.

4.6 Results

The previous section went into detail on analysing a dataset in the FDG glioma collection. It is now the case that we report the results of similar analyses being conducted on the whole dataset. The dataset is extremely large consisting of forty-two subjects. It seems even larger when you consider that every voxel in every subject was individually analysed as part of the mixture model approach, which is $128 \times 128 \times 35$ voxels $\times 31$ time points $\times$ 42 subjects. However the volume of work that this involves is worthwhile as interesting conclusions will be drawn from the dataset, which could not be drawn from one image alone. The larger data set will allow the trends in the parameters of interest for example be viewed and understood in ways not possible with a single dataset.

It is also worth highlighting that although they are not presented here kinetic maps for every subject have been generated and may prove valuable in the future.

As was the case in the previous section flux, flow and volume of distribution were calculated for each of the four regions being studied. These are compared in Figure 4.16, Figure 4.17 and Figure 4.18 for the compartmental model and the nonparametric model, for the compartmental model and the mixture model and for the mixture and nonparametric models respectively. The black line shown is

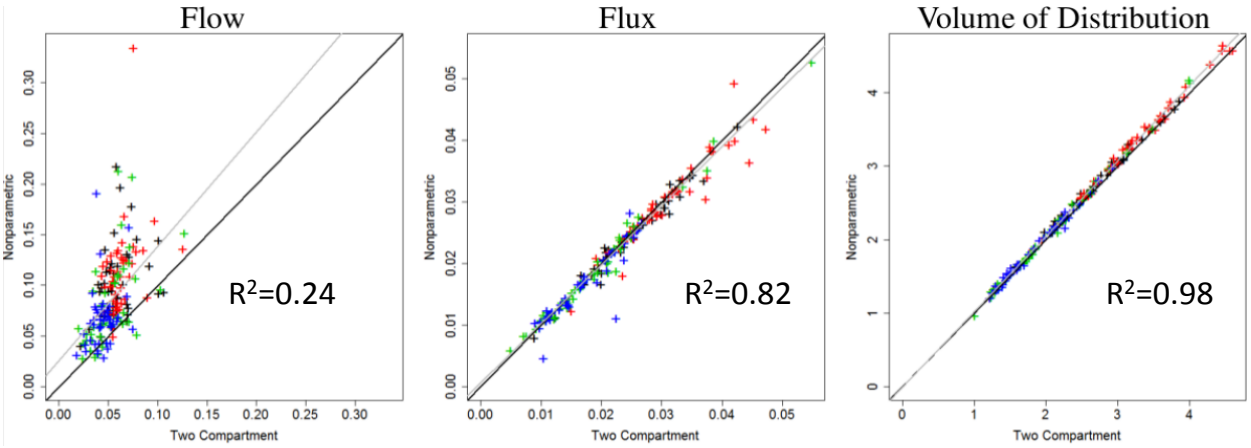


Figure 4.16: *Parameter estimates for flow, flux and volume in whole brain (black), grey (red), tumour (green) and white matter (blue) regions using both compartmental and nonparametric models and r-squared values.*

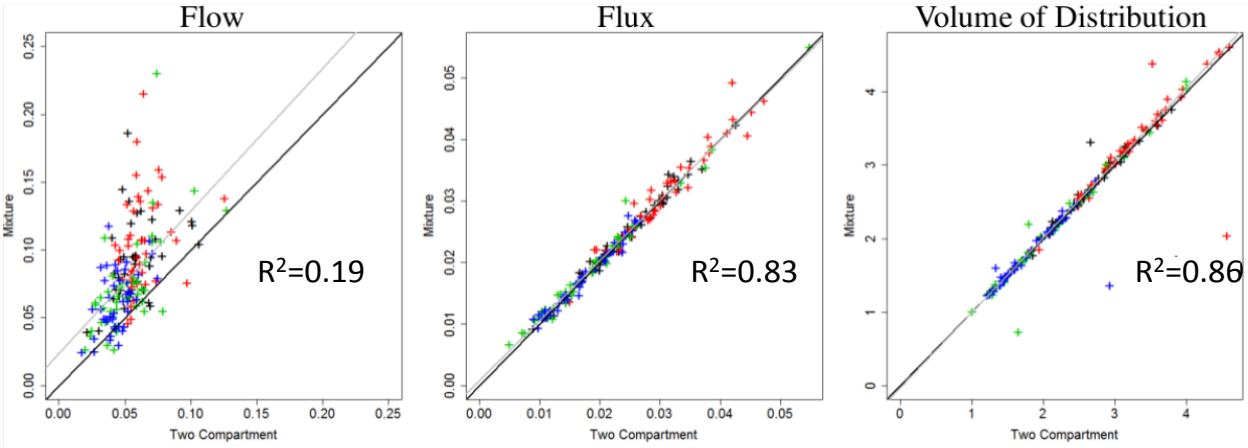


Figure 4.17: *Parameter estimates for flow, flux and volume in whole brain (black), grey (red), tumour (green) and white matter (blue) regions using the two compartmental and the mixture model and r-squared values.*

the line $y = x$ and the grey line is the linear regression line estimated by least squares. If the estimators were equal then these two lines would be identically equal. The R^2 values for the fit are also shown on this plot.

As in the case of the illustration, there is little difference in the case of flux and volume estimators for the two compartmental and nonparametric models. In fact the flux and volume estimates match the line so well that the line is almost invisible. However there is more variation between flow estimates and it does seem

that the nonparametric estimates for flow are higher than the compartmental model estimates. These results are consistent with the analysis of healthy subjects in [59]. It was seen in the illustration that the compartmental model estimator does not have as much flexibility as the nonparametric and mixture models, and tended to have a smaller estimate of flow. This is why the R^2 value for flow is so low. It is also worth highlighting that the estimated line appears almost parallel to the line $y = x$ although for flow the intercept appears greater than zero and the hypothesis that the slope is one is not rejected using the t test.

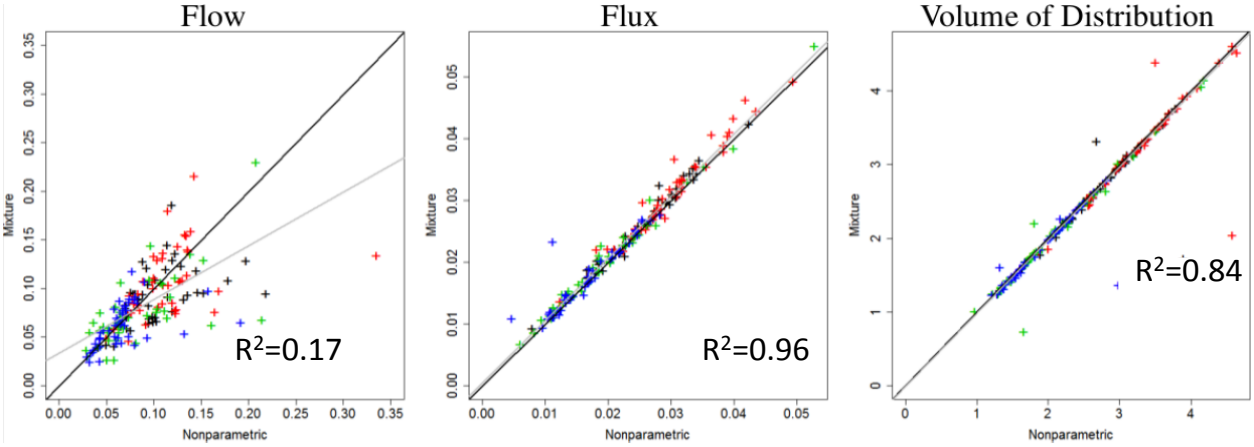


Figure 4.18: *Parameter estimates for flow, flux and volume in whole brain (black), grey (red), tumour (green) and white matter (blue) regions using the mixture model and nonparametric and r-squared values.*

Similarly, when examining the compartmental model and the mixture model. There is little difference in the case of flux and volume but there is more variation between flow estimates. In fact the stories are almost identical. Volume and flux estimates are very similar while flow has the mixture model generating larger estimates for flow as was the case with the nonparametric method. Similar to the nonparametric method the R^2 for flow is low and close to one for the flux and volume. The estimated lines tell a similar story to the previous case also.

Finally, when comparing the nonparametric and mixture models a similar picture is found for flux and volume, however in the case of flow there is again more variability. Here we look back at the illustration and Figure 4.12 and it is clear that although the flow values are closer together for two of the four regions, there are still clear differences between them in two of the four regions. It is also

the case here that the line fitted to the flow values is not parallel to the forty-five degree line and has a much smaller slope.

Using cross validation, residual sums of squares are calculated to compare the three models. These are presented in Figure 4.19. This figure shows the difference between the different models CVRSS for each individual ROI.

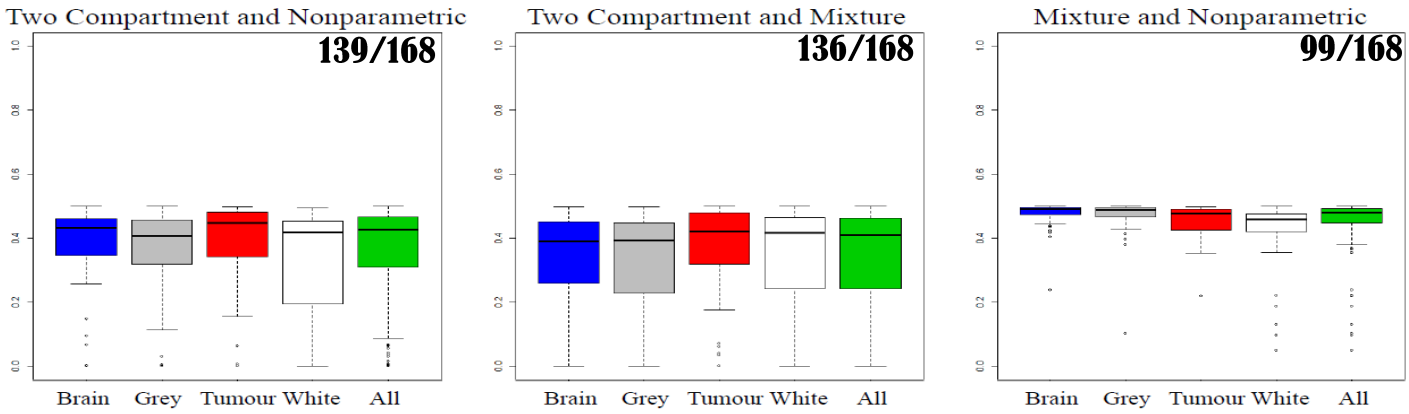


Figure 4.19: *Box plots comparing cross validation residual sums of squares for all three models. Blue represents the whole brain region, grey represents the grey matter region, red represents the tumour region, white represents the white matter region.*

In the case of comparing the nonparametric and compartmental models, the nonparametric model has a smaller CVRSS in forty-one, forty, thirty-one and twenty-seven of the forty-two regions for the brain, grey matter, tumour and white matter regions respectively. This totals to the nonparametric model outperforming the compartmental model in one hundred and thirty-nine of one hundred and sixty-eight regions. In the case of comparing the mixture and compartmental models, the mixture model has a smaller CVRSS in thirty-nine, thirty-nine, thirty-two and twenty-six of the forty-two regions for the brain, grey matter, tumour and white matter regions respectively. This totals to the mixture model outperforming the compartmental model in one hundred and thirty-six of one hundred and sixty-eight regions. In the case of comparing the mixture and nonparametric models, the nonparametric model has a smaller CVRSS in thirty-one,

Region	C-NP	C-M	M-NP
Brain	1 (<0.001)	3 (<0.001)	11 (0.003)
Grey	2 (<0.001)	3 (<0.001)	13 (0.019)
Tumour	11 (0.003)	10 (0.009)	21 (>0.01)
White	15 (0.088)	16 (0.160)	24 (0.441)

Table 4.2: *Number of regions where difference in the cross validation residual sums of squares for the models is smaller for the compartment model compared with the nonparametric model (first column), is smaller for the compartment model compared with the mixture model (second column) and is smaller for the mixture model compared with the nonparametric model (third column), P-values (two tailed) comparing these models to the binomial distribution with* $p = 0.5$ *are presented in brackets.*

twenty-nine, twenty-one and eighteen of the forty-two regions for the brain, grey matter, tumour and white matter regions respectively. This totals to the nonparametric model outperforming the nonparametric model in ninety-nine of one hundred and sixty-eight regions. This is summarised in the Table 4.2.

4.7 Discussion

This work has helped to demonstrate that the nonparametric and mixture models are viable alternatives to the compartmental model. It is also worth highlighting that this work was implemented using open source software tools such as AMIDE and R. Kinetic maps have been generated using the nonparametric method and an in depth analysis which has never previously been conducted on this dataset has been presented.

In carrying out kinetic analysis, the main goal is to separate flow and retention of tracer in tissue. One way that this is quantified is by estimating parameters which have some significance to flow and retention. Three key parameters are flow, flux and volume of distribution. It is clear from Figures 4.16, 4.17 and 4.18 that flow estimates are significantly different between the different models while the flux and volume of distribution estimates are quite similar. It is also worth

noting the comparison of the nonparametric models and the mixture models estimates for flow are also quite different. When differences occur it is necessary to decide on which model to chose for analysis.

The cross validation residuals lead us to significant evidence against the compartmental model in the case of both the mixture and nonparametric models. Similarly we find significant evidence against the mixture model in favour of the nonparametric model. However this is not as strong. This naturally leads to the question which model should be used for estimating parameters particularly in the case of flow. From the results in the previous section, the nonparametric model had the lowest cross validation residual sums of squares the majority of the time so it could be argued that this would be the best model to chose. However there are biases present in the model that were found in Chapter 3. An argument can be made that although these are statistically significant, they may be of little practical significance and therefore one can argue the nonparametric model works best on an ROI basis. In the case of FDG studies the most important parameter is often thought to be the flux and so an argument can be made that the choice of model is not important, however it is the authors opinion that the most correct model should always be used.

4.8 Conclusion

It is interesting that the flow parameter varies widely between models. One can assume that as the residue function is estimated by fitting a curve to data, then the entire time series has an effect on the curve. While the nonparametric and mixture models have a lot of flexibility, the compartmental model tend to have lower values at the start of the residue as the curve has only four parameters which can vary. The differences between the nonparametric and mixture models can probably be explained by a similar argument. The nonparametric has massive flexibility in estimating flow, while the mixture is constrained to be a linear combination of a relatively small number of basis elements. This perhaps leads to a question of whether the nonparametric method is overfitting the data due to the large number of parameters it has.

The flux values are quite similar. This was the case in the illustration earlier

in the chapter. The residue functions tend to level out towards the end of the scan and this is like a steady state solution in mathematics. As the model towards the end is the blood volume times the AIF plus the convolution of a curve which is flattening out and the AIF which is less spiked at later times, it stands to reason that the estimators will be similar here.

The volume of distribution is similar for all three estimators as well. Although the early time points for the compartmental model tends to be lower than the mixture and nonparametric models, this does not remain for long. Generally, the three models residues appear similar after just a few time points. As the opening time-bins have the smallest time duration, they do not have a massive impact on the area under the curve.

The compartmental model is outperformed by the nonparametric model. The mixture model also outperforms the compartmental model. It is not surprising that the compartmental model fares poorly in this work. If we rexamine the assumptions of the compartmental model, homogeneity, instant mixing and mass action, are clearly violated in the case of this dataset.

In the case of this study it does appear that the nonparametric model is outperforming the mixture model. This is not as clear cut as is the case with the compartmental model and the other two models but warrants some further investigations. It may be the case that since the nonparametric model is has more degrees of freedom, that it will always fit the data better. However the nonparametric model does outperform the other two models in the cross validation test and so one is lead to conclude that the nonparametric model is the appropriate model. It is highlighted that the nonparametric model did not outperform the mixture model by a massively large margin and to carry out more of this work on more data is left as future work.

In O'Sullivan *et al.* [59] an analysis similar to this is presented. Similar results for the parameter estimates were found in this case for healthy cerebral tissue. The results reported here also found evidence against the compartmental model. It is encouraging that both studies show similar results.

It is probably appropriate to mention that there are some downsides to the use of alternative methods such as the nonparametric and mixture models. Firstly, direct comparisson to historic studies completed with compartmental models is

not possible unless the historic data is analysed again. Secondly, although the new methods outperform the compartmental model in many ways, it does involve nonparametric regression, which many in the field of medical imaging will not be familiar with.

Chapter 5

Discussion, Conclusions and Future Work

5.1 Discussion and Conclusions

The work in this thesis is based on the analysis of dynamic PET region of interest time-course data. This has been studied by a variety of researchers over the years. Many different points of view have been taken. While many advocate for the compartmental models, their use in practice is questionable if not completely improper. However, this is seen frequently in the literature. Tests *in vitro* have shown these compartmental models to be appropriate but the case of tissue in a well stirred test tube and in a live subject (*in vivo*) are very different. This has largely motivated the efforts of this author, and those who published work in [14, 50, 59, 60].

In this work a Markov chain formulation was proposed to gain an insight into how PET radiotracers distribute throughout the body. In this formulation it was shown that in a limiting case where the number of tracer atoms was large the discrete Markov formulation approached the solution of the compartmental model differential equation compartmental models. The insights gained by this work and the work in [14, 50, 59, 60] lead us to consider alternative residues. This however in itself did not lead to any results which were particularly tangible.

We briefly mentioned the work of Cunningham and Jones [14] and Murase [50] which describe spectral analysis for residue estimation. Spectral analysis is

modelling the PET residue function by a large number of exponential functions. This can be thought of as a nonparametric exponential model or a many compartmental model. In their work, Murase and Cunningham and Jones show that it is possible to fit such a model to the data and suggest that this model has advantages over the compartmental model. These models have some nice properties. As the model is effectively a multi compartmental model, it has a residue function which is a survival curve, due to its components consisting of a mixture of exponentials. In the setup described in [14], there can be over two hundred variables to estimate. As a result of this, the optimisation is not simple. Cunningham and Jones propose using the simplex method for this as weighted least squares would likely be difficult to implement and not guarantee convergence. To put this in context the nonlinear least squares estimation for the two compartmental model consists of estimating six parameters and this is not trivial. Spectral analysis was in part the inspiration for the work contained in this thesis. Due to time constraints, this was never achieved in this project. Perhaps a limitation of this work is that this comparison has not been made.

In 2009, the case of examining alternative residue functions was rexamined by O'Sullivan *et al.* [60]. In this work, the authors proposed to estimate the probability density function for the transit time of the tracer in tissue instead of estimating the residue function directly. From this the residue was generated. This was implemented in a nonparametric form using cubic B splines to estimate the density. As the residue estimate is effectively one minus a cumulative density function it must be monotonic and start at one, the life table nature of the residue was preserved. A bootstrapping approach was used to compare the models, which was also used in this work to compare the compartmental and nonparametric models in Chapter 3. This work found strong evidence against the compartmental model in a dataset of ten distinct tissue regions of cerebral tissue in twelve normal subjects. Again the work in this thesis would be strengthened with this larger comparisson.

In 2012 an alternative nonparametric method was presented in Hawe *et al.* [28], which presented an analysis on a single PET dataset for the FDG and the ^{15}O water tracers. This analysis for FDG was the basis of Chapter 3 of this thesis. Firstly, this chapter shows that a nonparametric method for PET region

of interest time-course data is feasible and practical and in some ways is easier to implement than the compartmental model as the model is applied to data using the *orlm()* function in R, rather than an iterative least squares method, which can have issues with convergence. Secondly, Chapter 3 shows that the nonparametric model does not have the questionable assumptions underlying the compartmental model. In the example illustrated here, the nonparametric model outperformed the compartmental model. Another contribution of this chapter is the methods generated in comparing different non nested models using cross validation and the bootstrap.

A simulation study examining the efficacy and performance of the models was undertaken. It showed the convergence rates of both models to true parameters and the true residue. Convergence at a slower rate occurs for the nonparametric model. Small biases and estimates of variances for the parameters and the residue are found in perfect data.

Chapter 3 continues to examine the performance of the models in a simulation study. This study finds on perfect data that biases are small for parameter estimates for both models. The largest bias is one and a quarter percent for the nonparametric models estimate for flux. Rates of parameter convergence and convergence for the residue are also discussed. In almost all cases the compartmental model converges faster. This is hardly surprising as the simulated data is compartmental model data. It is worth noting that for volume of distribution, the convergence rates are very similar. A similar study which is based on the nonparametric residue was conducted and it was found that the nonparametric method outperforms the compartmental model here.

This model is new and has only been tested in brain tissues to date. The model cannot be said to be the correct model for the entire body until it has been tested in other regions. Also, the model is more complicated (more parameters) than the compartmental model so getting others in the medical imaging community to use it could be difficult. Also, previous analyses would all have to be redone with the nonparametric model to allow comparisons to historic data.

In the case of the examination of the extra sums of squares statistic, the traditional ESS statistic used in ANOVA does not follow the F-distribution. It was shown however that using the square root transformation, that the ESS

distribution looks similar to the F-distribution. For small numbers of simulations the p-values appear uniform. For the larger sample presented in Chapter 3, the p-values do not appear to follow a uniform distribution. This is due in part to the power of the Kolmogorov-Smirnov test. This result is interesting as it does show that the form of the statistic may not be the traditional one but there appears to be a relationship between these data and the F-distribution. However, to critique this work, it is necessary to highlight that this is of little practical use. In the absence of a rigorous closed form for the reference distribution, bootstrapping will be used to generate the reference distribution.

In 2014, O'Sullivan *et al.* [59] proposed a mixture model for the residue function and this was reviewed in Chapter 4. The basis of this was the nonparametric method proposed previously and a segmentation of the PET dataset. Again this model follows the rules associated with life table data. It also has the advantage of using the dataset itself to generate the residue function. It does have a drawback however, the modelling is more complicated than any of the other models presented in this thesis. It would also be a harder sell to the medical imaging community as splines are often not studied by those outside the field of statistics

The analysis of the large FDG dataset using these methods were applied to the glioma dataset. These data are analysed with the compartmental model, the nonparametric model and the mixture model. All three models are compared using cross validation. Strong evidence against the compartment model is found. In the case of O'Sullivan *et al.* [59], who conducted a similar analysis in normal subjects evidence against the compartmental model is found in favour of the other models. This is hardly surprising as these models do not have the assumptions of the compartmental model. The results suggest the nonparametric model is the best performing model, however comparing the nonparametric and mixture models is not so straightforward as the cross validated residuals are closer than in the case of the compartmental model. While the results here suggest the nonparametric model outperforms the mixture model it is not as strong a difference as was present when comparing the compartmental model and the other models so further study should be conducted to confirm this.

Finally, it seems appropriate to highlight that all the analyses in this work have been conducted using open source software. This is excellent as it means

that PET image analysis has been shown to be a field, which is accessible to any scientist who can source data to analyse with very little expense.

5.2 Future Work

The work in this thesis tells a story of a number of models, how they work, how they are implemented and applies them to data. There are however a few avenues which were not explored in this work.

Outside of the modelling perspective, there are other avenues open following this thesis. While this work was done exclusively with the FDG tracer, there are others of interest to the PET community. In Hawe *et al.* [28], a similar analysis adapted for the ^{15}O water tracer was illustrated. These methods could also be extended to other tracers which image receptor ligand interactions such as verapamil, dopamine or serotonin.

Another area that was mentioned on a few occasions in this thesis was spectral analysis. This was not implemented in this work. It would be interesting to add this model to the others used to estimate the PET residues and generate comparisons between this model and the compartmental, nonparametric and mixture models. An analysis such as this could be of interest to the PET literature in the context of reviewing the various models and highlighting which model works best. Similarly, the nonparametric model described in [60] should be analysed in a similar way.

Comparisons could be made between the mixture and compartmental and nonparametric models using the bootstrap. This was not done here as cross validation takes less time. Although it has not been presented in this volume, the bootstrapping comparison has been completed for all data comparing the compartmental and nonparametric methods. These results were very much in line with the cross validation comparison results. It remains as future work to carry out comparisons comparing the mixture model and the nonparametric model and the mixture model and the compartmental model using the bootstrap.

In Chapter 3 a simulation study on the nonparametric method is proposed. A study on the mixture model examining bias, variance and convergence of parameters could be of value to the medical imaging community. The study of all three

models convergence could result in a publication. Also of interest to the statistics community would be the software used to conduct these analyses, which could also be made available as part of an R package for modelling.

It is natural to consider the applications of this work outside of PET FDG studies. Their use could be extended to other applications such as MR and CT modelling where the flow of contrast agents and paramagnetics between blood and tissue are of interest. Dynamic MR scans frequently have an AIF present and much of the work here would be transferable to such analyses.

The work here can also be extended to other areas such as chemical reaction networks. I highlight that the work of Anderson and Kurtz [5] was directly applicable to PET data. The rate chemical reactions occur in systems where one chemical is introduced as a bolus would be particularly relevant. Also to other applications involving in pharmaceuticals such as chemical synthesis where first order ODEs are common. Additionally there are other applications of compartmental modelling such as epidemiology, where these techniques may be applicable.

Nomenclature

Δ **Delay**

v_D **Volume of Distribution**

2C **Two Compartmental Model**

AIF **Arterial Input Function**

AMIDE **A Medical Imaging Data Examiner**

ANOVA **Analysis of Variance**

CVRSS **Cross Validated Residuals Sums of Squares**

EEG **Electroencephalogram**

EKG **Electrocardiogram**

ESS **Extra Sums of Squares**

FDG **Fluorodeoxyglucose**

FDG-6-P **Fluorodeoxyglucose 6 Phosphate**

GE **General Electric**

ic.infer **R package**

K **Flow**

Kε **Flux**

LOR **Line of Response**

MM **Mixture Model**

MR **Magnetic Resonance**

MSE **Mean Squared Error**

nls() **Nonlinear Regression Package in R**

NP **Nonparametric**

PET **Positron Emission Tomography**

QQ **Quantile Quantile**

R **Statistical Software**

$R(t)$ **Residue Function**

ROI **Region of Interest**

RSS **Residual Sums of Squares**

SPECT **Single Photon Emission Computer Tomography**

SSReg **Sum of Squares for Regression**

TSS **Total Sums of Squares**

UCC **University College Cork**

UW **University of Washington**

Bibliography

[1] R.A. Adams. *Calculus: A Complete Course*. Addison-Wesley Longman, Incorporated, 2002.

[2] International Atomic Energy Agency. *Cyclotron produced radionuclides:guidance on facility design and production of [18F]fluorodeoxyglucose (FDG)*. IAEA Radioisotopes and radiopharmaceuticals series No. 3, Vienna, Austria., 2012.

[3] A. Alessio, E. Butterworth, J. Caldwell, and J.B. Bassingwaithe. Quantitative PET myocardial perfusion through integrated volume manipulation, display, and modeling. *Workshop on Multi-scale Modelling of the Heart*, 2008.

[4] Alliance Medical Scan Pricing. *http://www.alliancemedical.ie/our−pricing*. Accessed: 30-09-2014.

[5] D.F. Anderson, G. Craciun, and T.G. Hurtz. Product-form stationary distribution for deficiency zero chemical reaction networks. *Bulletin of Mathematical Biology*, 72(8):1947–1970, 2010.

[6] H. Anton and C. Rorres. *Elementary Linear Algebra (10th Edition)*. Wiley, 2010.

[7] AZ Tech Radiology. *http://www.aztechradiology.com/services/PET_Scanhtml*. Accessed: 18-06-2014.

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

[9] D. Bates and D.G. Watts. *Projection of the Residue function onto a basis of B-splines Nonlinear Regression Analysis and it's Applications*. Wiley, 1988.

[10] W.E. Boyce and R.C. DiPrima. *Elementary Differential Equations and Boundary Value Problems (8th Edition)*. Wiley, 2005.

[11] Cancer Center of Santa Barbara. *http://www.ccsb.org/upload/Image/Machines /PETScanner.jpg*. Accessed: 01-03-2014.

[12] J.-H. Choi, M.-J. Ahn, H.-C. Rhim, J.-W. Kim, G.-H. Lee, Y.-Y. Lee, and I.-S. Kim. Comparison of WHO and RECIST criteria for response in metastatic colorectal carcinoma. *Cancer Res Treat*, 37(5):290–293, 2005.

[13] D.R. Cox and H.D. Miller. *The Theory of Stochastic Processes*. Chapman and Hall/CRC, 1965.

[14] V.J. Cunningham and T. Jones. Spectral analysis of dynamic PET studies. *Journal of Cerebral Blood Flow and Metabolism*, 13:15–23, 2004.

[15] A. Berrington de Gonzlez, M. Mahesh, K.P. Kim, M. Bhargavan, R. Lewis, F. Mettler, and C. Land. Projected cancer risks from computed tomographic scans performed in the united states in 2007. *Arch. Intern. Med.*, 169 (22), 2009.

[16] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, 39:1–38, 1977.

[17] D.J.Brenner and E.J. Hall. Computed tomography an increasing source of radiation exposure. *N. Engl. J. Med.*, 357 (22), 2007.

[18] B. Efron and R.J. Tibshirani. An introduction to the bootstrap. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 158(2):339, 1995.

[19] E.A. Eisenhauer et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247, 2009.

[20] F.H. Fahey. Data acquisition in PET imaging. *Journal of Nuclear Medicine Technology*, 30(2):39–49, 2002.

[21] S.S. Gambhir, J. Czernin, J. Schwimmer, D.H.S. Silverman, R.E. Coleman, and M.E. Phelps. A tabulated summary of the FDG PET literature. *Journal of Nuclear Medicine*, 42(5 suppl):1S–93S, 200.

[22] S. Geisser. *Predictive Inference*. NY: Chapman and Hall, 1993.

[23] Get peace of mind not big hospital bills. *http://www.independent.ie/business/personal−finance/get−peace−of−mind−not−big−hospital−bills−26466731.html*. Accessed: 30-09-2014.

[24] K. Godfrey. *Compartmental Models and their Applications*. Academic Press, 1983.

[25] U. Gromping. Inference with linear equality and inequality constraints using R: The package ic.infer. *Journal of Statistical Software*, 33(10):1–31, 2010.

[26] R.N. Gunn, S.R. Gunn, and V.J. Cunningham. Positron emission tomography compartmental models. *Journal of Cerebral Blood Flow and Metabolism*, 21(6):635–652, 2001.

[27] R.N. Gunn, S.R. Gunn, F.E. Turkheimer, J.A. Aston, and V.J. Cunningham. Positron emission tomography compartmental models: a basis pursuit strategy for kinetic modeling. *Journal of Cerebral Blood Flow and Metabolism*, 22(12):1425–1439, 2002.

[28] D. Hawe, F.R. Hernandez-Fernandez, L. O'Suilleabhain, J. Huang, E. Wolsztynski, and F. O'Sullivan. Kinetic analysis of dynamic positron emission tomography data using open- source image processing and statistical inference tools. *WIRES Comput Stat*, 4:316–322, 2012.

[29] Health and Science The wonders of PET scan. *http://healthyscientist.blogspot.ie/2009/11/wonders−of−pet−scan.html*. Accessed: 30-09-2011.

[30] F.R. Hernandez-Fernandez, D. Hawe, J. Huang, M. Muzi, and F. O'Sullivan. Regional kinetic summaries of dynamic PET time-course data: Model the average or average the model? *IEEE conference of Medical Imaging*, Poster Presentation, 2011.

[31] R.J. Hicks. Beyond fdg: novel PET tracers for cancer imaging. *Cancer Imaging*, 4(1):22–24, 2004.

[32] A.R. Houghton and D. Gray. *Making Sense of the ECG, Third Edition.* Hodder Education, 2012.

[33] J. Huang and F. O'Sullivan. An analysis of whole body tracer kinetics in dynamic PET studies with application to image-based blood input function extraction. *IEEE Transactions on Medical Imaging*, PP(99):1–16, 2014.

[34] S.C Huang, M.E. Phelps, E.J. Hoffman, K. Sideris, C.J. Selin, and D.E. Kuhl. Noninvasive determination of cerebral metabolic rate of glucose in man. *Am J Physiol Endocrinol Metab*, 238:E69–E82, 1980.

[35] H.M. Hudson and R. Larkin. Accelerated em reconstruction using ordered subsets. *IEEE Transactions on Medical Imaging*, 13(4):601–609, 1994.

[36] R.B. Innis, V.J. Cunningham, J. Delforge, M. Fujita, A. Gjedde, R.N. Gunn, J. Holden, S. Houle, S.C. Huang, M. Ichise, Iida H, H. Ito, Y. Kimura, R.A. Koeppe, G.M. Knudsen, J. Knuuti, A.A. Lammertsma, M. Laruelle, J. Logan, R.P. Maguire, M.A. Mintun, E.D. Morris, R. Parsey, J.C. Price, M. Slifstein, V. Sossi, T. Suhara, J.R. Votaw, D.F. Wong, and R.E. Carson RE. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. *Journal of Cerebral Blood Flow and Metabolism.*, 27(9):1533–1539, 2007.

[37] Irish Health. *http://www.irishhealth.com/article.html?id=3912*. Accessed: 30-09-2014.

[38] E.L. Kaplan and P. Meier. Nonparametric estimation for incomplete observations. *JASA*, 53:457–481, 1958.

[39] S.S. Kety and C.F. Schmidt. The nitrous oxide method for the quantitative determination of cerebral blood flow in man: theory, procedure and normal values. *J Clinical Investigations*, 27(4):476–483, 1948.

[40] G. Koukourakis, G. Maravelis, S. Koukouraki, P. Padelakos, and V. Kouloulias. Overview of positron emission tomography chemistry: clinical and technical considerations and combination with computed tomography. *J BUON*, 14(4):575–580, 2009.

[41] E.R. Laws, I.F. Parney, W. Huang, F. Anderson, A.M. Morris, A. Asher, K.O. Lillehei, M. Bernstein, H. Brem, A. Sloan, M.S. Berger, S. Chang, and Glioma Outcomes Investigators. Survival following surgery and prognostic factors for recently diagnosed malignant glioma: data from the glioma outcomes project. *J Neurosurg.m*, 99(3):467–473, 2003.

[42] A.M. Loening and S.S. Gambhir. AMIDE: A free software tool for multimodality medical image analysis. *Molecular Imaging*, 2(3):131–137, 2003.

[43] J. Logan, J.S. Fowler, N.D. Volkow, Y-S. Ding G-J. Wang, and D.L. Alexoff. Distribution volume ratios without blood sampling from graphical analysis of PET data. *Journal of Cerebral Blood Flow and Metabolism*, 16(5):834–840, 1996.

[44] R.C. Semelka M.A. Brown. *MRI Basic principles and applications (4th Edition)*. Wiley-Blackwell, 2010.

[45] R. Maitra and F. O'Sullivan. Estimating the variability of reconstructed PET data: a technique based on approximating the reconstruction filter by a sum of gaussian kernels. *IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, 3:1411–1414, 1995.

[46] D.A. Mankoff, A.F. Shields, M.M. Graham, J.M. Link, J.F. Eary, and K.A. Krohn. Kinetic analysis of 2-[carbon-11]thymidine PET imaging studies: compartmental model and mathematical analysis. *JNM*, 39(6):1043–1055, 1998.

[47] P. Meier and K.L. Zierler. On the theory of the indicator-dilution method for measurement of blood flow and volume. *Journal of Applied Physiology*, 6(12):731–744, 1954.

[48] E.D. Morris, C.J. Endres, K.C. Schmidt, B.T. Christian, R.F. Muzic JR, and R.E. Fisher. Kinetic modeling in positron emission tomography. *in Emission Tomography: The Fundamentals of PET and SPECT, edited by M. Wernick and J. Aarsvold*, pages 499–540, 2004.

[49] M.Tubiana. Comment on computed tomography and radiation exposure. *N. Engl. J. Med.*, 358 (8), 2008.

[50] K. Murase. Spectral analysis: Principle and clinical applications. *Annals of Nuclear Medicine*, 17(6):427–434, 2003.

[51] M. Muzi, D.A. Mankoff, J.M. Link, S. Shoner, A.C. Collier, L. Sasongko, and J.D. Unadkat. Imaging of cyclosporine inhibition of p-glycoprotein activity using 11C-Verapamil in the brain: Studies of healthy humans. *Journal of Nuclear Medicine*, 50(8):1267–1275, 2009.

[52] M. Muzi, A. M. Spence, F. O'Sullivan, D. A. Mankoff, J. M. Wells, J. R. Grierson, J. M. Link, and K. A. Krohn. Kinetic analysis of 3'-deoxy-3'-18f-fluorothymidine in patients with gliomas. *Journal of Nuclear Medicine*, 47:1612–1621, 2006.

[53] E. Niedermeyer and F.L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincot Williams and Wilkins., 2004.

[54] J Nocedal and S.J. Wright. *Numerical Optimisation, 2nd ed.* Springer Publications, 2006.

[55] L. O'Suilleabhain, J. Huang, F. Hernandez, D. Hawe, E. Wolsztynski, M. Muzi, and F. O'Sullivan. Kinetic analysis of dynamic FDG and 15o-h2o PET studies by parametric and nonparametric methods: A statistical analysis. *Proceedings of the International Conference on Biomedical Engineering and Biotechnology*, 2012.

[56] F. O'Sullivan. Locally constrained mixture representation of dynamic imaging data from PET and MR studies. *Biostat*, 7:318–338, 2006.

[57] F. O'Sullivan and K. Roy Choudhury. An analysis of the role of positivity and mixture model constraints in poisson deconvolution problems. *Journal of Computational and Graphical Statistics*, 10(4):673–696, 2001.

[58] F. O'Sullivan, J. Kirrane, M. Muzi, J. O'Sullivan, A.M. Spence, D.A. Mankoff, and K.A. Krohn. Kinetic quantitation of cerebral PET-FDG studies without concurrent blood sampling: statistical recovery of the arterial input function. *IEEE Transactions on Medical Imaging*, 29(3):610–624, 2010.

[59] F. O'Sullivan, M. Muzi, D.A. Mankoff, J.F. Eary, A.M. Spence, and K.A. Krohn. Voxel-level mapping of tracer kinetics in pet studies: A statistical approach emphasizing tissue life tables1. *The Annals of Applied Statistics*, 8(2):1065–1094, 2014.

[60] F. O'Sullivan, M. Muzi, A.M. Spence, D.M. Mankoff, J.N. O'Sullivan, N. Fitzgerald, G.C. Newman, and K.A. Krohn. Nonparametric residue analysis of dynamic PET data with application to cerebral FDG studies in normals. *Journal of the American Statistical Association*, 104(486):556–571, 2009.

[61] F. O'Sullivan, E. Wolsztynski, J. O'Sullivan, T. Richards, E.U. Conrad, and J.F. Eary. A statistical modelling approach to the analysis of spatial patterns of FDG-PET uptake in human sarcoma. *IEEE Transactions on Medical Imaging*, 30(12):2059–2071, 2011.

[62] M.E. Phelps. *PET: Molecular Imaging and its Biological Applications (First Edition)*. Springer Publications, 2004.

[63] M.E. Phelps, S.C. Huang, E.J. Hoffman, C. Selin, L. Sokoloff, and D.E. Kuhl. Tomographic measurement of local cerebral glucose metabolic rate in humans with (f-18) 2-fluoro-2-deoxy-d-glucose: validation of method. *Annals of Neurology*, 6(5):371–388, 1979.

[64] Production and Synthesis of Radionuclides. *https://wiki.engr.illinois.edu/pages/viewpage.action?pageId=49744271*. Accessed: 30-09-2011.

[65] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[66] J.A. Rice. *Mathematical statistics and data analysis*. Cengage Learning, 2006.

[67] V.J. Robertson and K.G. Baker. A review of therapeutic ultrasound: effectiveness studies. *Phys Ther.*, 81(7):1339–1350, 2001.

[68] G.J.S. Ross. The efficient use of function minimisation in nonlinear maximum likelihood estimation. *Applied Statistics.*, 19:205–221, 1970.

[69] R. Smith-Bindman, J. Lipson, R. Marcus, K.P. Kim, M. Mahesh, R. Gould, A. Berrington de Gonzlez, and D.L. Miglioretti. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Arch. Intern. Med.*, 169 (22), 2009.

[70] L. Sokoloff, M. Reivich, C. Kennedy, M.H. Rosiers, C.S. Patlak, K.D. Pettigrew, O. Sakurada, and M. Shinohara. The [14C] deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *Journal of Neurochemistry*, 28(5):897–916, 1977.

[71] A.M. Spence, M. Muzi, M.M. Graham, F. O'Sullivan, K.A. Krohn, J.M. Link, T.K. Lewellen, B. Lewellen, S.D. Freeman, M.S. Berger, and G.A. Ojemann. Glucose metabolism in human malignant gliomas measured quantitatively with PET, 1-[C-11]glucose and FDG: analysis of the FDG lumped constant. *Journal of Nuclear Medicine*, 39(3):440–448, 1998.

[72] P.K. Spiegel. The first clinical x-ray made in america-100 years. *AJR*, 164(1):241–243, 1995.

[73] W.H. Sweet and G.L. Brownell. Localization of brain tumors with positron emitters. *Nucleonics*, 11:40–45, 1953.

[74] A. Tamir. *Applications of Markov Chains in Chemical Engineering*. Elsevier Publications, 1998.

[75] A. Taylor, D.M. Schuster, and N. Naomi Alazraki. *A Clinicians' Guide to Nuclear Medicine*. Society of Nuclear Medicine, 2000.

[76] M.M. Ter-Pogossian, M.E. Phelps, E.J. Hoffman, and N.A. Mullani. A positron-emission transaxial tomograph for nuclear imaging (PET). *Radiology*, 114(1):89–98, 1975.

[77] Y. Vardi, L.A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *Journal of the American Statistical Association.*, 80(389), 1985.

[78] E.H. Wacholtz. History and development of PET. *CE Websource.com, Technical Report*, 2012.

[79] A.D. Waxman, K. Herholz, D.H. Lewis, P. Herscovitch, S. Minoshima, M. Ichise, A.E. Drzezga, and Srand J.M. Mountz M.D. Devous. Society of nuclear medicine procedure guideline for fdg pet brain imaging. *Procedure Guideline for FDG-PET Brain Imaging v1.0*, 1:1–12, 2009.

[80] S. Weisberg. *Applied Linear Regression*. Wiley Series in Probability and Statistics, 2005.

[81] P.S. Talbot W.G. Frankle, M. Slifstein and M. Laruelle. A review on brain-imaging applications of SPECT - neuroreceptor imaging in psychiatry: Theory and applications. *International Review of Neurobiology*, 67:385–440, 2005.

[82] H. Zaidi. *Quantitative analysis in nuclear medicine imaging*. Springer Verlag, 2006.