

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA



**Incorporating accurate statistical modeling in PET
reconstruction for whole-body imaging**

Luis Manuel de Almeida Soares Janeiro

Thesis supervised by Professor Doutor Pedro Almeida, Faculdade de Ciências da Universidade de Lisboa, and by Doctor Claude Comtat, Service Hospitalier Frédéric Joliot, C.E.A., Orsay, France

Doutoramento em Biofísica

2007

Para a minha família

TABLE OF CONTENTS

Table of contents	i
Acknowledgments	v
Abstract	vii
Resumo	ix
Abbreviations	xv
 Introduction	 1
 1 – An introduction to Positron Emission Tomography (PET)	 5
1.1 – A brief history of PET	6
1.2 – Positron Emission Tomography overview	8
1.2.1 – Physical constraints and the three-gamma annihilation	10
1.2.1.1 – Positron range	10
1.2.1.2 – Annihilation photons non-collinearity	12
1.2.1.3 – Three photons positron annihilation	13
1.2.2 – Signal detection in PET	14
1.2.2.1 – Detectors	14
1.2.2.2 – Photomultiplier Tubes (PMTs)	17
1.2.2.3 –Block detector readout	18
1.3 – Data acquisition	19
1.3.1 – 2D and 3D mode for data acquisition	20
1.3.2 – Radial sampling	21
1.3.3 – The Michelogram	22
1.3.4 – Sinograms and projections	24
1.3.5 – Detected events in PET	26
1.4 – Data corrections	28
1.4.1 – Attenuation correction	30
1.4.2 – Correction for random coincidences	32
1.4.3 – Dead time correction	33
1.4.4 – Normalization	34
1.4.4.1 – Effects contributing to differences in the sensitivity	35
1.4.4.2 – Normalization procedures	36

1.4.5 – Scatter correction	37
1.5 – PET camera performance	40
1.5.1 – Spatial resolution	40
1.5.2 – Energy resolution	43
1.5.3 – Sensitivity	44
1.5.4 – Noise Equivalent Count (NEC)	45
2 – Image reconstruction algorithms in PET	47
2.1 – The Radon transform in PET	48
2.2 – A primary distinction between reconstruction algorithms	50
2.3 – Analytical reconstruction	52
2.3.1 – Direct Fourier Methods	53
2.3.2 – Filtered Backprojection	55
2.3.3 – The 3DRP algorithm	58
2.4 – Algebraic reconstruction	61
2.4.1 – Five components of an algebraic reconstruction method	64
2.4.1.1 – A finite parameterization of the image	64
2.4.1.2 – System matrix	65
2.4.1.3 – A model of the measurement uncertainty	67
2.4.1.4 – Objective function	70
2.4.1.5 – Numerical algorithm	72
2.4.2 – Algebraic Reconstruction Technique (ART)	74
2.4.3 – Maximum Likelihood – Expectation Maximization (ML-EM)	76
2.4.4 – Ordered Subsets – Expectation Maximization (OSEM)	81
2.4.4.1 – Block-iterative methods	85
2.4.4.2 – Raw-action methods	86
2.4.5 – Bayesian reconstruction	87
2.5 – Rebinning algorithms	92
2.5.1 – FORE (Fourier Rebinning)	93
2.5.2 – Other rebinning techniques	96
3 – The weighted OSEM approach and the adopted implementation	98
3.1 – Weighted OSEM	99
3.1.1 – NEC transform and NEC-OSEM	102
3.2 – The effects of FORE on the data variance	104
3.3 – The adopted OSEM implementation	105

4 – Numerical observer studies comparing different OSEM approaches for 3D whole-body PET	110
4.1 – Data simulation	111
4.1.1 – The anthropomorphic phantom	111
4.1.2 – The analytical simulator (ASIM)	112
4.1.3 – Simulation conditions	114
4.1.3.1 – Calculation of the slice variance reduction factors	119
4.2 – Image reconstruction	120
4.3 – Image quality assessment	1239
4.3.1 – NPWMF observer study	123
4.3.2 – Study of contrast versus noise	125
4.3.3 – Methodology to fix the amount of post-smoothing and the number of iterations	125
4.4 – Results	126
4.4.1 – Preliminary results for fixing the amount of post-smoothing and the number of iterations	126
4.4.2 – Study of contrast versus noise	130
4.4.3 – NPWMF observer study	132
4.5 – Discussion	136
 5 – Comparing NEC-based OSEM with other weighted OSEM approaches for clinical 3D whole-body PET imaging	 139
5.1 – Data acquisition and description	141
5.2 – Data pre-processing	141
5.3 – Data corrections and the estimation of the mean and the variance	144
5.3.1 – The estimation of the mean	145
5.3.2 – The estimation of the variance	147
5.4 – FORE and the weighting factors	152
5.5 – Image reconstruction and final results	156
5.5.1 – Image reconstruction	156
5.5.2 – Reconstructed images	157
5.6 – Discussion and conclusions	161
 Final discussion and conclusion	 163
 Appendix	
A.1 – Coordinate systems in 2D and 3D	171

A.1.1 – 2D case	171
A.1.2 – 3D case	171
A.2 – The Radon Transform	174
A.3 – The Backprojection Operator	176
A.4 – Central Section Theorem	177
A.5 – The Likelihood Function	180
A.6 – A reconstruction techniques diagram	182
A.7 – The Non-Prewhitening Matched Filter (NPWMF) numerical observer	183
A.8 – The EXACT HR+ PET scanner	185
List of figures	189
List of tables	193
Bibliography	195

Acknowledgments

My first word goes to the teaching staff of the Institute of Biophysics and Biomedical Engineering (IBEB), Faculty of Sciences of the University of Lisbon. Prof. Pedro Miranda, Prof. Eduardo Ducla Soares and Prof. Alexandre Andrade have always been generous, attentive, patient and I have always had their helpful comments and suggestions whenever they were pertinent. Their doors were always opened and this gave me additional motivation and responsibility. I also do not forget their personal support and how important it was.

I am also grateful to all my colleagues of the IBEB, for the mood within the Institute. This is priceless and makes IBEB a perfect place for working and studying. In special, I would like to thank Nuno Matela and Mónica Martins for all the work we have developed together. We began together and we had many fruitful discussions along the journey. Technical discussions, of course, but not solely... I thank them for their support, their comments, their generosity, their altruism and, specially, for their personal support. I felt it.

My gratitude goes also to the administrative staff of the Institute. Beatriz and Ana were always there. They know each one of us, and have a special way to handle, and solve, our problems. For those who are abroad, this is, again, priceless... and most of the times they were the first persons I met when I was back to the Institute. Their support was enormous.

I would also like to acknowledge my colleagues and all the persons with whom I had the privilege to work with at the Service Hospitalier Frédéric Joliot (SHFJ), namely Doctor Régine Trébossen, Charles Pautrot, Frédéric Bataille and Sébastien Jan, for their comments and helpful suggestions. I am specially in debt to Doctor Maria João Ribeiro, for all the support and for the kindness of having made possible the study we have done with real clinical data. Her expertness in the field was essential.

For all the comments, insights and shared knowledge, I am also in debt to Prof. Paul Kinahan, of the University of Washington, in Seattle. My gratitude to Prof. Paul Kinahan is extended for having accepted, and partially financed, my stay in Seattle, within the Imaging Research Laboratory. I felt home many thousand miles away and I met an extraordinary group of persons. Thank you to Adam Alessio for the fruitful discussions and his very clear way of exposing ideas.

My gratitude to Doctor Fernando Tomé and Doctor Maria Leonor Arsénio Nunes is enormous: for their personal support, their experience, their knowledge and culture. In Paris their presence was essential and I grew up with their example.

Doctor Claude Comtat was my supervisor in Orsay. I have no words to acknowledge him for everything. From the timeless discussions in Paris, no matter where – and many times dans “une terrasse”! - to the complete support at the SHFJ. Their critical comments, suggestions, perspectives and solutions have always astonished me. And still do. Many things I’ve learned, I’ve learned from him. I still keep the schematics he drew to make things easier, and everything matches. I remember to rewrite some texts many times, because there was always a better and clear way to put the things. Fortunately, I also keep some old versions, so I can compare. He was perfectly right! In any case, I have always believed that Claude would be there in case I would be unable to find the solution. And I was right.

He also introduced me many thing in Paris. We shared opinions about many, many different things. He often told me: “take your time”. And how necessary it was! From a personal point of view, I had his entire support. My debt to Doctor Claude Comtat is complete.

Finally, and most of all, I would like to thank Doctor Pedro Almeida. Once was the Master, now is the PhD. He has always been my supervisor. In every situation I have always felt his support. The most important thing, nevertheless, was the motivation he gave me. It was fundamental and, probably, the key for concluding all this work. He knew exactly what I was doing at any time, and helped me to solve each and every problem I had. I am not only referring to technical problems. I am, above all, referring to other more essential questions we have deeply discussed. I believe that, for reaching the end, technical and scientific hindrances pose minor problems and difficulties.

Fortunately, or in consequence, our friendship goes very beyond the scope of this work. I admire his example, and I hope to follow it.

This PhD work was made possible by the support of the fellowship BD/SFRH/4989/2001 from Fundação para a Ciência e Tecnologia (FCT). I would like to clearly state my gratitude to FCT.

Abstract

The thesis is devoted to image reconstruction in 3D whole-body PET imaging.

OSEM (“*Ordered Subsets Expectation maximization*”) is a statistical algorithm that assumes Poisson data. However, corrections for physical effects (attenuation, scattered and random coincidences) and detector efficiency remove the Poisson characteristics of these data. The Fourier Rebinning (FORE), that combines 3D imaging with fast 2D reconstructions, requires corrected data. Thus, if it will be used or whenever data are corrected prior to OSEM, the need to restore the Poisson-like characteristics is present.

Restoring Poisson-like data, i.e., making the variance equal to the mean, was achieved through the use of weighted OSEM algorithms. One of them is the NECOSEM, relying on the NEC weighting transformation. The distinctive feature of this algorithm is the NEC multiplicative factor, defined as the ratio between the mean and the variance. With real clinical data this is critical, since there is only one value collected for each bin – the data value itself. For simulated data, if we keep track of the values for these two statistical moments, the exact values for the NEC weights can be calculated.

We have compared the performance of five different weighted algorithms (FORE+AWOSEM, FORE+NECOSEM, ANWOSEM3D, SPOSEM3D and NECOSEM3D) on the basis of tumor detectability.

The comparison was done for simulated and clinical data. In the former case an analytical simulator was used. This is the ideal situation, since all the weighting factors can be exactly determined. For comparing the performance of the algorithms, we used the Non-Prewhitening Matched Filter (NPWMF) numerical observer.

With some knowledge obtained from the simulation study we proceeded to the reconstruction of clinical data. In that case, it was necessary to devise a strategy for estimating the NEC weighting factors. The comparison between reconstructed images was done by a physician largely familiar with whole-body PET imaging.

KEYWORDS: Positron Emission Tomography (PET), statistical image reconstruction, NEC (“*Noise Equivalent Count*”) Transformation, weighted OSEM (“*Ordered Subsets Expectation Maximization*”), tumor detectability.

Resumo

O trabalho de investigação subjacente à presente centra-se no estudo de alguns algoritmos de reconstrução de imagem, utilizados em Tomografia por Emissão de Positrões (PET)¹.

De entre as duas abordagens possíveis, são utilizados algoritmos algébricos², por serem aqueles que permitem não só modelar o processo de emissão-deteção da radiação, como, também, incorporar na própria estrutura do algoritmo informação relativamente à natureza estatística dos dados³ a serem reconstruídos, o que no caso vertente é fundamental.

De entre os algoritmos algébricos estatísticos⁴, adoptámos os OSEM (*Ordered Subsets Expectation Maximization*), uma vez que se encontra já amplamente difundido na prática clínica.

Porém, uma das premissas intrínseca ao OSEM é assumir que os dados a reconstruir obedecem a uma estatística de Poisson. Ora, em muitos casos, verifica-se que isto não é verdade, ou seja, que o OSEM procede à reconstrução de imagem a partir de dados que não satisfazem aquele requisito. Esta é, então, a questão central da presente tese: proceder à reconstrução de imagem relativa a exames PET de corpo inteiro, baseada no algoritmo OSEM e preservando, na medida do possível, o modelo estatístico que lhe subjaz, isto é, dados distribuídos segundo uma estatística de Poisson.

A principal razão para que os dados deixem de seguir uma estatística de Poisson prende-se com as correcções a que são sujeitos, depois de serem organizados na forma de projecções ou sinogramas, mas antes de se proceder à sua reconstrução. Estas pretendem compensar a existência de determinados efeitos físicos (como a atenuação, dispersão de Compton ou existência de coincidências fortuitas) e geométricos (como sejam os que decorrem das dimensões finitas do *scanner* ou da sua própria geometria, corrigidos naquilo a que se chama “normalização”), que fazem com que nem todos os fotões detectados em coincidência

¹ Optamos por preservar o acrónimo em língua inglesa, por se aquele que, na prática corrente, se utiliza também em Portugal.

² A alternativa seria utilizar algoritmos analíticos de reconstrução, como por exemplo a Retro projecção Filtrada (FBP).

³ Quando nos referimos a “dados” a reconstruir, estamos a considerar os valores obtidos nas diferentes projecções (sinogramas), segundo linhas-de-resposta (LORs) com orientações várias, e que constituem a informação a partir da qual se procede à reconstrução tomográfica.

⁴ O ART é um exemplo de um algoritmo algébrico não estatístico.

correspondam a fótons emitidos na mesma aniquilação electrão-positrão, e cujas trajectórias façam entre si um ângulo de, aproximadamente, 180° .

Em princípio, o facto de os algoritmos algébricos permitirem modelar o processo de emissão e detecção de fótons, torna legítimo presumir que as correcções acima referidas podem ser incorporadas no processo de reconstrução, e feitas em simultâneo com este. Levantam-se, porém, duas questões, uma das quais, de fundo.

Em primeiro lugar, considerando que estamos a trabalhar com dados 3D, isto é, obtidos não apenas para LORs segundo planos transaxiais⁵, surge a necessidade de proceder a reconstruções rápidas destes dados. Ora, mesmo com os computadores actuais, isto revela-se uma tarefa computacionalmente exigente e demorada, se se optar por fazer a reconstrução directamente em 3D. A alternativa é efectuar uma operação de *rebinning*⁶ e proceder à reconstrução em 2D, com o correspondente ganho em termos de velocidade de reconstrução. Um dos algoritmos de *rebinning* que melhor desempenho apresenta, nomeadamente no compromisso entre resultado obtido e complexidade de implementação, é o denominado *Fourier Rebinning* (FORE). Porém, para que possa ser utilizado com os dados 3D, estes têm de ser previamente corrigidos relativamente aos efeitos referidos atrás. Ou seja: uma vez que é inevitável que os dados sejam previamente corrigidos e as correcções fazem com que aqueles deixem de apresentar uma estatística de Poisson, então a utilização do algoritmo FORE resulta, de forma incontornável, na obtenção de dados (2D) não Poisson e, portanto, não conformes com a premissa do OSEM.

Em segundo lugar, mesmo quando se procede a reconstruções directamente em 3D (conhecidas por “*fully 3D reconstructions*”), na maioria dos casos (na prática clínica) os dados são também corrigidos para compensar os efeitos da atenuação, dispersão de Compton, coincidências fortuitas, bem como normalizados.

Isto significa que, em ambos os casos (seja por obrigação decorrente do uso do FORE, seja pela prática habitual), é preciso arranjar uma estratégia para devolver aos dados as características de uma distribuição de Poisson. Isto significa arranjar uma maneira de fazer com que os dados corrigidos voltem a ter valores médios iguais à variância (característica essencial da distribuição de Poisson).

A utilização de versões do OSEM denominadas OSEM ponderado (“*weighted OSEM*”), designação que advém do facto de os dados corrigidos serem multiplicados e/ou adicionados por factores de ponderação, abrem a possibilidade de, após a referida ponderação, os dados a

⁵ Permitindo ter anéis não contíguos em coincidência.

⁶ Operação que consiste em rearranjar os dados 3D de modo a transformá-los em dados 2D, tentando preservar ao máximo as vantagens da aquisição 3D (nomeadamente relativamente à sensibilidade).

reconstruir voltarem a ter valores médios iguais à variância (em cada bin). No entanto, se a possibilidade existe, nem todas as versões do OSEM ponderado resultam do mesmo modo. Depende do factor de ponderação usado. No caso de se recorrer à transformada NEC (*“Noise Equivalent Count”*) e se tomarem como factores de ponderação os coeficientes envolvidos nesta transformada, os dados assim ponderados apresentam, de facto, em cada bin, um valor médio igual à variância. Ou seja, em primeira aproximação, voltam a exhibir as características de uma distribuição de Poisson. Para tanto, o factor multiplicativo NEC, para um determinado bin, é definido do seguinte modo:

$$NEC_i = \frac{\langle y_i \rangle}{\sigma^2(y_i)}$$

em que $\langle y_i \rangle$ e $\sigma^2(y_i)$ são, respectivamente, o valor médio e a variância desse bin. Ao OSEM ponderado que recorre aos coeficientes da transformada NEC chamamos NECOSEM. Outros algoritmos existem que recorrem a ponderações que também transformam os dados com o mesmo resultado, como é o caso do SPOSEM3D.

Apresentado que foi o contexto teórico em que assenta o trabalho de investigação a que a presente tese se reporta, importa agora descrever sucintamente o que foi feito e quais os objectivos que nos propusemos atingir.

Na prática, se pudéssemos resumir a maior parte do trabalho que fizemos, poderíamos arriscar fazê-lo dizendo que comparámos exaustivamente o desempenho de várias versões ponderadas do algoritmo OSEM, ao reconstruir dados provenientes de aquisições 3D, com vista à obtenção de imagens de corpo inteiro.

Em função do exposto nos parágrafos anteriores, e independentemente do processo de obtenção dos dados 3D, as comparações de desempenho efectuadas têm por base duas perspectivas diferentes. A primeira separa reconstruções FORE seguidas de OSEM 2D (tendo sido usados os algoritmos FORE+AWOSEM e FORE+NECOSEM) de reconstruções directas em 3D (ANWSESEM3D, SPOSEM3D e NECOSEM3D). Na segunda perspectiva, a distinção é feita entre algoritmos cuja ponderação devolve aos dados as características de Poisson⁷ (FORE+NECOSEM, SPOSEM3D e NECOSEM3D), e aqueles em que isso não ocorre, ou seja, em que pese embora os valores dos dois momentos fiquem mais próximos⁸, não são, de facto, iguais (FORE+AWOSEM e ANWSESEM3D).

⁷ Independentemente de serem reconstruções 2D ou 3D.

⁸ Uma vez que a ponderação leva em consideração a atenuação, que é o factor que mais contribui para o desvio da estatística de Poisson.

Quanto à proveniência dos dados 3D, foram efectuados dois estudos complementares: um em que se usaram dados simuladas; e um outro, após aquele, em que se reconstruíram dados reais obtidos numa das câmaras PET HR+ do S.H.F.J..

Quanto ao primeiro, e mais exaustivo, recorreu-se ao simulador analítico ASIM e a um fantôma numérico para gerar os dados necessários⁹. As potencialidades do simulador, as características do fantôma, bem como as dimensões dos tumores a ele adicionados e respectivas actividades, haviam sido previamente validadas em estudos envolvendo características muito semelhantes às do nosso próprio estudo [Lartizien, 03; Lartizien, 01].

Uma vez que este simulador permite guardar informação relativamente à média e à variância em cada bin, os coeficientes NEC podem ser calculados exactamente. Analogamente, ao ser possível simular a atenuação, conhecemos também os valores exactos dos coeficientes de atenuação (ACFs) a partir dos quais se estabelecem os factores de ponderação quer no AWOSEM quer no ANWOSEM3D. Isto é, do ponto de vista do OSEM ponderado, estamos nas condições ideais: é possível determinar exactamente os factores de ponderação dos dados.

Como algoritmos de reconstrução foram usadas as seguintes versões do OSEM ponderado: FORE+AWOSEM e FORE+NECOSEM, no que diz respeito a reconstruções 2D; e ANWOSEM3D, SPOSEM3D e NECOSEM3D, quanto a reconstruções 3D (“*fully 3D*”).

As imagens reconstruídas sofreram ainda uma filtragem adicional com um filtro gaussiano.

Uma vez que a quantidade de informação gerada e utilizada neste estudo foi imensa, era impossível que a avaliação da qualidade de imagem fosse feita por observadores humanos. A solução encontrada foi, portanto, recorrer a um observador numérico cujo desempenho, para a tarefa clínica e tipo de exames em causa, fosse comprovadamente semelhante ao dos observadores humanos. A escolha recaiu no observador numérico “*Non-Prewhitening Matched Filter*” (NPWMF), cuja adequação para este tipo de estudos tinha sido previamente verificada [Lartizien, 04].

Relativamente a resultados, poderíamos resumi-los do seguinte modo: em primeiro lugar, usando o observador NPWMF, foi possível constatar a existência de um valor óptimo para a largura a meia altura do filtro gaussiano de pós-filtragem da imagem reconstruída. Esse valor óptimo não é exactamente o mesmo para todos os órgãos estudados, mas 5.5 mm é o melhor compromisso. Ao invés, este valor óptimo não se verifica quando se estuda a variação no índice de detectabilidade do NPWMF em função do número de iterações. Logo, para estabelecer uma

⁹ O volume de dados que foi necessário gerar tornava impraticável a utilização de um simulador de Monte Carlo.

equivalência entre a qualidade das imagens reconstruídas pelos diferentes algoritmos usados, foi necessário recorrer a uma outra figura-de-mérito (FOM). Essa foi a relação contraste-ruído. Assim, tendo-se tomado como referência a imagem reconstruída, após 4 iterações, pelo FORE+AWSOEM - algoritmo e número de iterações usados com frequência na prática clínica – foi verificar-se, para os outros algoritmos e usando a relação contraste ruído, para quantas iterações se obtinham valores idênticos desta FOM. A conclusão a que se chegou é que 4 iterações era o valor comum para todos, com excepção do SPOSEM3D que, por incluir na sua estrutura um factor aditivo, apresenta uma convergência mais lenta. Neste caso as imagens equivalentes eram obtidas ao fim de 15 iterações.

Com esta informação compararam-se, depois, os desempenhos dos diferentes algoritmos, e chegou-se às seguintes conclusões: o FORE+AWOSEM tem um desempenho global superior ao FORE+AWOSEM; o mesmo se pode constatar quando se comparou o ANWOSEM3D com o FORE+AWOSEM. Quanto às vantagens do ANWOSEM3D sobre o FORE+NECOSEM, não são tão óbvias. Os resultados da comparação do SPOSEM3D (15 iterações) com as 4 iterações do FORE+AWOSEM e FORE+NECOSEM, são idênticos aos que se verificaram no caso do ANWOSEM3D.

Quanto ao NECOSEM 3D, verificou-se ter um desempenho melhor que o FORE+AWOSEM, mas quando comparado com o FORE+NECOSEM as diferenças não se revelaram muito significativas. Relativamente aos dois outros algoritmos 3D não há, no essencial, diferenças de desempenho. Ora, isto significa, na prática, que a complexidade adicional inerente à utilização do NECOSEM3D não tem tradução prática num incremento na detectabilidade de tumores. Ou seja, os benefícios não justificam os custos da sua utilização. De facto, há problemas relacionados com a definição dos valores dos coeficientes NEC em bins relativos a LORs que atravessam a fronteira do objecto. Neste casos é necessário usar uma estratégia que passa por recorrer à carta de atenuação para perceber onde estão as fronteiras do objecto e, em bins que estão nesta região, estabelecer um valor fixo para os coeficientes NEC. Este procedimento, porém, pressupõe um ajuste caso a caso que nem sempre resulta, sublinhando a complexidade intrínseca ao NECOSEM3D.

Com o conjunto de informações obtidas a partir do estudo de simulação, passou-se, de seguida, à utilização deste algoritmos para reconstruir imagens a partir de dados reais, obtidos com pacientes que fizeram exames PET de corpo inteiro no scanner HR+ do S.H.F.J..

Do ponto de vista da utilização dos algoritmos OSEM ponderados, a principal diferença relativamente aos estudos de simulação é que deixamos de estar nas condições ideais, ou seja, deixamos de saber com exactidão os valores dos factores de ponderação a usar.

No caso da atenuação, o problema resume-se, simplesmente, a ter de lidar com valores afectados de ruído. No entanto, no caso dos coeficientes NEC o problema é substancialmente mais complexo, uma vez que, para cada bin, não há senão informação relativamente à contagem nesse bin, ou seja, não há qualquer tipo de informação relativamente ao valor médio ou variância que lhe correspondem. Por conseguinte, para usarmos o NECOSEM tivemos de arranjar um processo para obter os valores destes dois momentos, o que, basicamente, consistiu em fazer a convolução das projecções com um *kernel* gaussiano. Foi também necessário identificar os bordos do objecto, para, nos bins relativos à fronteira, se atribuir um valor fixo ao coeficiente NEC. Deste modo tentou contornar-se o problema das flutuações dos valores dos coeficientes NEC nesta região. A delimitação dos contornos do objecto foi efectuada com recurso ao mapa da atenuação, estabelecendo-se que LORs para as quais a atenuação fosse superior a 5% são LORs inequivocamente no interior do objecto e, portanto, o coeficiente NEC pode ser calculado de acordo com a equação que se referiu anteriormente. Para os restantes bins, ao coeficiente NEC é atribuído o valor 1. Este limiar de 5% usado para delimitar os contornos da atenuação foi, porém, estabelecido em função dos casos concretos em estudo, mas não podemos argumentar contra a utilização de outros valores, como 8% ou 10%. É, por conseguinte, um parâmetro estabelecido ad-hoc.

As imagens foram reconstruídas usando o FORE+AWOSEM, o FORE+NECOSEM e o ANWOSEM3D. Para as analisar, solicitou-se a colaboração de uma médica com vasto treino na observação deste tipo de imagens. O resultado da análise por ela efectuada pode resumir-se ao seguinte: não se vislumbram diferenças significativas entre as imagens reconstruídas pelos três processos. Dado o facto de as imagens terem sido analisadas por um único médico, esta deve revestir-se apenas de carácter preliminar, sendo necessária uma análise mais vasta, envolvendo mais médicos, de modo a que se possam tirar ilações com carácter mais definitivo. Independentemente disto, consideramos que, em função das aproximações envolvidas na estimação dos coeficientes NEC, seria porventura mais surpreendente se a conclusão prévia apontasse no sentido oposto.

Pensamos ainda ter dado um contributo para demonstrar que, pese embora os problemas práticos que envolve, é possível reconstruir imagens a partir de dados clínicos reais, usando o FORE+NECOSEM.

PALAVRAS-CHAVE: Tomografia por Emissão de Positrões (PET), reconstrução estatística, transformada NEC (*Noise Equivalent Count*), OSEM (*Ordered Subsets Expectation Maximization*) ponderado, detecção de tumores.

Abbreviations

3DRP	3D Reprojection Algorithm
ACF	Attenuation Correction Factor
ART	Algebraic Reconstruction Technique
BGO	Bismuth Germanate
BMI	Body Mass Index
DFM	Direct Fourier Methods
DOI	Depth-of-Interaction
FBP	Filtered Backprojection
FDG	2-[¹⁸ F]Fluoro-2-deoxy-D-Glucose
FFT	Fast Fourier Transform
FORE	Fourier Rebinning
FOV	Field-of-View
FWHM	Full-Width at Half Maximum
HR+	ECAT EXACT HR+ (CTI/Siemens)
LOR	Line of Response
LSO	Lutetium Oxyorthosilicate
MAP	Maximum <i>a Posteriori</i>
ML-EM	Maximum Likelihood – Expectation Maximization
MRI	Magnetic Resonance Imaging
NEC	Noise Equivalent Count
NEMA	National Electrical Manufacturers Association
NPWMF	Non-Prewhitening Matched Filter
OSEM	Ordered Subsets – Expectation Maximization
PET	Positron Emission Tomography
PMT	Photomultiplier Tube
PSF	Point Spread Function
SIR	Statistical Iterative Reconstruction
SPECT	Single Photon Emission Computed Tomography
SNR	Signal-to-Noise Ratio
WLS	Weighted Least-Squares

Introduction

Positron Emission Tomography (PET) is a functional medical imaging technique that can be used to image biochemical or physiological processes within the body. Since it relies on isotopes undergoing a positron decay, it belongs to the field of Nuclear Medicine.

As many other imaging techniques, it is rapidly moving forward on many aspects: radiochemistry, instrumentation, image reconstruction, image processing and visualization, as well as also taking part in the ever more thrilling field of image fusion (nowadays through PET/CT).

The research work on the basis of the present thesis was devoted to one of the developing fields referred above: image reconstruction. In PET, this is equivalent to say image reconstruction from projections, which is, in fact, the reason for its classification as a tomographic technique.

There are two approaches for image reconstruction from projections: analytical and algebraic. Statistical¹ algebraic methods, such as OSEM (Ordered Subsets Expectation Maximization), have, among others, the advantage - over the analytical ones - of incorporating a model for the measurement uncertainty. In other words, OSEM - the adopted reconstruction algorithm - intrinsically incorporates a statistical model for the data it will be handling during the reconstruction process. This model is a Poisson model, i.e., OSEM assumes Poisson data.

The justification for relying on OSEM for image reconstruction is the following: it is an algebraic method; it is rapid² (when compared to other statistical algorithms such as the ML-EM); it is relatively easy to implement; it doesn't have additional dependencies on other parameters³ (like regularization terms) which are often not set in a straightforward manner; and, in consequence, it is widespread in clinical routine.

However, whilst rapid, OSEM is not so rapid as one would like, in special when used for directly reconstructing 3D data (data containing information for LORs within planes that are not normal to the scanner axis). In that case, the huge amount of data makes the image reconstruction much slower. Hopefully, rebinning techniques were devised to sort 3D data into 2D data. Not all of them are equally accurate, but the Fourier Rebinning (FORE) represents a

¹ ART is an example of a non-statistical algebraic reconstruction algorithm.

² Not as rapid as the analytical algorithms, but with the advantages of being algebraic.

³ Such as most of the Bayesian reconstruction methods.

good compromise between accuracy and implementation feasibility for the rebinning of 3D into 2D data.

Assuming that we will be just dealing with 3D data⁴, the image reconstruction can proceed in two different ways: directly using a fully 3D version of OSEM without rebinning data – more time consuming and computationally demanding; or using FORE to convert 3D data into 2D data, and reconstruct through a 2D OSEM implementation – more rapid, but subjected to the constraints imposed by the use of FORE.

The critical constraint associated with the use of FORE are not the approximations done during the rebinning process. The unavoidable constraint imposed by FORE is the need to correct collected data, prior to the use this algorithm, for geometrical (normalization) and physical effects: attenuation, scattered and random photons. The result is that corrected data are no longer Poisson data. This aspect is the cornerstone of the present thesis: the use of OSEM for the reconstruction of corrected data, adopting a strategy to restore the Poisson-like characteristics of these data, as required by the theoretical foundations of OSEM.

It should also be mentioned that whilst FORE requires data previously corrected for the effects referred above, in practice, 3D data are also often corrected for the same effects. Therefore, the problem of having non-Poisson data is not specific of a FORE+OSEM2D reconstruction, but should be taken into account in almost all of the 3D OSEM implementations.

Having hitherto referred the motivation for the research work, it is now important to describe how we have proceeded and what are the objectives we would like to reach.

Restoring the Poisson-like characteristics of the data, i.e., making the variance equal to the mean, was achieved through the use of weighted OSEM algorithms. The weights, when multiplied⁵ by the data, grant them this propriety. At least, these weights transform data in order to approximate these two moments. In fact, not under all the weighted OSEM versions used in our work (AWOSEM and NECOSEM, for 2D; ANWOSEM, SP-OSEM and NECOSEM3D, for 3D) data are given the Poisson-like characteristics. In AWOSEM and ANWOSEM, weighted data do not have a mean equal to the variance. These two moments are closer than they would be without weighting, but the objective is not completely fulfilled.

The NEC transformation is the approach adopted for restoring the Poisson-like characteristics of data. The weighted OSEM version relying on the NEC weighting

⁴ No septa in the scanner.

⁵ Or added, or, even in a more general approach, multiplied and added. In the latter case, two different terms exist: one multiplicative term, and one additive term.

transformation is known as NECOSEM. The distinctive feature of such algorithm is the definition of the NEC multiplicative weight: for each data bin this factor is calculated as the ratio between the mean value and the variance. With real clinical data this is the source of a fundamental problem: since there is only one value collected for each bin – the data value itself – there is no knowledge about the mean and the variance for that bin. For simulated data, on the contrary, if we keep track of the values for these two moments, the exact values for the NEC weights can be calculated.

The goal of the research work described in the present thesis is to incorporate an accurate statistical modeling in the OSEM reconstruction algorithm used for whole-body PET imaging. As referred above, this is done through the use of weighted versions of the OSEM algorithm. However, since not all of the weighted versions restore the Poisson-like characteristics of data required by OSEM, we have compared their performance. In addition, the fact that our raw data are 3D data, lead us to use the two different pathways for image reconstruction: fully 3D reconstruction (using complete and non-complete weighting restoration of Poisson-like characteristics of data), and FORE + 2D OSEM (once again, using complete and non-complete restoring).

The comparison of the performance was done using simulated data and clinical data. In the former case, the use of an analytical simulator allowed us to have complete control over the data characteristics. In other words, we know exactly the values needed to calculate all the weights involved in the reconstruction process. We are, therefore, from the weighting point of view, under the best performing conditions of the algorithm. Simulation studies are one of the two main blocks into which the developed research work can be grouped.

From simulated data, and apart from the conclusions concerning the reconstruction algorithms' performance, some information is obtained in respect to specific parameters, such as the FWHM of the post-smoothing filter applied to images and the number of iterations one can use to compare algorithms.

With that knowledge, we proceeded to the application of some of the algorithms to clinical data: the second major section of the work. The key point when using clinical data is that there is no exact information about the weighting values. This is the case even for the attenuation, because the Attenuation Correction Factors (ACFs) are affected by noise. Nevertheless, for each LOR (data bin) there is a measured value for the attenuation and normalization procedure, so weighting factor based on this parameters can be directly computed. In order to use a NEC based reconstruction, on the contrary, it is necessary, first, to estimate the mean value and the variance for each data bin. This was done, but the results, i.e., the NEC

weights estimated that way, contain a certain degree of uncertainty and we are, for sure, no more under the optimal conditions obtained with simulated data.

For comparing the algorithms' performances, two different strategies were devised: the huge amount of data resulting from simulation studies made human observer studies impracticable. So, we used a numerical observer – the NPWMF observer - whose performance for similar studies and similar conditions mimics the performance of humans. Lartizien [Lartizien, 04] has shown in a previous study, for similar imaging conditions, that the NPWMF detectability index allows good correlation with human observer detectability. The NPWMF detectability index was the main tool to access the performance of the different algorithms working over simulated data. For clinical data, the use of numerical observers is not possible, and the comparisons between images were done by a physician who is largely familiar with the images obtained for this sort of whole-body PET scans.

The thesis is organized as follows: Chapter 1 is an overview of some introductory topics related to the PET imaging technique. Chapter 2 is dedicated to the theory of the image reconstruction in PET. Because this is a very large field of investigation involving different approaches and a mired of specific questions, we concentrate on those aspects having a more or less direct connection with our work, or that are relevant in order to put it into context.

Chapter 3 is the link between the theoretical and the experimental part. The weighted OSEM approaches are detailed, as well as the NEC transform and its application to the restoring of the Poisson-like characteristics of the data. The main features of the specific OSEM implementation used in our work are also described.

Chapter 4 is devoted to the simulation studies, using the ASIM analytical simulator and the NPWMF numerical observer. It describes, as referred, the first block of the experimental work sustaining this thesis.

The second block, i.e., the reconstruction of clinical data is the core of Chapter 5. Some procedures adopted in that case are a consequence of part of the results obtained for simulated data.

We finish with a global discussion where some conclusions are drawn and future developments suggested.

The very last part of the thesis is an appendix containing additional material that complements the text in the other chapters.

1 – An introduction to Positron Emission Tomography (PET)

This chapter is conceived as an introduction to the medical imaging modality discussed in this thesis: the Positron Emission Tomography (PET). This is one of the two major basic modalities that go under the designation of Nuclear Medicine imaging techniques (the other is SPET, which stands for *Single Photon Emission Tomography*).

Before proceeding with a general description of the technique, one interesting way to introduce Positron Emission Tomography is to inspect the paragraph above. Let us start by the name itself: *Positron Emission Tomography*. Alone, it provides the following information:

- it is an emission imaging technique. This means that, independently from the type of radiation involved in the technique, the radiation is emitted from inside the object (the body of the patient). This contrasts with conventional radiology or CT (*Computed Tomography*), since these are transmission techniques, i.e., rely on the transmission of a radiation beam throughout the patient;
- positrons are the radiation type at the core of PET. This is a fact, but it should be mentioned in advance that positrons are not the type of particle to be detected. More: the emitted beam of radiation is not a beam of positrons (as to be seen later, emitted particles from the body are γ photons). So, it is better to clearly separate *Positron Emission* into two separate meaningful words: *Positron* plus *Emission*.

However, since positrons have the leading role in PET and these particles result from the energetic rearrangement of an unstable nucleus, it becomes clear why PET is one of the Nuclear Medicine imaging techniques: it has a nuclear decay process on its basis. Once more, this physical evidence distinguishes PET, for example, from CT, because the latter uses electromagnetic radiation resulting from an atomic process;

- it is a tomographic technique. In fact, PET is the synthesis of a physical process (the emission and detection of radiation with origin in the nucleus¹) – the tracer principle -

¹ This is also the principle of scintigraphy, which is not a tomographic imaging technique.

and a mathematical procedure: the reconstruction of a 3D object (activity distribution) from its 2D projections (detected radiation) - the tomographic principle². Considering this point, PET and CT are under the same classification: both are tomographic modalities, which is neither the case of scintigraphy nor of conventional radiology.

1.1 – A brief history of PET

Positron emission from radioactive nuclei was discovered in 1933 by Joliot and Thibaud. In 1946, the importance, for medicine, of positron emitting radionuclides produced in a cyclotron was suggested [Jones, 03a].

The roots of PET, however, are fixed in 1951. In this year, two major events are reported where, even though very far from the tomographic technique used nowadays, the principles of this technique can be identified: Wren and his colleagues arranged two sodium iodine scintillation counters, in opposite positions, to detect a positron emitter, ^{64}Cu , in brain tumors [McCready, 00; Nutt, 02]. The idea was to record differential uptake of radiopharmaceuticals by the brain; Sweet and colleagues, at the Massachusetts General Hospital, used two sodium iodine crystals, coupled together in coincidence mode and rectilinear scanning, to localize brain tumors [McCready, 00; Nutt, 02]. These two works represent the first attempts to record positron emission data for a medical application.

The Anger camera, introduced in 1954, using NaI(Tl) or plastic scintillators, was soon applied for coincidence measurements of positron emitters [McCready, 00; Lundqvist, 98]. During the 60s, clinical studies were performed with this device, although it was not yet possible to obtain tomographic images.

In 1973, Robertson, from the Brookhaven National Laboratory, built the first ring tomograph. However, because of limited sampling, lack of attenuation correction and of a proper image reconstruction algorithm, it was unable to obtain true reconstructed cross sectional images. In the same year, M. Phelps, at Washington University, built the first PET tomograph: PETT I (*Positron Emission Transaxial Tomography*). This attempt was also unsuccessful in producing proper reconstructed images, because it employed lead collimators, a limited sampling and did not provide attenuation correction [Nutt, 02].

² Both principles have been awarded with the Nobel Prize: the tracer principle, in 1943 (de Hevesy); the tomographic principle, in 1979 (Hounsfield and Cormack) [Lundqvist, 98].

Henceforth, PET enters its modern era. By the beginning of 1974, PET II, a scanner conceived for phantom studies, was developed by Phelps and colleagues, and the principles of PET published based on studies performed with this scanner (as well as with PET II $\frac{1}{2}$, an upgraded version) [Nutt, 02].

Still in 1974, Phelps, Hoffman and colleagues, at Washington University, constructed PET III for human studies. It was composed of 48 NaI(Tl) detectors, organized as a hexagonal array, and the gantry had a rotation capacity of 60°. The first images of blood flow, oxygen, and glucose metabolism of Fluorine 18 (F-18) bone scans from PET III represented the first published human PET images using the Filtered Backprojection (FBP) algorithm [Nutt, 02].

Following PET III, the first commercial PET scanner was designed by EG&G ORTEC in collaboration with Phelps and Hoffman. It was named ECAT II (*Emission Computed Axial Tomograph*) and delivered in 1976. ECAT II used a total of 96 NaI(Tl) crystals with a diameter of 3.75 cm and had a PDP-11 computer with 32 Kbytes of memory [Nutt, 02].

After 1976, with the principle of positron emission tomography undoubtedly established, the commercial delivering of PET scanners begun. Benchmarks in the history of this imaging technique moved, therefore, to the field of detectors, radiopharmaceuticals, reconstruction algorithms and camera design, as well as image fusion. Some of these benchmarks are [Nutt, 02]:

- the first evaluation of BGO for use in PET, performed by Cho, Farukhi and Derenzo. The first tomograph using BGO was designed in 1978 by Chris Thompson and his group at the Montreal Neurological Institute. In the same year, EG&G Ortec produced NeuroEcat, the first commercial tomography to use BGO;
- the synthesis of FDG by A. Wolf and J. Fowler's group, and the first PET imaging with FDG, by Phelps and its group at UCLA, using the ECAT II scanner. Latter, Hamacher and colleagues developed a new synthesis method for FDG, which is still in use today;
- the discovered and grown of the first LSO crystals, in the period of 1989 – 1992, and its latter (1995) refinement and cost effectiveness;
- the development of the PET/CT scanner, combining a ring PET scanner with a CT scanner in the same gantry, allowing morphological and functional data fusion.

1.2 – Positron Emission Tomography Overview

From a medical/biological point of view, PET is an interesting technique because it can be used to obtain information *in vivo* about the way an organ is functioning (for example, its metabolism or perfusion) as well as to gain insight into some biochemical processes at the cellular level. This is the main reason why PET (and SPET, also) is classified as a functional image technique³. However, this classification is, probably, better understood if matched against anatomical imaging techniques – such as the conventional use of MRI or CT⁴ –, where just structural (thus anatomical) information is obtained. New imaging techniques, such as the PET/CT⁵, SPET/CT, or, more recently, PET/MRI [Lucignani, 06], can superpose functional information over an anatomic image, thus being referred as fusion imaging techniques.

In Figure 1.1 we show an example of a typical whole-body [¹⁸F]-FDG image obtained with the widespread ECAT EXACT HR+ scanner.

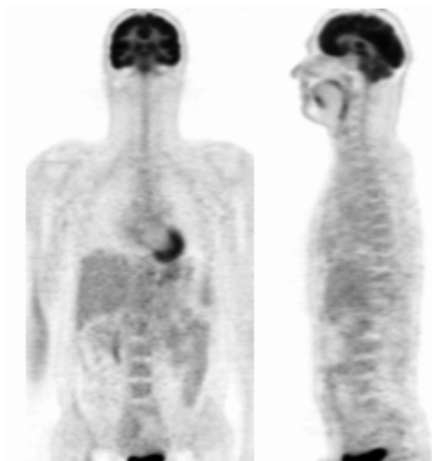


Figure 1.1

Typical whole-body FDG PET image, obtained from the HR+ scanner, for a patient with a body mass index equal to 23.7 and a dose of 338.5 MBq: a) coronal cross section; b) sagittal cross section.

The principle of PET is based on the detection, in coincidence, of radiation emitted from the interior of the body of the patient. So, from the previous sentence, some questions can be formulated: how is the radiation introduced into the body?; which radiation is involved in the process?; how is the radiation detected; and what does mean “detection in coincidence”?

The answer to the first question is beyond the scope of this thesis. However, in short, everything starts by the definition of the biological process under study and with the

³ Another widespread functional image modality is the fMRI – Functional Magnetic Resonance Imaging.

⁴ Not considering fMRI or functional imaging in CT [Dawson, 06].

⁵ PET/CT is an imaging modality resulting from the fusion of a PET and CT systems, under the same scanner. This leads to images where the missing information in CT (functional information) becomes available through the PET component. The weakness of PET images – poor spatial resolution – is, on the other hand, overtaken by the CT anatomical image.

identification of some molecule that takes part in that process. Then, one atom of the chosen molecule is replaced by a positron emitter radionuclide. The final labeled molecule – the tracer – is, finally, introduced into the body, either injected, inhaled or by some other mean prescribed from biological and medical considerations. The tracer is the source of information in PET.

The question, of which radiation is involved in PET, can be misleading. In fact, if the radionuclides introduced into the body are positron emitters, the detected radiation are γ photons. This is, moreover, inherently related with the answer to the third question, since not only the type of radiation to be detected should be considered, but also the fact that these γ photons result from a positron/electron annihilation process.

If, in PET, γ photons result from a positron/electron annihilation, then⁶: these photons have an energy equal to 511 keV; are emitted in pairs; each pair corresponds to one annihilation; and, for a certain annihilation, the two photons are emitted back-to-back (in opposite directions) simultaneously. So, in principle⁷, using the appropriate detectors, if two photons are detected in coincidence – within a very short timing window⁸ –, this means that the annihilation took place somewhere along the line connecting the two detectors struck by the photons. This line is called a line-of-response (LOR) (Fig. 1.2). Considering all the LORs, it is possible to reconstruct an image of the distribution of the points, inside the body, where annihilations took place. Since we are assuming that annihilations occur at the exact position where positrons are emitted, the former image is, therefore, an image of the distribution of the tracer locations. In short: the detection, in coincidence, of two 511 keV γ photons emitted, back-to-back, from the point of the positron annihilation is the main physical process beneath the possibility to reconstruct an image of the tracer inside the body. From the perspective of the final user of this image technique, this is the raw data conveying information about the biological process under study.

With the overall principle of PET described, it is possible to identify four main constraints by which this technique is affected. The first one is the biological accuracy with which a chosen tracer molecule defines the specific biological pathway under study. The second one is the physical accuracy, or the accuracy with which one can assume that the annihilation takes place at the exact location where the positron is emitted. The third one is a technical constraint and is related with the ability to effectively register an annihilation via the detection, in

⁶ This is not strictly true, but useful for a first approach. The noncolinearity and the three photons annihilation will be discussed latter on this chapter.

⁷ Under ideal assumptions such as the absence of random or scattered events or considering that the annihilation takes place exactly at the same location where the positron is emitted. This is, however, far from reality, but with the correct approach, the principle remains valid.

⁸ Typically 6 to 12 ns wide [Zanzonico, 04].

coincidence, of the two corresponding photons. The last one can be established at a mathematical/processing level, and is associated with trustworthy image reconstruction.

In addition to the *in vivo* functional information conveyed by PET images, this technique is also regarded as a quantitative imaging tool: individual voxel values in the reconstructed object represent the local radiotracer concentration in the body, in absolute units of radioactivity concentration, within reasonable accuracy and precision. This accurate and precise mapping of the radiotracer concentration in the body is important for two main reasons: it provides a true reflection of the physiological process under study and enables the use of tracer kinetic methodology to model the time-varying distribution of the labeled component in the body, allowing the estimation of relevant physiological parameters [Meikle, 03]. However, in order to provide the meaningful quantitative information, acquired data should be corrected for some different effects, prior to reconstruction. This is the main reason for data correction, described latter in this chapter.

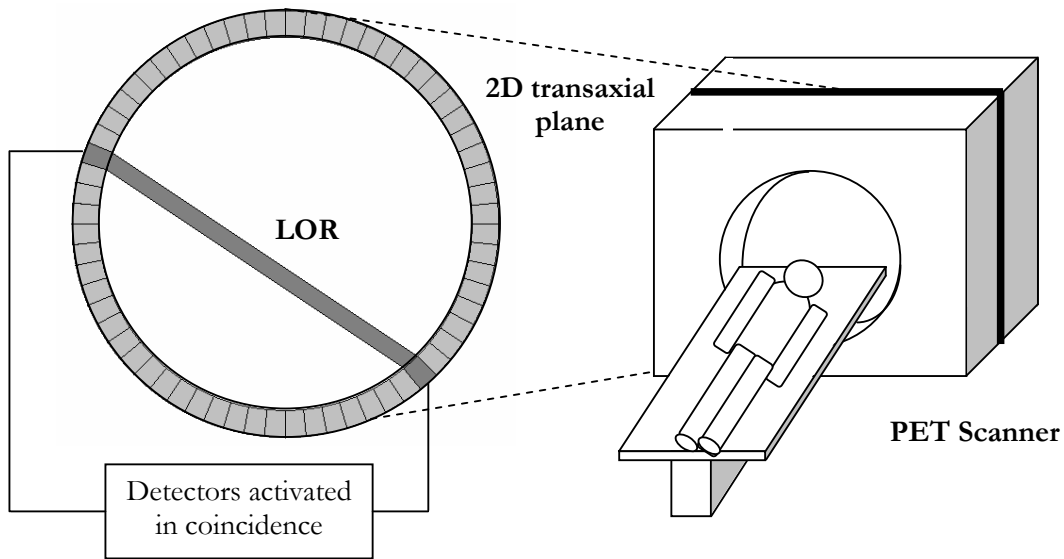


Figure 1.2 – Schematic representation of a LOR within a transaxial plane.

1.2.1 – Physical constraints and the three-gamma annihilation

1.2.1.1 – Positron range

The positron-electron annihilation occurs at the end of the positron range, when the positron has dissipated almost all of its kinetic energy and both the positron and the electron are

at rest⁹. However, for a given radionuclide, positrons are emitted over a spectrum of kinetic energies ranging from 0 to a characteristic maximum (endpoint) energy, E_{max} ; the associated average positron energy, \bar{E} , is approximately one-third of the endpoint energy: $\bar{E} = 1/3 E_{max}$ [Cherry, 03]. So, in order to lose this amount of energy, positrons must travel a finite distance from the decay nucleus, which is done along a very tortuous path (Figure 1.3), in consequence of multiple large-angle deflections. The distance traveled by the positron in tissues depends on the electron density of the medium and, for a given medium, on the positron's energy. This distance ranges from 0 up to a maximum value called the extrapolated range, R_e , corresponding to the highest energy positrons. This extrapolated range is the maximum distance that the positron would travel if it was not significantly deflected and traveled in a straight line to the end of its range¹⁰. In addition, it is often considered the *effective positron range*: the average distance from the emitting nucleus to the end of the positron range, measured perpendicular to a line defined by the direction of the annihilation photons [Cherry, 03].

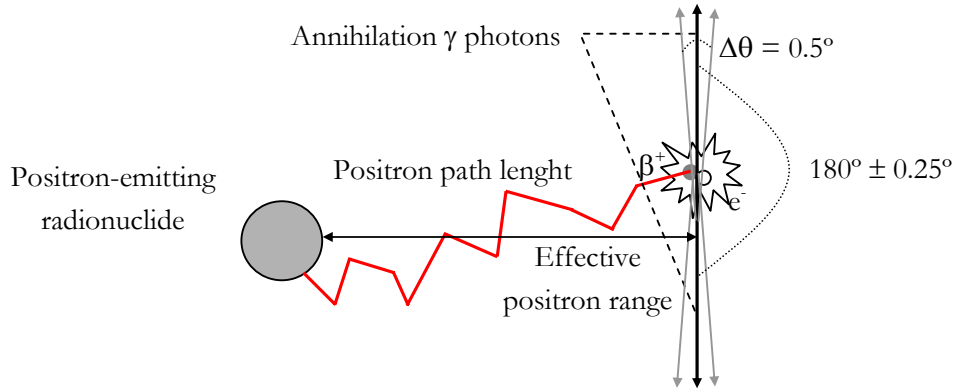


Figure 1.3 – Physical aspects of positron-electron annihilation: positron range and non-colinearity ($\Delta\theta$).

For positrons emitters used in PET, the maximum energies vary from 0.58 MeV to 3.7 MeV, the maximum (or extrapolated) ranges, (R_{max}), from 2 mm to 20 mm, and the root-mean-square (rms) ranges (R_{rms} , and $FWHM_{Range} \approx R_{rms}$) from 0.2 mm to 3.3 mm [Zanzonico, 04]. This latter parameter is important since the positron range distributions are cusp-like shaped, with long tails, and, therefore, the FWHM is not the best indicator of positron range [Cherry, 03].

⁹ The total energy of the pair positron-electron is 1.022 MeV and the total moment is zero. Thus, in accordance to the principles of energy and momentum conservation, the total energy of the two annihilation γ photons must be equal to 1.022 MeV and their total moment zero.

¹⁰ This is equivalent to the positron path length.

Some relevant values associated with the physical properties of positron-emitting radionuclides used in PET can be found in Table 1.1.

A Monte Carlo method for the calculation of the positron annihilation point distribution when traversing a dense medium can be found in [Levin, 99]; a model for predicting the annihilation density distribution arising from a point source positron emitter is proposed in [Palmer, 92].

Radionuclide	Half-life ($T_{1/2}$)	Maximum β^+ energy (MeV)	β^+ range in water (mm)	
			Max. range (R_{max})	Rms range (R_{rms})
^{11}C	20.4 min	0.96	3.9	0.4
^{13}N	9.96 min	1.2	5.1	0.6
^{15}O	2.05 min	1.7	8.0	0.9
^{18}F	1.83 h	0.64	2.3	0.2
^{62}Cu	9.74 min	2.9	15	1.6
^{64}Cu	12.7 h	0.58	2.0	0.2
^{66}Ga	9.49 h	3.8	20	3.3
^{68}Ga	1.14 h	1.9	9.0	1.2
^{76}Br	16.1 h	3.7	19	3.2
^{84}Rb	1.3 min	3.4	18	2.6
^{86}Y	14.7 h	1.4	6.0	0.7
^{124}I	4.18 d	1.5	7.0	0.8

Table 1.1 – Some physical properties of positron-emitting nuclides used in PET (adapted from [Zanzonico, 04]).

1.2.1.2 – Annihilation photons non-collinearity

Since it is assumed when defining a LOR that the annihilation photons are emitted back-to-back, it is important to consider the physical constraint of the non-collinearity of these two annihilation photons.

The deviation from collinearity is due to the residual momentum and kinetic energy of the orbiting electron. How large is the non-collinearity for annihilation γ photons traveling in water or in the living tissues is often not obvious. For example: Zanzonico ([Zanzonico, 04]) refers to Berko [Berko, 56] to say that the deviation, i.e. $\Delta\theta$ in Fig. 1.3, can go up to 0.5° ; however, in the latter paper, some studies are present for different metals, but none for water or any other relevant material in the scope of PET. Thus, it is not clear where that value comes from. In another paper by Ollinger and Fessler ([Ollinger, 97]), it is assumed that the magnitude of the deviation is on the order of one degree or less; Raylman [Raylman, 96] assumes 4 mrad

($\approx 0.23^\circ$), approximately, for the angular deviation of photons emitted in free annihilations. For Humm [Humm, 03], if an approximated calculation of the deviation from the co-linearity is made just by considering the thermal motion of the particles and the conservation of the Fermi moment, this would give $180^\circ \pm 0.25^\circ$. In addition, the same author mentions that the distribution of the angular deviation in water is assumed to be Gaussian with a FWHM approximately equal to 0.5° .

In sum, the value of $\pm 0.25^\circ$ for the non-colinearity spread of annihilation photons is usually assumed. A study of the influence of such non-colinearity on PET image spatial resolution can be found in [Sánchez-Crespo, 06].

1.2.1.3 – Three photons positron annihilation

The two-gamma decay is not the only process for positron annihilation. In fact, an S -state electron-positron pair can exist in the singlet state of zero-spin angular momentum or in one of three triplet states of unit spin angular momentum. From the general symmetry properties, it was demonstrated that the single state annihilates with two photons, whereas the triple states have to annihilate via three photons¹¹ [Berko, 56]. In the same paper, Berko shows an expression for the spin-averaged cross section for a 2γ annihilation, as well as a ratio for 2γ and 3γ spin average cross sections, in the case of a free positron: $\sigma_{2\gamma} / \sigma_{3\gamma} = 372$. This result is drawn from a previous paper by Ore [Ore, 49], where the supporting calculations are detailed¹².

However, in matter, prior to annihilation, a positron can capture an electron to form an hydrogen-like system called positronium (Ps), which is a short-lived bound state of a positron and an electron. Due to spin statistics, 25% of all positroniums are formed in the single state ($S = 0$) called para-positronium (p-Ps), while the remaining form the triplet state ($S = 1$) called ortho-positronium (o-Ps) [Kacperski, 04]. Still discussing the behaviour in matter, interactions with the surrounding electrons prevail, leading to a direct annihilation of the positron with one of the electrons (pick-off process), the conversion of o-Ps into the p-Ps which then decays rapidly via 2γ , or to chemical reactions producing short-lived Ps compounds. As a result, the effective yield of 3γ annihilation in matter depends on the rates of Ps formation and quenching. So, from a practical perspective, this means that in non-metallic materials, like water, the 3γ yield is usually of the order of 0.5% [Kacperski, 04].

¹¹ 4γ and 5γ gamma decays [Chiba, 98; Adachi, 94] have also been detected experimentally, but their probabilities are very low and, thereby, of little interest for PET [Kacperski, 04].

¹² Other relevant papers regarding three-photon annihilation are, for example, the paper by DeBenedetti [DeBenedetti, 54] and the one by Ferrell [Ferrell, 56], where the theory of positron annihilation is revised.

For positron emission tomography, the principle behind the detection in coincidence of the three annihilation photons is ideal (perfect collimation): using energy and momentum conservation laws and a set of nonlinear equations to be solved numerically (rather than complex image reconstruction methods), it would be possible to determine the decay site, independently for each event. Kacperski [Kacperski, 04] explores the triple coincidence detection for PET, in parallel with the conventional coincidence of two back-to-back annihilation photons.

1.2.2 – Basics of signal detection in PET

The great majority of commercial PET scanners available nowadays, use, as signal transducers - i.e., for converting γ photons into an electrical signal - scintillation crystals optically coupled to photomultiplier tubes.

As radiation crosses the scintillator, the crystal becomes excited, causing the emission of light. This light is transmitted to the photomultiplier where it is converted into a weak current of photoelectrons. Such electrical signal is then, usually, driven to a pre-amplifier and subsequent shaping circuitry and, finally, enters the digital processing stages.

1.2.2.1 – Detectors

As referred above, in commercial PET scanners, detector is almost synonymous of a scintillator material.

The main physical process behind the behavior of these scintillators is called fluorescence: the prompt¹³ emission of visible radiation from a substance, following its excitation. If the emission is delayed because of metastable excited states involved in the process, this is known as phosphorescence. The delay time between absorption and reemission can be from a few microseconds up to hours, depending on the scintillation materials. Another emission process is delayed fluorescence: it results in the same spectrum as prompt fluorescence, but is also characterized by a longer time delay between excitation and light emission [Knoll, 89; Leo, 94]. In any case, light emission is small: typically a few hundred to a few thousand photons for a single γ -ray interaction within the range of 70 keV to 511 keV [Cherry, 03].

The ideal scintillator should satisfy the following requirements [Knoll, 89; Leo, 94]:

- high efficiency for conversion of exciting energy to fluorescent radiation. For a high efficiency, scintillator materials should have both a high effective atomic number (Z_{eff}) and

¹³ Within 10^{-8} s [Leo, 94].

a high density. Greater light yield (number of photons per MeV) can be translated into a more linear response, better energy resolution and a better spatial resolution;

- linear conversion. i.e., the light yield should be proportional to absorbed energy;
- to be transparent to the wavelength of its own emission. There are usually absorptive losses in the crystal due to the partial overlap of its emission and absorption bands;
- a decay time of the induced fluorescence as short as possible, for fast signal generation;
- good optical quality and capable of being manufactured in sizes large enough to be of interest as a practical detector;
- refraction index as near as possible to that of glass (≈ 1.5) to permit an efficient coupling to the PMT. The emission wavelength of the scintillation light should match the light response characteristic of the PMT's photocathode¹⁴.

No material meets simultaneously all the criteria just stated. Consequently, the choice of a particular scintillator is the result of a trade-off between these factors. Scintillation materials used in nuclear medicine can be of two general types: inorganic substances in the form of a solid crystal, or organic substances dissolved in a liquid solution [*Cherry, 03*]. Inorganic crystals tend to have the best light output and linearity, but have relatively slow response times; organic scintillators are generally faster but yield less light [*Knoll, 89*].

Material	Density (g/cm ³)	Z _{eff}	Attenuation length for 511 keV γ (mm)	Prob. of PE	Light output (ph/MeV)	Decay time (ns)	Scintillation emission λ (nm)	Refractive index
BGO	7.1	75	10.4	40	9000	300	480	2.15
LSO	7.4	66	11.4	32	30000	40	420	1.82
NaI:Tl	3.67	51	29.1	17	41000	230	410	1.85
CsI:Tl	4.51	52	22.9	21	66000	900	550	1.80
GSO	6.7	59	14.1	25	8000	60	440	1.85
LGSO					23000	40	420	
LuAP	8.3	64.9	10.5	30	12000	18	365	1.94
YAP	5.5	33.5	21.3	4.2	17000	30	350	1.95

(BGO: Bi₄Ge₃O₁₂; LSO: Lu₂SiO₅:Ce; GSO: Gd₂SiO₅:Ce; LGSO: Lu_{1.8}O₂SiO₃:Ce; LuAP: LuAlO₃:Ce; YAP: YAlO₃:Ce. PE: Photoelectric effect.)

Table 1.2 – Some properties of scintillator materials used for PET (adapted from [*Humm, 03*]).

¹⁴ The widespread bi-alkali PMTs have a maximum sensitivity in the range 390-410 nm. So, useful scintillators should have their maximum intensity in the wavelength range between 380 and 440 nm.

Inorganic scintillators – by far the most used in PET - are crystalline solids whose scintillation properties are due to the characteristics of their crystal structure, and, thus, are scintillators only in crystalline form. In addition, most of the inorganic scintillators are impurity activated¹⁵: their crystalline structure contains small amounts of atoms of other elements that disturb the normal electronic band structure and are responsible for the scintillation effect.

The scintillation mechanism in inorganic materials depends on the energy states characteristics of their crystal lattice [Knoll, 89; Leo, 94].

Table 1.2 shows the main properties of scintillator materials used for PET. A detailed description of these properties and their implication in the scintillator performance can be found, for example, in [Humm, 03; Melcher, 00; Balcerzyk, 00]. Since the performance of a PET system strongly depends on the crystal, the most used types deserve a short reference:

- **BGO**: it is a pure inorganic scintillator that requires no activator. BGO is relatively hard, rugged and non-hygroscopic. Its high density (7.1 g/cm^3) makes BGO the detector material with the greatest attenuation coefficient for 511 keV gamma rays. The main disadvantages of this crystal are: its low light yield, in result of which it exhibits an inferior energy resolution; poor timing properties; and a fluorescence which is dependent on the temperature, requiring a more stringent environmental regulation for the operation of BGO scanners [Humm, 03].
- **LSO**: its main advantages are a good light yield (30000 ph/MeV); a short decay time (≈ 40 ns); and good mechanical properties. For 511 keV it has a sensibility 1.5 times lower than BGO, which is a disadvantage. It is also referred a non-proportionality of the light output to the deposited energy, possible different decay times in samples obtained from large and different ingots, and the presence of the naturally long-lived isotope (^{176}Lu) within the crystal, which leads to a measurable background count rate that may have impact for some applications [Humm, 03].
- **GSO**: one of its major advantages is a decay time of 60 ns. It also has good mechanical properties for cleaving. One important disadvantage is a low light yield, although very stable [Humm, 03].

Organic scintillators are aromatic hydrocarbon compounds containing linked or condensed benzene-ring structures [Leo, 94]. As mentioned before, they have a fast decay time and, in opposition to inorganic crystals, their fluorescence process is due to transitions in the

¹⁵ Self-activated scintillators, such as BGO, are those where the activator atoms are a major constituent of the crystal. There are also inorganic scintillators, as the NaI, that, in some conditions – low temperature, for example – are scintillators in their pure state.

energy level structure of a single molecule. Thus, such fluorescence can be observed from a given molecular species, independent of its physical state [Knoll, 89]. An example of a crystal organic scintillator is the anthracene. Plastic scintillators are usually considered as another type of organic scintillators.

1.2.2.2 – Photomultiplier Tubes (PMTs)

A Photomultiplier Tube (PMT) is a device that produces a pulse of electrical current when stimulated by very weak light signals, such as those produced by a scintillation crystal.

The front surface coupled to the scintillator is the entrance window, which, in the opposite side, is connected to the photocathode. From this photocathode, when struck by photons of visible light, are ejected photoelectrons. The conversion of light into electrons requires, nevertheless, a subsequent amplification stage, because the number of photoelectrons ejected from the photocathode is extremely small.

The charge amplification is granted by the dynodes/resistors arrangement plus a very stable high voltage supply, and is performed in a very linear way. Therefore, the output signal remains proportional to the number of original photoelectrons over a wide range of amplitude.

Some critical aspects regarding the PMT are:

- the photocathode's material, that must be sensible to the light produced in the scintillator crystal. For each material, sensibility is a function of the wavelength of incident light, and the curve describing this dependency is the spectral sensitivity characteristic. Some examples can be found in [Fhykt, 02], [Leo, 94] or [Knoll, 89]. The most used photocathode materials are antimony-caesium (SbCs), silver-oxygen-caesium (AgOCs), and the bi- and tri-alkali compounds SbKCs, SbRbCs, and SbNa₂KCs [Fhykt, 02]
- the work function¹⁶ characteristic of the photocathode's material. Incident photons must have more than a minimum value of energy (imposed by the work function) to release photoelectrons from the photocathode. Materials with low work functions should be chosen, but within an important trade-off: the work function should not be too small, in order to prevent the escape of electrons due to their thermal kinetic energy, i.e., to prevent the increase of thermionic noise.
- The dynode's material. A good material for a dynode should have a high secondary emission factor, stability of secondary emission effect under high currents, and a low thermionic emission (low noise). The latter effect is so much important as it is usually

¹⁶ The inherent potential barrier that exists in the interface between material and vacuum.

the most significant source of random noise from a PMT [Leo, 94]. Thermionic electrons are also the larger contributors for the dark current¹⁷.

The electron multiplication should be linear for pulses within the range of interest. A critical factor that strongly affects this required linearity is any deviation of dynode voltages from their equilibrium value, during the course of operation. Therefore, the high-voltage supply must be very stable: typically, a 1% increase in high voltage applied to the tube increases the amount of current collected at the anode by about 10% [Cherry, 03].

PMTs are also often wrapped in a metal foil for magnetic shielding, since the focusing of the electron beam from one dynode to the next can be affected by external magnetic fields [Cherry, 03].

A special type of PMTs are position-sensing photomultipliers, that provide some information about the location of the point of interaction between the incident photon and the photocathode [Knoll, 89; Humm, 03].

An alternative to PMTs is a special type of semiconductor photodiodes, know as avalanche photodiodes (APDs) [Knoll, 89; Gordon, 94]. An APD is a p - n junction in a silicon wafer that can be operated near breakdown voltage under reverse bias. The distinctive feature of such device is the acceleration electrons generated by interactions with ionizing radiation are subject to, due to the high electric field. These electrons gain enough velocity to generate free carriers by impact ionization, which are themselves sufficiently accelerated between collisions to create additional electron-hole pairs along the collection path. This process results in internal gain, increasing the signal and, therefore, improving the SNR.

1.2.2.3 – Block detector readout

The basic unit of a PET scanner is the block detector. A typical block detector is a piece of a scintillator crystal scored with a fine saw to create pseudo-individual crystals. This effect of pseudo-individual crystals is achieved since the scores provide a barrier to the optical dispersion of light inside the whole crystal. Moreover, scores are filled with a reflective material to reduce optical cross-talk between individual elements [Cherry, 03]. The number of individual elements per block is a characteristic of each PET scanner [Humm, 03].

The standard arrangement for coupling the PMTs assembly with the block detector is represented in Fig. 1.4 a: the light from each block is collected by 4 PMT, centered over each quadrant. Under this arrangement, if the signal from the four PMTs are A , B , C and D , then, to

¹⁷ It can be defined as the anode current measured without photocathode illumination [Leo, 94].

determine the detector element where the photon detection took place, the signals from the PMTs are combined as follows [Cherry, 03; Moisan, 97]:

$$X = \frac{(A+B)-(C+D)}{A+B+C+D} \quad \text{and} \quad Y = \frac{(A+C)-(B+D)}{A+B+C+D}$$

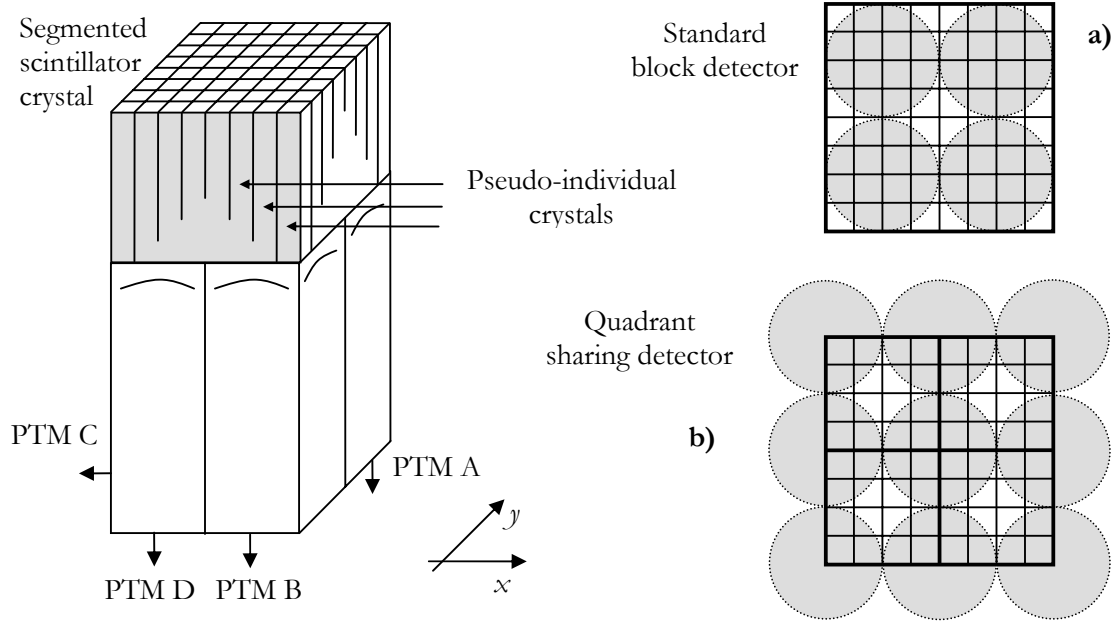


Figure 1.4 – A schematic representation of a block detector. a) Standard PMT arrangement in a standard block detector; b) Quadrant-sharing detector.

Another approach is denominated quadrant sharing and it is used for smaller detector blocks (Fig. 1.4 b): each block is also monitored by four PMTs, but each of these PMTs is centered in the corner of four different blocks.

1.3 – Data acquisition

In this section we discuss the process of collecting data and the two possibilities for doing that collection, known as 2D mode and 3D mode. We will also refer the strategy of data store: sinograms and projections.

1.3.1 – 2D and 3D mode for data acquisition

An important concept behind the distinction between the 2D and 3D acquisition modes is the *transaxial plane* (Fig. 1.2). A transaxial plane is a plane perpendicular to the scanner axis. So, neglecting the finite thickness of a ring, a transaxial plane can also be identified as the plane within which lies a detector ring. This means that along the direction of its axis, the scanner can be viewed as a stack of transaxial planes, each one corresponding to a detector ring.

Using the concept introduced in the former paragraph, the difference between a 2D and a 3D acquisition can be stated as follows: in a 2D acquisition data are collected for LORs that are within the same transaxial plane, i.e., each ring can be treated separately. In a 3D acquisition mode, in addition to transaxial planes, data are also acquired for LORs connecting detector elements in different rings: within oblique LORs, corresponding to polar angles different from zero. Thus, 3D mode contains all the information of a 2D acquisition, plus the information coming from the oblique LORs. This extra, and redundant data¹⁸, are useful for reconstruction purposes¹⁹, but can be huge enough to become unaffordable in terms of storage space.

The physical process for selecting just LORs within the same transaxial plane relies in the use of septa between adjacent rings, which is in practice equivalent to performing an axial collimation. So, in 2D mode these septa are present, while in 3D mode there are no septa at all. The existence of these septa is not shared by all the PET scanners available. There are scanners without septa that can only perform 3D acquisitions. Other scanners have retractable septa, allowing the user to choose among a 2D (with septa) or a 3D acquisition (septa retracted).

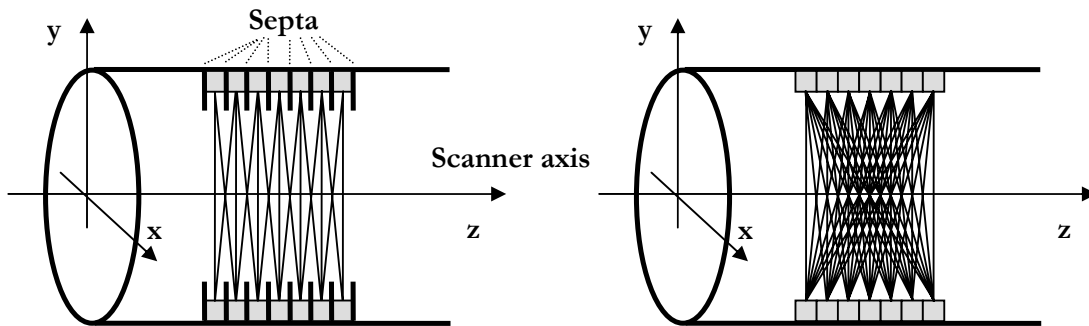


Figure 1.5 – 2D and 3D data acquisition schemes for PET.

We must nevertheless point out that the distinction that has been made between 2D and 3D acquisition modes is conceptual. In fact, to increase the axial sampling, the length of the

¹⁸ The image can be reconstructed just taking the 2D information, i.e., LORs within transaxial planes.

¹⁹ In particular, to improve the signal-to-noise ratio, as it will be discussed in the next chapter.

septum is adequately chosen in order to allow, even under a 2D acquisition mode, the acquisition of data along LORs connecting detector elements belonging to first neighbor rings (Fig. 1.5). These LORs lie within planes corresponding to polar angles slightly different from zero – the cross planes or inter-planes. Two cross planes corresponding to two adjacent detector rings cross the scanner axis at a point exactly halfway between the points of intersection of the scanner axis and the transaxial planes associated with each of these two rings. Thus, these cross planes can be added together to form a single virtual²⁰ transaxial plane located halfway between the real adjacent transaxial planes. This is the strategy for incrementing the axial sampling: if the scanner has n detector rings, the use of cross planes leads to a total of $2n - 1$ transaxial planes.

Using 3D mode raises a different question: if, theoretically, with a n ring scanner, it is possible to define n^2 data planes, in practice, however, due to the large amount of data this would produce, there are groups of planes that are mashed (added together) into a single plane. This compression strategy corresponds to a loss of information, but it guarantees that the large amounts of data can be stored and processed within reasonable times.

1.3.2 – Radial sampling

For describing the radial sampling one should describe how data are acquired within a plane. For the sake of simplicity, let us assume a given transaxial plane.

Since information corresponds to data acquired along LORs with different spatial orientations, either the identification of the detectors at both extremities of each LOR or the orientation of the LOR, using a reference frame, is stored. This latter approach is usually adopted, using a xOy reference framework defined within the plane. As depicted in Fig. 1.6, using that framework it is possible to define an azimuthal angle, ϕ . LORs with the same direction as the yy' axis correspond to an azimuthal angle equal to 0° ; other orientations are labeled with an azimuthal angle different from zero. ϕ can vary between 0° and 180° .

The width of the transaxial FOV is defined by the acceptance angle in the plane [Bailey, 03a]. This is determined by the electronics and is related with the permission of a detector to be in coincidence with other detectors in the opposite side of the ring (Fig. 1.8). The greater the acceptance angle, the larger the number of detectors which form the “fan”. The size of the intersection between the fan and the xx' axis determines the width of the FOV.

²⁰ In the sense that data were not acquired for LORs lying within such a transaxial plane.

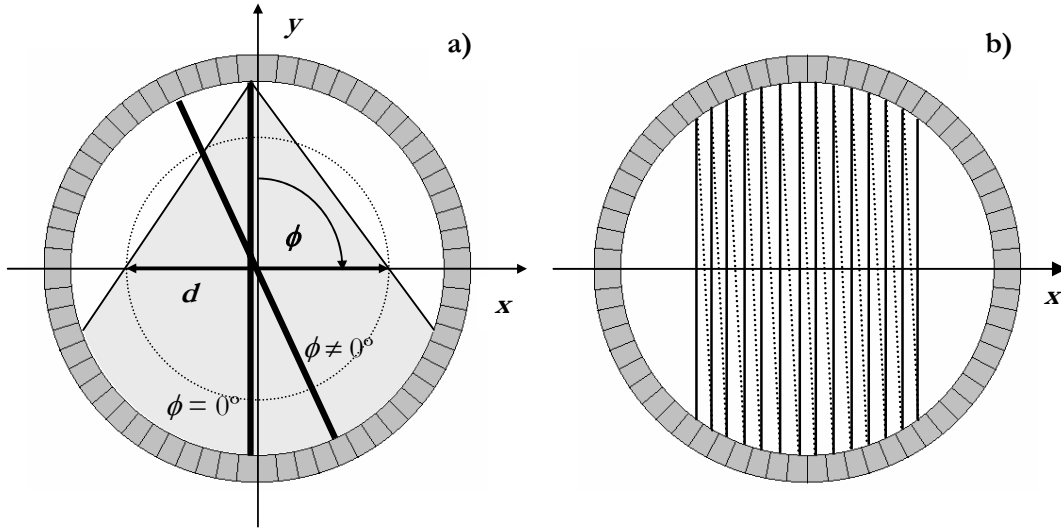


Figure 1.6 – The azimuthal angle and the radial sampling within a plane. (d – width of the FOV)

The use of a circular ring geometry leads to distortions in the sampling that should be considered before using data [Bailey, 03a], which are due to the fact that:

- the distance between the opposing detectors decrease towards the edges of the FOV;
- the lines of response are not evenly spaced in the projections; they get closer together for LORs farthest from the center of the scanner.

In spite of this, the sampling frequency along the radial direction is increased in the following way: in addition to LORs connecting detectors which, for a certain azimuthal angle, are directly in opposing positions (solid line in Fig. 1.6 b), i.e., $(D_a : D_b)$, $(D_{a+1} : D_{b+1})$, $(D_{a+2} : D_{b+2})$, etc., a double sampling is achieved by also considering LORs connecting a detector and the adjacent to the one which is in a directly opposing position (dashed lines in Fig. 1.6 b): $(D_a : D_b)$, $(D_{a+1} : D_b)$, $(D_{a+1} : D_{b+1})$, $(D_{a+2} : D_{b+1})$, $(D_{a+2} : D_{b+2})$, etc.

1.3.3 – The Michelogram

The Michelogram is a tool for representing the axial sampling of acquired data, devised by C. Michel. It is a grid combined with two axes, each one with a number of unitary marks equal to the number of rings in the scanner. The first ring in one of the extremities of the scanner is ring zero, and the other rings are sequentially numbered. Each point in the grid of the Michelogram corresponds to a pair of rings. Since LOR planes are defined within a ring or whenever two rings are put in coincidence, then each point in the grid is a conceptual

representation of a plane. Fig. 1.7 shows three different Michelograms corresponding to three different types of acquisitions.

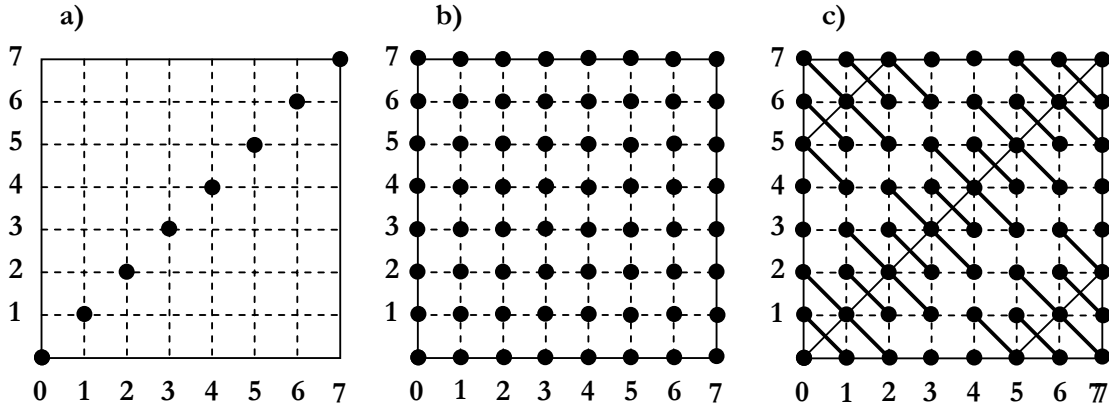


Figure 1.7 – Three examples of Michelograms corresponding to three different acquisitions: a) 2D acquisition; b) 3D acquisition without mashing; c) 3D acquisition with “mash”. Axial location is along the bottom left to top right diagonal direction.

The Michelogram in Fig. 1.7 a) represents a 2D acquisition: coincidences are allowed just for LORs inside each transaxial plane (each ring is exclusively in coincidence with itself), so points are along the diagonal. Fig. 1.7 b) is for a 3D acquisition with no restrictions, i.e., any ring can be in coincidence with any other. If the scanner has n rings, information will be available for n^2 different planes and, thus, there are n^2 points in the Michelogram. The acquisition represented by the Michelogram in Fig. 1.7 c) corresponds to a 3D acquisition with “mash”: a group of planes with the same axial position and neighbor values of ring differences mashed upon one single plane. Lines connecting points in the sinogram describe the mashing data were subjected to. This strategy leads to bands, or segments, in the sinogram. The intersection of the diagonal line at the center of the segment with each of the “mashing lines” indicates the axial position of the corresponding plane upon which data have been added. If such a point of intersection is not over a grid point, the scanner axis crosses the plane not within a real ring, but in a point halfway between two adjacent rings.

Whenever the mashing strategy is adopted, there is also another keyword: span. The span is the ring difference between the diagonal lines at the center of two neighbor segments. In the case of Fig. 1.7 c), the span is equal to 5.

1.3.4 – Sinograms and projections

Sinograms and projections are alternative methods for storing and viewing raw PET data. Fig. 1.8 schematically represents the process of data acquisition within a transaxial²¹ plane. For this single plane, there will be as many 1D projections as samplings for the azimuthal coordinate. Each of these 1D projections is a row along the radial coordinate, x' , with the acquired data stored in bins (their number is determined by the radial sampling).

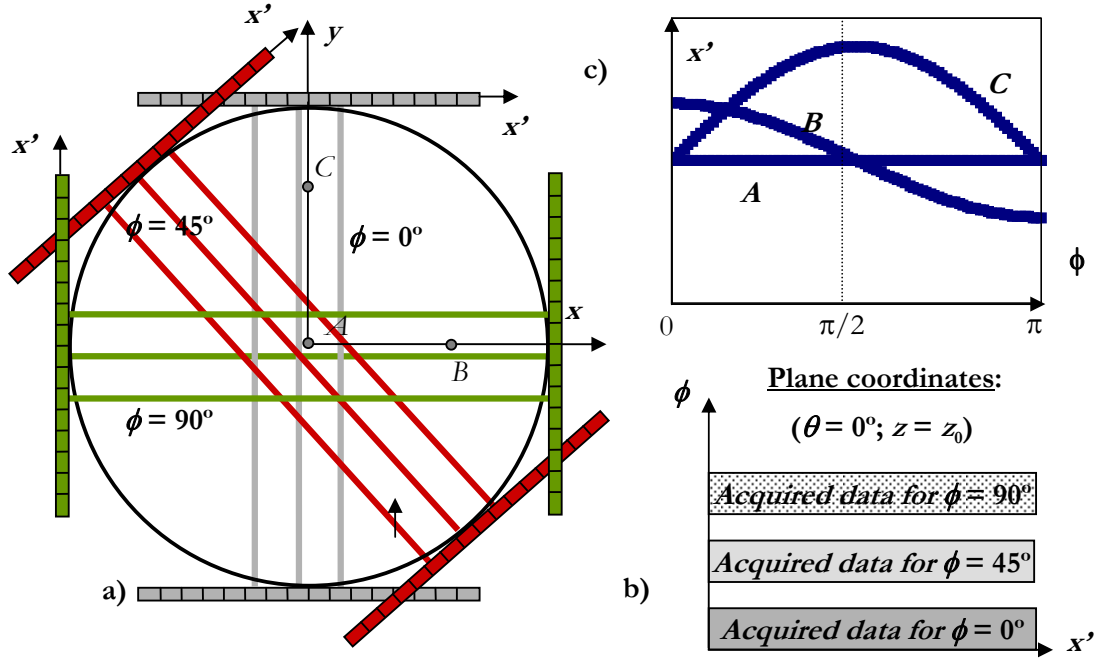


Figure 1.8 – a) Data acquisition within a transaxial plane ($\theta = 0^\circ; z = z_0$); b) Sinogram structure for this plane; c) The variation of the radial coordinates of three point sources as a function of the azimuthal angle.

A sinogram is a way to display the projections, where projection data at one view (azimuthal angle) are put in one row of the sinogram and the vertical direction represents the view angle: Fig. 1.8 b). The presence of a point source inside the FOV results in a sine wave in the sinogram (except if the source is positioned over the scanner axis, such as source A of Fig. 1.8 a). In Fig. 1.8 c) are represented three of such sine waves, each one describing the variation of the radial coordinate of a point source, as a function of the azimuthal angle. A detailed mathematical description of a sinogram structure can be found in [Barrett, 04].

²¹ A transaxial plane is used just for the sake of simplicity. A similar procedure can be adopted for planes characterized by a polar angle different from 0° .

A 2D projection for a certain azimuthal angle (view), is the image that directly results from the data stored in LORs described by the same azimuthal angle and the same polar angle.

According to the usual coordinate system (Fig. 1.9), an oblique sinogram can be parametrized as follows [Defrise, 02]: $s(x', \phi, z, \theta)$, where: $0 \leq \phi \leq \pi$, $-\theta_{\max} \leq \theta \leq \theta_{\max}$; and $|z| < (L/2 - R|\tan \theta|)$.

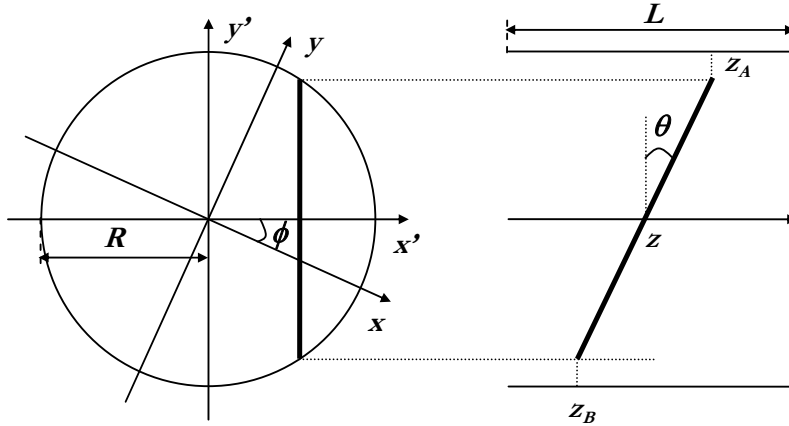
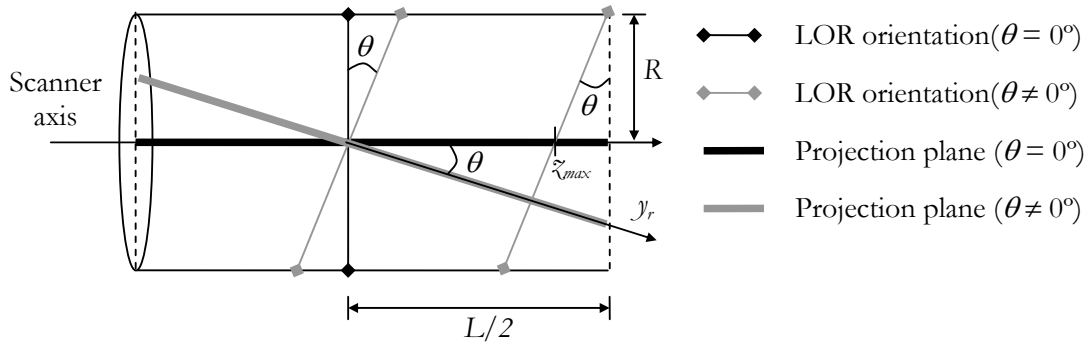


Figure 1.9

Coordinate system for sinogram parameterization. R is the ring radius, and L the scanner length.

The limits for the z coordinate of the sinogram are better understood when establishing the relationship between the axial projection coordinate (y_r) and axial sinogram coordinate (z). In Fig. 1.10 are represented two projection planes and the corresponding LOR orientation as a function of θ (the azimuthal angle can be assumed as 0° , for the shake of clarity).

Figure 1.10 – Link between projection coordinates (y_r) and sinogram coordinates (z).

From Fig. 1.10 one can observe that: $L/2 - z_{\max} = R \cdot \tan \theta$. This explains the range for the axial sinogram coordinate, z . In addition, it is also clear the link between the axial coordinates within a projection and a sinogram: $y_r = z \cdot \cos \theta$.

1.3.5 – Detected events in PET

The detected events in PET can be grouped into five categories: trues, randoms, scattered, single and multiple events (Fig. 1.11). Among them, the trues, randoms, and scattered events are known as *prompts*, i.e., coincidence events. From an ideal PET perspective, just the trues are not spurious: the other result from undesirable physical processes. Thus, for a correct image reconstruction from acquired data, it is necessary to identify the trues and adopt a strategy to handle the other events, either by performing some sort of correction or simply by discarding them. In this section the origin of these events is described, while in the next section the strategies for handling them will be presented.

The *trues*, on the base of which PET was conceived, correspond to photons detected within a coincidence time window, that have suffered no scatter and are the result of the same annihilation (Fig. 1.11 a). *Multiple* events result from more than one annihilation and correspond to the detection, within the same coincidence window, of three or more γ photons (Fig. 1.11 b). Since there is an ambiguity in deciding which photons make a valid pair (result from the same annihilation), these events are usually discarded by the system [Bailey, 03a].

A *single* event corresponds to the detection of a single photon which is unpaired within its coincidence window. This means that, for some reason (attenuation, FOV finite size, photon energy outside the energy window), one of the annihilation photons is lost. According to Badawi [Badawi, 00a], the singles flux is strongly dependent on the detector ring diameter, but only weakly dependent on the axial FOV, when this is greater than ~ 25 cm.

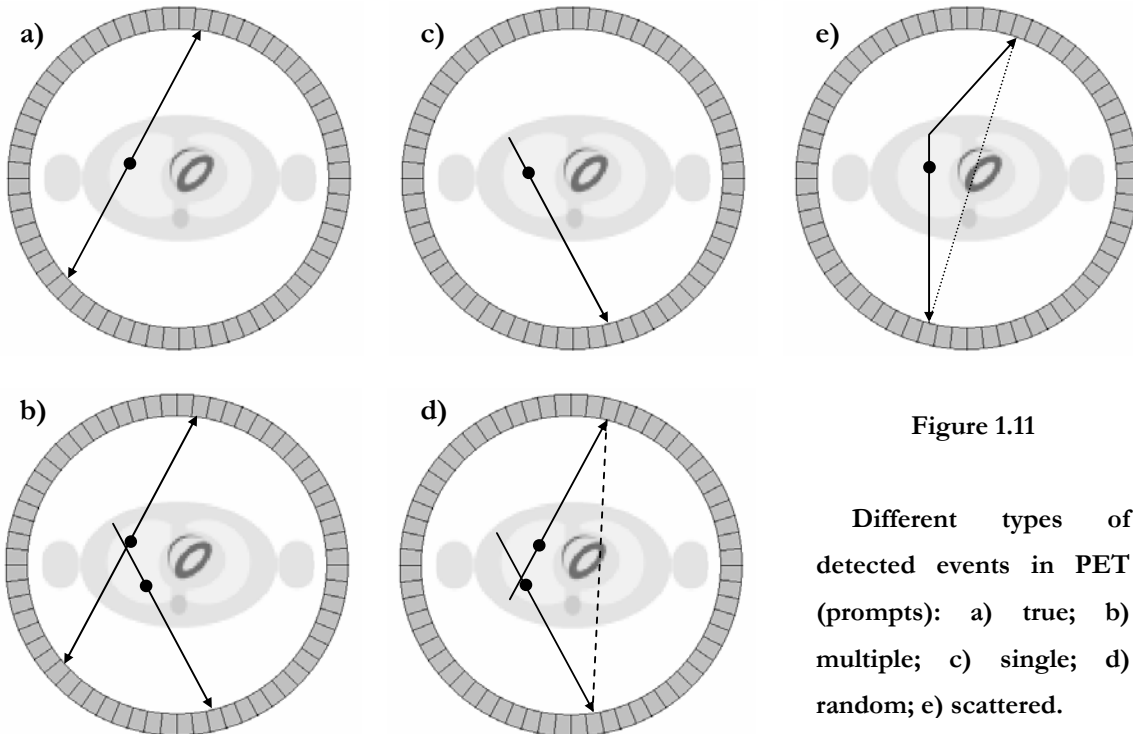


Figure 1.11

Different types of detected events in PET (prompts): a) true; b) multiple; c) single; d) random; e) scattered.

A *random* coincidence is the result of two positron annihilations taking place within the same coincidence window. However, while in a multiple event more than two photons are detected, in a random, from the four emitted photons, just two of them are detected (the other two are lost). Whenever there are two photons detected in coincidence, this is assumed as a valid event contributing to the image. The problem is that the LOR connecting the two detectors hit by these photons (Fig. 1.10 d) does not correspond to a real LOR, in the sense that no annihilation took place within this LOR. In another way: a random event induces the formation of an artificial LOR which is uncorrelated with the distribution of the tracer.

In face of this, two important things must be considered: how does the random event rate vary and from what does it depend; and, assuming the inevitability of having among the prompts a certain number of random events, how should they be handled before proceeding with the image reconstruction, i.e., how to correct the prompts for the randoms. The latter point will be discussed in the following section. Considering the first point, the random event count rate is a function on the number of disintegrations per second, and this count rate between two detectors, a and b , R_{ab} , is given by [Bailey, 03a; Hoffman, 81]:

$$R_{ab} = 2\tau \cdot S_a S_b \quad \text{Eq. 1.1}$$

where S_a and S_b are the single event rates for detectors a and b , respectively, and 2τ is the width of the coincidence timing window. Since $S_a \approx S_b = S$, then the random event rate increases approximately proportionally to S^2 . When the dead time is small, this means that the random event rate is roughly proportional to the square of the activity concentration.

It is also important to note that, unlike the trues, random coincidences can arise from activity outside the FOV [Spinks, 98]. Thus, the random coincidence count rate depends in a complicated way on both the source and detector geometry [Cherry, 03], and, according to Badawi [Badawi, 00a] is strongly dependent on both the axial FOV and the detector ring diameter. From Eq. 1.1 it is possible to verify the following:

- the greater the total amount of activity used in a study, the higher the ratio of random-to-true coincidence rates, due to the quadratic dependence of the random coincidences;
- the decreasing of randoms rate in proportion to the width of the coincidence window.

In actual PET scanners, the random-to-true ratio ranges from about 0.1 – 0.2 for brain imaging, to more than 1 for applications where large amounts of activity exist outside the FOV [Cherry, 03].

A *scattered* event occurs when one or both annihilation photons detected in coincidence have undergone Compton interaction. Due to the relatively poor detectors' energy resolution, there are scattered photons whose energy fits within the energy window operated by the scanner. Thus, when both photons (scattered and unscattered) are detected in coincidence, they will be treated as a prompt. As is the case of randoms, the consequence of this is the definition of an artificial LOR (connecting the two detectors hit by the photons), which is not correlated with the real activity distribution: if the photons suffers Compton scattering, it will be deflected from its original trajectory and will be no longer back-to-back with the other photon produced in the same annihilation.

Scattered photons are so more relevant as they can be scattered from outside into the FOV [Sossi, 95; Ferreira, 98] or even scattered by physical parts of the scanner, such as the gantry or others. An external end-shield was suggested [Hasegawa, 00] for screening radiation from outside the FOV. The use of septa between planes [Thompson, 88] is a strategy for reducing the likelihood of accepting photons scattered inside the FOV.

In 3D mode the number of scattered photons is much higher than in 2D mode: the scatter fraction – the ratio of scattered events to the total recorded coincident events [Thompson, 88] – is about 10% in the latter case and 30%-40% in 3D mode [Cherry, 03]. The scatter-to-true rate does not depend on the amount of activity administered, because both the scatter and the true coincidence rates increase linearly with this parameter. It also does not depend on the width of the coincidence time window [Cherry, 03]. In clinical studies, the scatter-to-true coincidence ratio ranges from 0.2 to 0.5 for brain imaging and from 0.4 to 2 for abdominal imaging [Cherry, 03]. The scatter fraction was found to be strongly dependent on the detector ring diameter, but only weakly dependent on the axial FOV [Badavi, 00a].

1.4 – Data corrections

This section is devoted to the description of some corrections applied to the data before image reconstruction. As stated before, these corrections are absolutely essential to grant PET imaging the accuracy and precision required for being considered a quantitative imaging tool.

If these corrections are to be grouped, one possibility is to organize them into four groups. One of them, includes the attenuation correction as a single member, since it arises due to the object itself, independent from all the rest. In a second group, the corrections for scatter

and dead time. These are corrections that, in principle, wouldn't be needed with an ideal scanner (ideal energy resolution and electronic processing capacity). In practice, this is not the case, so scattered events, whose rate is dependent from many factors such as the scanner configuration or the activity distribution, are detected and the respective correction required. A third group of corrections is directed towards the randoms: this sort of events can be reduced by narrowing the coincidence window, but the width of this coincidence window is always limited by the finite time speed of γ photons, i.e., it should be large enough to allow the detection in coincidence of the two photons produced in one positron annihilation, reflecting time-of-flight effects. This means that there is no possibility to completely avoid random events, whose rate is dependent, as seen, on the scanner geometry and activity distribution. The last group is reserved for the normalization and arc effect corrections, intrinsically inevitable whatever the adopted geometry. Even if ideal parameters were conceivable, they would still be needed. In practice, the arc effect correction is often included under the normalization.

Among the corrections listed above, the most important effect data should be corrected for is attenuation, which can affect both the visual quality and the quantitative accuracy of PET data [Kinahan, 03].

Before proceeding with a short description of the main strategies for correcting data, it is important to expose a model for the way these corrections affect the prompts. Mathematically, this model can be formulated as follows [Meikle, 03]:

$$y_i = (p_i - s_i - r_i) \cdot A_i \cdot D_i \cdot N_i \quad \text{Eq. 1.2}$$

where:

- p_i : the prompts measured along LOR i ;
- s_i : the scattered events estimated for LOR i ;
- r_i : the random events estimated LOR i ;
- A_i : the attenuation correction factor for LOR i ;
- D_i : the dead time correction factor for LOR i ;
- N_i : the normalization factor for LOR i ;
- y_i : the number of events for LOR i , after all corrections have been applied to prompts;

Sometimes, dead time correction is included under the normalization correction and, therefore, the corresponding factors are merged.

1.4.1 – Attenuation correction

As it will also be seen in Chapter 2, the way photons attenuation can be treated in PET, is not only very convenient from the point of view of image reconstruction²², but leads to a correction approach which is easy to implement.

When a photon beam passes through an absorber material, the number of photons transmitted throughout that absorber over a distance l , is given by (for homogeneous materials) [Knoll, 89]: $I(l) = I_0 \cdot \exp(-\mu \cdot l)$, where I_0 is the number of transmitted photons without the absorber, and μ is the linear attenuation coefficient, a quantity that is dependent on both the absorbing material and the photon energy. In the context of diagnostic imaging, the attenuation coefficient reflects, essentially, the sum of the probabilities associated with the photon interaction by photoelectric absorption and Compton scattering: for photon energies in the range of 30 to 50 keV, the dominant effect is the photoelectric effect; for energies between 200 and 1000 keV, it is the Compton scattering. Since PET imaging occurs at 511 keV, photon attenuation is determined by Compton scattering [Kinahan, 03].

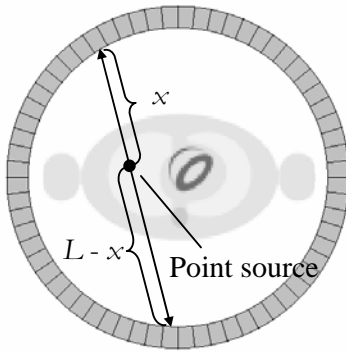


Figure 1.12 – The attenuation path length of two annihilation photons emitted from a point source inside the object. L is the total length of the LOR.

Considering a point source located inside the body (inside a PET scanner, Fig. 1.12) the probability that both annihilation photons will reach the detector (at two detector elements in opposing positions) is given by the product of the probabilities for their individual detection [Cherry, 03]:

$$P_{\text{det}} = e^{-\mu x} \cdot e^{-\mu(L-x)} \Leftrightarrow P_{\text{det}} = e^{-\mu L} \quad \text{Eq. 1.3}$$

where x is the trajectory's length for one photon and L is the total length of the LOR. Therefore, the probability that both photons will reach the detector is independent of the source location along the LOR. The problem of correction for attenuation in the body translates into determining the probability of attenuation for sources lying along different LORs.

²² It results in a multiplicative term, which can be easily handled by the reconstruction algorithm. This is strong contrast with SPECT, where the attenuation requires much more complicated approaches.

The widespread approach to correct for attenuation is the use of transmission measurements [Bailey, 98]. This approach is the combination of two measurements, both of them using a rotating source of radiation. The first is called the *blank scan* and is performed without the object inside the FOV. The idea is to determine the number of counts measured for each LOR (source and detector located in the two opposing extremities), when there is no attenuation along the path (except that resulting from the air). The second is the *transmission scan*, which is performed with the object placed inside the FOV, i.e., when photons suffer attenuation along the LORs. The *attenuation correction factor* for a certain LOR (corresponding to a certain set of values for coordinates x', ϕ, θ and z, i , when the transmission is done at 511 keV, is given by:

$$A_i = \frac{Blank_i}{Trans_i} \quad \text{Eq. 1.4}$$

where $Blank_i$ and $Trans_i$ are the counts in the blank and transmission scans for LOR i .

In order to obtain the transmission scans referred above there are three types of sources that can be used: positron sources, γ -ray sources and X-ray sources. The usual positron source for transmission scans is the $^{68}\text{Ge}/^{68}\text{Ga}$ source. One advantage of using this type of source is the possibility to reject random or scattered events, based on the knowledge of the source location (“triple-point” method or “sinogram windowing”) [Kinahan, 03; Meikle, 03]. In the group of γ -ray sources, ^{137}Cs is the most used [Kinahan, 03]. The third possible type of transmission sources are X-ray sources²³. Among the advantages of this last type is a low statistical noise and a fast acquisition; as disadvantage, there is the need²⁴ to convert the measured attenuation coefficients into the appropriate values for 511 keV [Kinahan, 03].

In alternative to the transmission approach, there is also the possibility of using CT reconstructed images of the object and proceed with a conversion method (such as segmentation or scaling), to convert the Hounsfield units into attenuation correction factors [Kinahan, 03].

To conclude, it should also be addressed the relevance of this correction, in particular for whole-body PET. In their paper [Bai, 03], Bai *et al.* summarize the main points raised by this discussion. Some perceived advantages of reconstructing without attenuation correction are: avoiding noise amplification, reduction of patient scanning time, avoidance of potential artifacts due to patient motion between transmission and emission scans; and improved contrast/noise ratios for lesions. However, the same authors argue, using methods for high-flux transmission imaging and a statistical transmission image reconstruction, the transmission scan does not significantly extend the total scan time, nor is emission image unduly amplified. Moreover, there

²³ This is the natural approach for PET/CT.

²⁴ Also true for the ^{137}Cs .

are important disadvantages resulting from not performing attenuation correction, such as inaccuracies in the uptake, shape and location of lesions, since the attenuation effect in PET is nonlinear. This is the case, for example, when scanning the thorax [Bai, 03]. So, whilst there are valid reason for reconstructing PET images without attenuation correction [Kinahan, 03]: checking for the presence of artifacts introduced by patient motion occurring between the emission and transmission scans and/or mechanical problems with the transmission scan; it is suggested [Bai, 03] that all studies should at least be reconstructed with attenuation correction to avoid missing regions of elevated tracer uptake.

1.4.2 – Correction for random coincidences

Random coincidences are due to the finite width of the coincidence timing window. The most evident consequence of randoms on a reconstructed image is the introduction of a relatively uniform background, reducing contrast and distorting the relationship between image intensity and activity of the object. Narrowing the coincidence timing window to avoid randoms can be part of the solution, but it must take into account the trade-off between minimizing the acceptance of randoms and the loss of sensitivity of true coincidences.

The common approach for correcting random coincidences is the use of a “*delayed window*”. This method estimates the random events rate, based on the fact that random events are not time correlated with each other. The principle, independent from being performed on-line (using the widespread delay line) or off-line (using a time tag associated with the photon arrival to the detector), is the following: whenever a photon strikes the detector, it triggers the opening of a coincidence timing window; if another photon strikes the detector ring within this window, this will be assumed as a prompt, since the detection is performed in coincidence. However, this second photon could be correlated with the first (resulting from the same annihilation - true event) or not (random event). If the trigger results in the overture of a coincidence window with the same width but delayed by a period of time much greater than the window itself, any photon striking the detector within that delayed window does not result, for sure, from the same annihilation as the one that has triggered the process: they are uncorrelated photons. Since the coincidence timing window has a fixed width, the estimated rate of uncorrelated events will be, on average, the same for that particular delayed window as for the former window without any delay, because the scanning conditions are assumed to be the same during the delay period. Thus, the delayed window counts provide an estimation of the random events rate.

With the knowledge of this rate, random events can be subtracted on-line from the prompts, during acquisition. This procedure, however, introduces noise in the measurement, because randoms are subtracted from the prompts, but the corresponding standard deviations for these measurements are summed.

In alternative, to avoid the increase of noise due to random subtraction, two techniques have been proposed for reducing the variance of the estimated random coincidences [Brasse, 05]: one is referred as the “smoothed delays” estimation method, and relies on the acquisition of the delayed coincidences in a separate sinogram, rather than on-line subtraction. The other method is the “singles-based” estimation of random coincidences [Casey, 86], under which the mean random coincidence rate $\langle R_{ij} \rangle$ for each LOR (detectors i and j) is calculated using Eq. 1.1.

An overview of variance reduction methods in 3D PET, is given by Badawi [Badawi, 99a].

1.4.3 – Dead time correction

The dead time is an intrinsic characteristic of every radiation counting system. It is related with the time required to process individual detected events, as the consequent individual pulse produced by the radiation detector has finite time duration. This duration is primary determined by the integration time, i.e., by the period of time the front-end electronics is integrating charge produced by the PMT.

There are many sources of dead time. One of them is the pulse pile up: radioactive decay is a random process and there is always a finite probability that successive events occur within the integration time corresponding to the electronic existence of the pulse triggered by the first event. In that case, because the second pulse occurs before the first has disappeared, the two pulses will overlap and form a single, distorted, one. This is critical under high count rates. So, to avoid the incorrect energy information carried by this distorted pulse, it should be rejected²⁵ and the two valid events discarded. This is an example of a dead time loss. Another source of dead time is the “reset” time, during which part of the system is unable to accept further events [Meikle, 03]. Multiple coincidences do also contribute to dead time, because the impossibility to ascertain which is the correct coincidence pair forces the rejection of all the events.

Due to dead time losses, the measured count rate is less than the true count rate. The possibility to measure the dead-time behavior as a function of count rate for a PET scanner, is

²⁵ This rejection is done using an upper energy level discriminator. Nevertheless, it is possible that the superposition of two individual events (scattered events, for example) results in a pulse whose energy lies within the energy window. In that case, the pile-up distorted pulse will be accepted and misinterpreted.

referred by Meikle [Meikle, 03]: the performance of a “decaying source” experiment using a uniform source containing a known quantity of a short-lived positron emitter such as ^{18}F or ^{11}C .

With respect to dead time, counting systems can be classified as “*paralyzable*” or “*nonparalyzable*”. Under a nonparalyzable system, if an event occurs during the dead time, τ , of a preceding event, the second event is ignored, without any effect on subsequently occurring events. If the same situation takes place in a paralyzable system, that second event is not counted as well, but introduces its own dead time during which subsequent events can not be recorded (there is an extension of dead time) [Cherry, 03].

The relationship between the measured count rate, R_m , and the true count rate, R_t , depends on the type of dead time. For a nonparalyzable system, the relation is given by

$R_m = \frac{R_t}{1 + R_t \tau_{np}}$, where τ_{np} is the nonparalyzable dead time constant. For a paralyzable system:

$R_m = R_t \exp(-R_t \cdot \tau_p)$, where τ_p is the paralyzable dead time constant [Cherry, 03; Knoll, 89;

Tai, 98]. A graphical comparison of the evolution of the measured and true rates, for the two dead time models, can be found, for example, in [Cherry, 03]. It is important to note the different behavior, at very high count rates, between two systems ruled by these models: while a

nonparalyzable system increases asymptotically towards a maximum value, $R_m^{\max} = \frac{1}{\tau_{np}}$, a the

measured count rate for a paralyzable system, at very high count rates, approaches zero, a phenomenon that is known as “counter paralysis” [Cherry, 03].

In practice, a system is made by different components. If one of them is dominant, it will determine the dead time model for the whole system. If, on the contrary, there is no dominant component, the dead time behavior will be a hybrid of the two models [Cherry, 03].

A possible approach to correct for dead time is to use the decay source measurements, to build a look-up table of dead time factors. This approach, however, does not account for spatial variations in source distribution that may account for different measured-to-true count rate ratios in different subsystems. Therefore, one solution is either to measure the dead time or to estimate analytically the model that better describes this parameter for each subsystem and, then, to establish a model incorporating the performance of each one of these subsystem [Meikle, 03].

1.4.4 – Normalization

The need for normalization arises in response to the question: when exposed to the same radiation source, will all detector pairs in coincidence record the same number of counts? In

other words, will the number of counts be the same for all the LORs? The answer to this question is no: some LORs will record more counts than other – different sensitivities –, for many different reasons, some of them discussed below. The normalization is, therefore, a correction applied to data in order to remove this undesirable effect.

1.4.4.1 – Effects contributing to differences in the sensitivity

An immediate source of differences in sensitivity is the efficiency of a detector element: while being part of the same detector block, detector elements can have different efficiencies due to their position within the block, physical variations among these crystal and light guides, as well as variations in the gains of the photomultipliers coupled to the block [Meikle, 03].

Other effects that contribute to differences in sensitivity are [Meikle, 03]:

- the accuracy with which detectors are aligned in the gantry;
- asynchronicity between detector pairs;
- the sum of adjacent data elements, either to simplify the reconstruction or reduce the size of data sets;

Another important effect that is usually included under the normalization procedure is the correction for the arc effect. This effect is a consequence of the ring configuration of the PET scanner: lines of response close to the edge of the FOV are narrower and more closely

spaced than those at the center (Fig. 1.13: $d_1 < d_2$) [Defrise, 98; Meikle, 03].

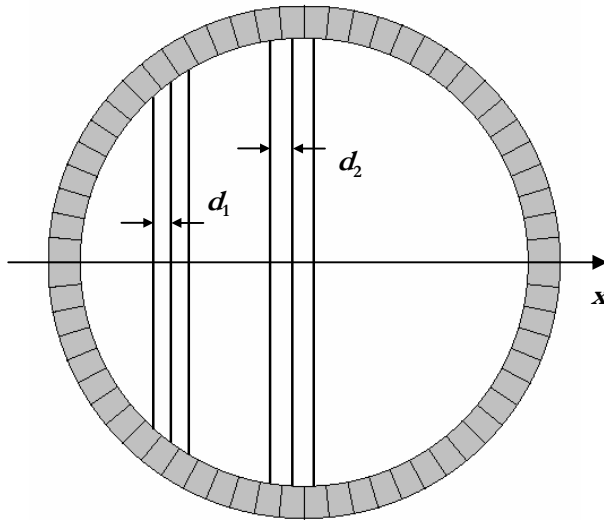


Figure 1.13

Illustration of the arc effect in a transaxial plane of a ring tomograph. LOR separation, d , is smaller at the edge of the FOV ($d_1 < d_2$).

The narrowing of the space between LORs results in a tighter acceptance angle and, thus, in a reduced sensitivity. However, for a transaxial plane, this effect is partially compensated by the fact that the separation between opposing detectors is less towards the edge of the FOV, contributing to an increase in the acceptance angle [Strul, 03]. Another effect of narrowing the space between LORs is the transaxial unequal sampling: the sampling distance is smaller at the

end of the FOV than at the center. Mathematically, this oversampling at the edges is described by equation Eq. 1.5 [Defrise, 98]:

$$d_i = \Delta r \sqrt{1 - \left(\frac{x_i}{R}\right)^2} \quad \text{Eq. 1.5}$$

where R is the scanner radius, Δr is the center-to-center spacing of the detector elements in the transverse direction: $\Delta r = 2\pi R/N_D$ (N_D is the number of detector elements in the transverse direction), d_i is the LOR spacing, and x_i is the transverse distance from the centre.

1.4.4.2 – Normalization procedures

The simplest and more obvious approach to perform normalization is to expose all detectors to an appropriate source, whose activity is well known. This is often done by using a planar or rotating source of ^{68}Ge [Meikle, 03], but other alternatives do exist [Bailey, 98b; Ferreira, 01]. With these sources it is possible to directly measure the efficiency of each detector. For a certain LOR, the normalization coefficient is assumed to be proportional to the inverse of the counts in each detector the LOR is connecting. This process is known as “direct normalization”. However, if it is theoretically obvious, it has some constraints and drawbacks [Meikle, 03]: in order to obtain statistically relevant data, very long scans (typically several hours) are required; the source – typically a rotating rod source or a rotating planar source - must have a very uniform activity concentration to grant a uniform exposure of all LORs. Otherwise, the resultant normalization coefficients will be biased: since the amount of scatter and its distribution in the normalization scan may be very different from that associated with a regular emission scan, this can result in bias and artifacts.

A different approach is to split the normalization into different components and treat each one of them separately: the component-based model for normalization [Casey, 95; Badawi, 99b]. The central idea is to express the normalization coefficients as a product of factors associated with the effects normalization should account for. One possible normalization model based on this principle is the following [Badawi, 99b]:

$$N_{ij} = \epsilon_u \epsilon_v b_u b_v c_{u \bmod D} c_{v \bmod D} d_{uvr} f_{uv} g_{uvr} h_{ij}$$

where N_{ij} are the normalization coefficients, i and j are detector indices, u and v are ring indices, D is the number of crystals along one side of a block detector, r is the radial position of a given LOR, k describes the relative position of the LOR within the block detector, and:

- ϵ_u : are the intrinsic crystal efficiencies, relating the average sensitivity of each crystal to the sensitivities of other crystals;

- b_u : are the axial block profile factors, describing the systematic variation in efficiency with the relative axial position of the crystal within each block;
- $c_{\text{mod}D}$: are the transaxial block profile factors;
- d_{uvk} : are the crystal interference factors, describing the modulation of the radial geometric factors by the relative position of the crystal in the block;
- f_m : are the axial geometric factors, describing the relative mean efficiency of the LORs in each sinogram compared with others after block profiles have been taken into account;
- g_{uvr} : are the radial geometric factors, describing the systematic variation in LOR efficiency with radial position in a particular projection;
- h_{win} : are the time-window alignment factors

This is one example of a possible decomposition of the normalization coefficients. The way each factor in the model above is measured is beyond the scope of this thesis. Detailed studies on this subject are, for example: [Ollinger, 95], [Badawi, 99b], [Ferreira, 01] or [Badawi, 00b].

1.4.5 – Scatter correction

Scattered events are detected in consequence of the poor energy resolution of PET scanners, which doesn't allow the discrimination of scattered from unscattered photons, based on their energies. An interesting discussion about the effectiveness of the energy resolution on the scatter reduction is provided in [Thompson, 93]: the author argues that is unlikely that improving the energy resolution, alone, reduces the detected scattered radiation up to the point where it can be neglected or provides the scatter reduction achieved by using inter-slice septa.

The use of scatter correction is especially important since the fraction of scattered events in PET is usually very high, especially in 3D mode: for a scan of the abdomen, it can be 60% to 70% [Cherry, 03].

In order to correct data for scattered events, there are some facts to be taken into account [Meikle, 03]: the scatter distribution is relatively featureless, containing mainly low spatial frequencies; the energy spectrum below the photopeak has a large contribution (whilst not exclusively) from scattered events; scattered coincidences that fall within the photopeak window are mainly due to photons that have scattered only once; LORs that don't cross the object (assuming the randoms have been previously subtracted) are due to scattered events, since an LOR from a true event represents an annihilation which took place inside the object.

From the knowledge of the former facts, many different approaches have been suggested for scatter correction. These approaches can be divided into four categories [Meikle, 03], shortly described in the following paragraphs: empirical scatter corrections, multiple energy window techniques, convolution and deconvolution approaches, and simulation-based scatter correction.

Empirical scatter corrections rely especially in the following facts: coincidences recorded outside the object are entirely due to scatter (assuming previously random subtraction); the scatter distribution is characterized by low spatial frequencies. Therefore, fitting an analytical function to the scatter tail outside the object, in projection space, it is possible to estimate the underlying distribution of scattered events. This approach is effective for neurological studies and among its advantages are its simple implementation and computational efficiency [Meikle, 03]. The main drawback is the fact that the scatter distribution is not always approximated by a smooth analytical function, particularly in the thorax, which can result in over or under subtraction (when using the approach described by Eq. 1.2).

Another empirical scatter correction is to make two measurements of the same object: one with the septa extended and the other with the septa retracted. If feasible (the required additional scan can be impractical), measures obtained with the septa extended, after scaling to account for differences in sensibility, can yield a measurement of the scatter in direct planes for a 3D acquisition. Assuming that the scatter distribution doesn't depend markedly on the polar angle (which is not always the case, especially for large axial FOV scanners), scatter contribution to oblique planes can be estimated by interpolation. At the end, a scatter distribution will be available for all the 3D planes and can be used for scatter correction [Meikle, 03].

Multiple energy window techniques rely on the energy resolution (whilst poor) of the detection system and in two important observations: most Compton scattered events are recorded in a region of the energy spectrum below the photopeak; and there exists a critical energy above which only unscattered events are recorded [Meikle, 03]. Dual Energy Window methods, as the name suggests, use, in addition to the photopeak window, an auxiliary energy window: below the photopeak, to directly estimate the scattered coincidences, an approach that is known as *Dual Energy Window* (DEW) method [Grootoonk, 96]; or above the photopeak, in order to collect just unscattered events – *Estimation of Trues Method* (ETM) [Ferreira, 02]. In both cases, the measurements in the auxiliary energy window are used to estimate the scatter contribution to the photopeak. The main advantage of these methods is that they consider scatter arising from activity outside the axial FOV. The main drawback are: scatter estimates based on Poisson measurements, which are noisy; the dependence of the spatial distribution of

the scatters from the energy of the scattered γ , i.e., the scattered distribution is not the same for the different windows [Meikle, 03].

The triple energy window method (TEW) is the natural extension of the DEW, that, in the implementation of Shao [Shao, 94], has two lower energy windows. This method has the main advantages and disadvantages of the DEW, but it improves the latter by reducing the sensitivity of the scatter correction to variations in source distribution and source size.

The third category for scatter correction are the convolution approaches. The analytical relation between scattered and unscattered data in 2D projections –the simplest but also the more intuitive – is the following, assuming that a spatial relationship exists between the scatter and the photopeak distributions [Bailey, 94; Meikle, 03]:

$$P_{sc}(s) = \int_{-\infty}^{+\infty} P_{unsc}(t) b(s-t, t) dt$$

where P_{sc} is the scatter distribution in a given slice, $P_{unsc}(s)$ is the one-dimensional projection of the true activity, and $b(s, x)$ is the scatter contribution to radial position s due to a source positioned in x . So, the core of these methods relies on the determination of function $b(s, x)$ and on the assumption that the substitution of P_{unsc} by P_{obs} – the measured projection data – is under a reasonable degree of accuracy²⁶.

The 3D extension of this method takes into account cross-plane scattering, which is achieved through the definition of a two-dimensional scatter response function and a two-dimensional convolution with the projections [Bailey, 94; Meikle, 03].

Deconvolution approaches, instead of a convolution, are aimed to estimate the distribution of unscattered data from measured projection, using a deconvolution operation.

Simulation-based correction methods use the knowledge about the photon interaction with matter to estimate the scatter contribution to projection. This can be done either via an analytical or Monte Carlo simulation. Extensive revisions of both methods as well as the other types referred in this section can be found, for example, in [Ferreira, 01] or [Meikle, 03].

All these corrections that have been described affect the performance of the PET scanner. The characterization of that performance, through the use of some figures-of-merit, is discussed in the following section.

²⁶ This model is for 2D data, where the scatter fraction is not so high as in 3D PET.

1.5 – PET camera performance

Aiming to standardize performance measurements of PET scanners, the National Electrical Manufacturers Association (NEMA) produced document of performance standards for PET scanners, 1994: the *NU 2-1994: Performance Measurements of Positron Tomographs*. In parallel, the European Union also developed measurements for PET scanners, leading, in 1998, to the International Electrotechnical Commission *EC Standard 61675-1: Radionuclide Imaging Devices – Characteristics and Test Conditions. Part 1. Positron Emission Tomographs*. It should be noted, however, that in spite of some overlap between the NEMA and IEC standards, many of the specific details of the measurements are different (for example, regarding phantoms, acquisition procedures, or data processing methods).

More recently, the NEMA standard has been updated, resulting in the standard that, while superseding the former NU 2-1994, it does not resolve the differences between the NEMA and IEC standards [Daube-Witherspoon, 02]. This is the *NEMA Standard Publication NU 2-2001: Performance Measurements of Positron Emission Tomographs; 2001*.

In practice, both the standards remain in use. However, it is easier to find in the literature studies describing scanners' performance comparisons using the NEMA 2-2001 standard. This is true even for studies carried on in Europe²⁷.

1.5.1 – Spatial Resolution

The spatial resolution of a system represents its ability to distinguish between two point sources in an image, i.e., it corresponds to the shortest distance these point sources can be placed so they will be identified in the image as two separated sources. This important parameter is usually characterized by imaging a point source or a line source and measuring the corresponding FWHM of the point spread function or line spread function, respectively.

In PET, the spatial resolution is usually measured within a transaxial plane as well as along the axial direction. For a transaxial plane, the resolution can also be divided into radial FWHM and tangential FWHM, for point (line) sources with an offset from the camera's axis. Among the factors that influence the spatial resolution in PET, are [Bailey, 03a; Stickel, 05]:

- non-zero positron range;
- non-collinearity of the annihilation photons;

²⁷ For example, [Bergmann, 05]

- distance between detectors;
- width of the detectors;
- stopping power of the scintillation detector;
- incident angle of the photon onto the detector;
- depth of interaction of the photon in the detector;
- optical decoding (the accuracy of the crystal identification by the four PMTs)
- number of angular samples;
- reconstruction parameters.

Due to the limited sampling in the axial direction – one sample per plane – it is inappropriate to measure axial resolution on such systems from profiles of reconstructed data as there are insufficient sampling points for granting an accurate estimation [Bailey, 03a].

The non-zero positron range is particularly relevant since there is no mean to avoid this physical effect. Since there are no effective methods to remove positron range blurring, this physical effect imposes, in fact, a limit on spatial resolution²⁸.

Positron annihilation distributions, in water and other living tissues, exhibit a cusp-like shape [Sánchez-Crespo, 04] (very sharp peak with long-range, long-intensity tails). The sharp peak preserves high spatial frequencies, but the tails are responsible for the blurring effects seen on PET images [Palmer, 05; Levin, 99]. A seminal paper concerning the contribution of the positron range to the spatial resolution of PET systems was published in 1986, by Derenzo [Derenzo, 86]. However, as Levin [Levin, 99] notes, the results referred by Derenzo should be taken carefully, as they were obtained for polyurethane foam, using a bin size ≥ 5 mm. Some recent works, referring values for quantifying the blurring due to positron range are:

- [Levin, 99]: in terms of the FWHM, for a ^{18}F point source, the contribution to the degradation of spatial resolution is about 0.1 mm; for ^{11}C , this contribution increases to 0.19 mm.
- [Palmer, 05]: considering an hypothetical imaging system with an intrinsic spatial resolution of 1.5 mm FWHM, for a line source of ^{18}F the FWHM was blurred to 1.67 mm; in the case of line sources of ^{11}C , ^{15}O and ^{13}N , the FWHM was blurred to 1.88 mm, 2.50 mm and 2.06 mm, respectively.
- [Sánchez-Crespo, 04]: the most distinctive feature of this study relies on the fact that the authors have investigated the positron range contribution to spatial resolution losses, not only for different radionuclides but also for various human tissues. It were

²⁸ The importance of the positron range depends from the radionuclide being used as well as from the tissue inside which the positron is moving [Sánchez-Crespo, 04].

found large differences, in image blur due to positron range, between human compact bone, soft tissue and lung tissue. The authors refer, for example, that in a PET camera with 3 mm system resolution, the loss of spatial resolution due to the positron distance of flight accounts for 1.3% in bone and 3.5% in soft tissue but up to 22.5% in lung tissue for ^{18}F . The corresponding values were, respectively, 2.9%, 9.3% and 45.0% for ^{11}C and 37.9%, 68.9% and 167% for ^{82}Rb . One of the major conclusions of this study is that positron range becomes a limiting factor in lung tissue, regardless of the choice of radionuclide, for high-resolution PET cameras with a spatial resolution in the order of 1 – 2 mm.

The use of very large magnetic fields (> 5 Tesla) to reduce the positron range effect in the PSF is explored in [Christensen, 95]. In that paper it is concluded that the spatial resolution of PET improves when positron annihilation takes place in a strong magnetic field.

The non-collinearity of the annihilation photons is the other unavoidable physical contribution to the degradation of spatial resolution. The effect on spatial resolution, FWHM_{180} , is linear dependent on the separation of the detector elements, D , and mathematically expressed by the following equation [Cherry, 03]: $\text{FWHM}_{180} = 0.0022 \times D$

The non-collinearity related blurring for a 80 cm (D) diameter whole-body PET scanner is about 2 mm, but for a 12 cm diameter small-animal PET scanner reduces to 0.3 mm ([Zanzonico, 04].

Considering, for example, the individual contribution to the degradation of spatial resolution due to the positron range ($\text{FWHM}_{\text{range}}$), non-collinearity ($\text{FWHM}_{\text{ncol}}$) and to the detector system (FWHM_{det} , which is itself the result of a combination of different factors), the combined blurring (FWHM_{sys}) results from the geometrical mean of the aforementioned factors.

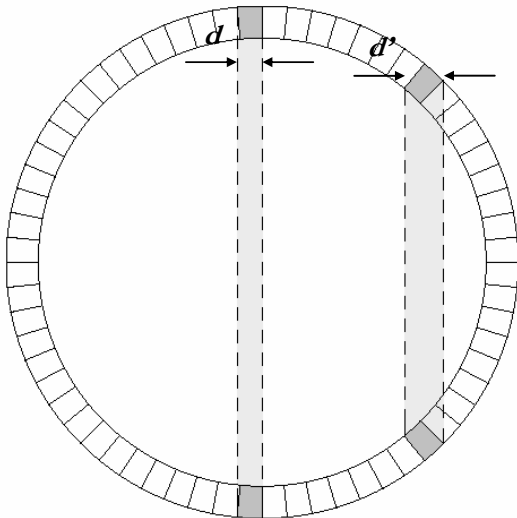


Figure 1.14

Parallax effect: the apparent width of the detector element (LOR) increases with the increasing of the radial offset.

Another important factor that should be taken into account is the depth at which each photon interacts with the crystal (DOI or parallax error - Fig. 1.14). This effect is due to the finite distance a photon travels within a crystal before depositing all of its energy.

For a source located at the centre of the scanner and a discrete detector, the intrinsic spatial resolution is essentially determined by the width of the detector element, Δ : $\text{FWHM}_{\text{det}} = d/2$ [Cherry, 03]. However, due to factors such as the depth of interaction of the γ rays within the crystal, for most of the PET detector systems there is no information about the depth at which the interaction took place within the crystal. In consequence, for a source in a location with an offset from the centre of the FOV, the apparent width of the detector, d' , increases. The result is a decrease of the spatial resolution along the radial direction, as one moves away from the centre to the borders of the FOV. For a whole-body PET scanner with 4 mm wide detectors and a diameter of 80 cm, the DOI effect contributes to a 40% degradation of the resolution at a distance of 10 cm from the center of the FOV [Cherry, 03]. A thin crystal with a high stopping power will contribute for the reduction of the distance traveled by the photon inside the detector and, therefore, to the reduction of the parallax effect. It should also be noted that the DOI effect depends on the scanner geometry, being different for a ring, hexagonal or octagonal system.

One possible strategy to deal with the DOI effect is to use the phoswich detector method that involves stacking thin layers of different detector materials with different decay times. The signal decay time will be used to identify the layer in which the interaction took place [Bailey, 03b].

1.5.2 – Energy resolution

In the context of nuclear medicine, the energy resolution can be defined as the precision with which a system can measure the energy of incident photons. Good energy resolution is necessary to achieve good image contrast and to reduce the background counts in the image. Primary causes for the degradation of the energy resolution are random statistical variations, including [Cherry, 03]:

- statistical variations in the number of scintillation light photons produced per keV of radiation energy deposited in the crystal;
- statistical variations in the number of photoelectrons emitted from the photocathode;
- statistical variations in the electron multiplication stage (dynodes) of the PMT.

In PET, in addition to the former definition – which is the single event energy resolution – it is also possible to define the *coincidence energy resolution* [Bailey, 03a]. One useful way to measure this coincidence energy resolution is to set one energy window for accepting a wide range of energies (for example, between 100 and 850 keV), while another is made narrow and stepped in small increments over the energy range. Using this method, it was demonstrated that the energy resolution for a line source of $^{68}\text{Ge}/^{68}\text{Ga}$, in air, is approximately 20% at 511 keV for a BGO scanner [Bailey, 03a].

The intrinsic energy resolution of some oxyorthosilicates scintillators (YSO, LSO, GSO and LGSO) can be found in [Balcerzyk, 00]. A model to predict the energy resolution of multicrystal encoding detectors for PET is proposed in [Vozza, 97].

1.5.3 – Sensitivity

The sensitivity of a PET scanner quantifies the ability of a scanner to detect the coincident photons emitted from inside the FOV. It is determined by four main factors: the scanner geometry, the detector efficiency, the acquisition energy window, and the dead time.

The scanner geometry establishes the total solid angle covered by the scanner over its FOV. For a ring scanner with diameter D and detectors whose thickness is b , if the small interdetectors area is ignored, the geometric efficiency decreases linearly from approximately d/D , at the centre of the scanner, to 0, at its borders, resulting in an average geometry efficiency given by $d/2D$ [Cherry, 03]. So, small diameter scanners with a large extension in the axial direction usually have higher sensitivities.

The detector efficiency is related with the probability that a photon whose trajectory intersects the detector will interact with it, i.e., will be detected. This is dependent from the detector material and its thickness (typically ranging from 20% to 70% [Stickel, 05]): for a single detector, the detector efficiency is given by $\varepsilon = 1 - \exp(-\mu_l x)$ [Cherry, 03], where μ_l is the linear attenuation coefficient for the detector material and x is the detector thickness. In the case of a coincidence system, ε should be replaced by ε^2 . An higher detector efficiency contributes to a higher scanner sensitivity.

The third factor affecting the sensitivity is the acquisition energy window, because events falling outside this window will be rejected²⁹.

²⁹ However, as referred, increasing the energy window will increase the scatter fraction

Finally, the dead time: it is another source of count losses, because some of the events are rejected by the processing chain. If sensitivity is measured at low activity, dead-time can be neglected. However, at realistic activity levels, this is not the case, and the sensitivity is lower due to the dead-time.

A high-sensitivity scanner collects more coincidence events in a fixed amount of time for a fixed activity within the scanner FOV. This is relevant, in practice, since it usually translates into an improved SNR for the reconstructed image, due to a reduction in the effect of statistical fluctuations [Bailey, 03b].

For a point source located at the centre of the FOV, PET systems sensitivities range from 0.2% to 0.5% (0.002 to 0.005 cps/Bq) in 2D mode, to 2% - 10% (0.02 – 0.10 cps/Bq) in fully 3D mode [Cherry, 03]. In face of these values it is noteworthy that while higher in comparison to other techniques (such as SPET), they are low in absolute terms: most of the potential signal available from the radiotracer is lost [Bailey, 03a].

Bailey et al. [Bailey, 97] introduced the effective sensitivity, defined as the product of the absolute sensitivity, C_{Abs} (cps.MBq⁻¹) and the ratio of the measured NEC (Noise Equivalent Count, as will be discussed in section 1.5.4) to the ideal true count rate with no scatter, random or dead time, T_{ideal} (Cps): $C_{eff}(a) = \frac{NEC(a)}{T_{ideal}(a)} \cdot C_{Abs}$. Therefore, the effective sensitivity is a function of the count rate, includes scatter, system sensitivity, dead time and random coincidences, and permits comparisons between different tomographs as well as the same tomography under different conditions.

1.5.4 – Noise Equivalent Count (NEC)

The Noise Equivalent Count Rate (NECR) [Strother, 90] is a mean for comparing the count rate performance of different scanners or the same scanner operating at different conditions. By definition, the NEC of a variable x is equal to the mean of the Poisson distribution with the same SNR as x [Nyys, 01]. In another way, it is the count rate which would have result in the same SNR in the data in the absence of scatter and random events [Bailey, 03a] - therefore, lower than the observed count rate. The widespread mathematical expression is the following [Strother, 90; Badavi, 96; Daube-Witherspoon, 02]:

$$NECR = \frac{T^2}{T + S + k.R} \quad \text{Eq. 1.6}$$

where:

- T : is the true event rate;
- R : is the randoms rate;
- S is the scatter rate;
- k : is a factor dependent on the way randoms are estimated: $k = 2$ if the scanner does on-line subtraction of randoms (i.e., the estimate is noisy); $k = 1$ otherwise (noise-free estimate).

So, for properly comparing different imaging systems, the parameters included in Eq. 1.6 should be established coherently. In particular, the k factor in the denominator must take into account the procedure adopted for the estimation of randoms. The randoms and scatter rates are given by [Daube-Witherspoon, 02]:

- $$R = R_T - \frac{T}{1 - SF}$$
- $$S = \frac{SF}{1 - SF} T$$

where SF is the intrinsic scatter fraction [Daube-Witherspoon, 02] – the ratio of scattered events to the total events, which are measured at a low count rate such that random coincidences, dead-time effects, and pileup are negligible - and R_T is the total count rate.

For 3D scanners, since the trues and scatter count rates are proportional to the activity, while the randoms count rate is proportional to the square of the activity, there exists an optimum activity for each scanner that corresponds to the maximum of the NEC [Zanzonico, 04]. A list of the peak noise equivalent counts for different scanners is shown in [Humm, 03]³⁰.

³⁰ A plot of the NEC curve for the HR+ scanner as a function of the activity concentration can be found in [Bailey, 03a].

2 – Image reconstruction algorithms in PET

In this chapter give an overview of some image reconstruction algorithms used in Positron Emission Tomography. A detailed description of all of these algorithms is beyond the scope this particular thesis. The ML-EM and OS-EM, the last one extensively used in our work, will observe special attention. Different algorithms may be preferred depending on factors such as the signal-to-noise-ratio, processing time constraints and, most importantly, the specific task for which the image is being reconstructed [Defrise, 03]. With this fundamental premise established, it is possible, from a theoretical point of view, to enumerate advantages of an algorithm over another, and vice-versa. Moreover, for a specific task, some algorithms have a better performance than others.

Also interesting are the four trends that, according to Jones [Jones, 03b], contribute to the increase of the computational demands in emission tomography: the incorporation of increasingly detailed physical models of the image formation; the increasing number of detectors in a single system – the authors state that the number of LORs in clinical tomographs doubles roughly every then months! -; the trend toward fully 3D image reconstruction; and the trend to combine multiple imaging modalities in single studies, such as in the case of PET/CT.

There is an extensive bibliography devoted to reconstruction algorithms, going back, at least, to the nineteen seventies. From those days now, some of the algorithms have been abandoned in practice, others were recuperated (i.e., iterative algebraic algorithms), a broad class of them is still in use (analytical algorithms) and many more were introduced and are still being investigated. The optimal reconstruction algorithm, however, still doesn't exist.

Guides to tomographic reconstruction algorithms are very useful. Examples are: [Ollinger, 97], [Defrise, 01], [Zeng, 01], [Lewitt, 03], or, more recently, [Qi, 06].

2.1 – The Radon transform in PET

As we mentioned before, positron emission tomography is based on the detection in coincidence of two 511 keV photons, approximately emitted back-to-back, resulting from the annihilation of the positron with some free electron in the object. Once detected, under the former co-linearity approximation, it is possible to assume that the annihilation took place somewhere in the LOR connecting the two detectors stricken by this pair of photons.

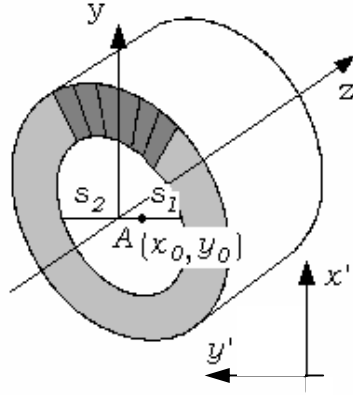


Figure 2.1

A single point source in the field-of-view of a PET scanner. $\overline{S_1}$ is the attenuation path of one annihilation photon; $\overline{S_2}$ is the attenuation path of the other photon.

Let us now, for the sake of simplicity, consider a transaxial plane and admit the existence of a single point source with activity $A(x_0, y_0)$, located at (x_0, y_0) (Fig. 2.1). Since we are assuming a single point source within the whole field-of-view placed somewhere along the line $\overline{S_1 S_2}$, during a timing interval T_e the corresponding element in the matrix (sinogram or projection), $\mathcal{E}(x', \phi)$, will record a number of events equal to:

$$\mathcal{E}(x', \phi) \propto A(x_0, y_0) T_e P_1 P_2$$

where P_1 and P_2 are the detector striking probabilities for each photon. Considering that the object is itself an attenuation medium and the attenuation coefficients are, in each point, given by $\mu(x, y)$, the former probabilities are given by:

$$P_1 = \exp\left(-\int_{s1} \mu(x, y) dy'\right) \quad \text{and} \quad P_2 = \exp\left(-\int_{s2} \mu(x, y) dy'\right)$$

and, then:

$$\mathcal{E}(x', \phi) = \mathcal{A}(x_0, y_0) T_e \exp\left(-\int_{s_1} \mu(x, y) dy'\right) \exp\left(-\int_{s_2} \mu(x, y) dy'\right) = \mathcal{A}(x_0, y_0) T_e \exp\left(-\int_s \mu(x, y) dy'\right)$$

with $s = s_1 + s_2$. Generalizing, if all the point sources along the line are to be considered (instead of a single source), the equation above leads to [Tofl, 96]:

$$\mathcal{E}(x', \phi) = \int_s \mathcal{A}(x_0, y_0) T_e \exp\left(-\int_s \mu(x, y) dy'\right) dy'_0 \quad \text{Eq. 2.1}$$

The most important fact in Eq. 2.1 is that the exponential depends only from the whole path traveled by both photons, s , and not from the particular point where the emission took place. This path is equal to the length of the straight line connecting the two opposite detectors and is equal for all the emissions occurring along this line. As a consequence, the attenuation factor behaves just as a multiplicative factor, so we can write:

$$\mathcal{E}(x', \phi) = T_e \exp\left(-\int_s \mu(x, y) dy'\right) \int_s \mathcal{A}(x_0, y_0) dy'_0 \Leftrightarrow \frac{\mathcal{E}(x', \phi)}{T_e \exp\left(-\int_s \mu(x, y) dy'\right)} = \int_s \mathcal{A}(x, y) dy'$$

or:

$$\frac{\mathcal{E}(x', \phi)}{T_e \exp\left(-\int_s \mu(x, y) dy'\right)} = \int_{-\infty}^{+\infty} \mathcal{A}(x' \cos \phi - y' \sin \phi, x' \sin \phi + y' \cos \phi) dy' \quad \text{Eq. 2.2}$$

The second member of equation 2.2 is the Radon transform (cf. Appendix A.2) of the function describing the activity of each point, for x' and ϕ .

We have, therefore, introduced the Radon transform in the context of PET and stated its relation with the measured data, for the 2D case. The tomographic reconstruction is achieved using a mathematical method to reconstruct the activity distribution (a 3D function) from projections (2D), i.e., a method to solve the following equation, which is a generalization for the 3D case of the problem we have been discussing [Defrise, 02]:

$$s(x', \phi, z, \theta) = \int f(x' \cos \phi - y' \sin \phi, x' \sin \phi + y' \cos \phi, z + y' \tan \theta) dy' \quad \text{Eq. 2.3}$$

2.2 – A primary distinction between reconstruction algorithms

There are many different approaches for image reconstruction from projections. However, according to their characteristics, and in particular to their underlying strategy for solving equation 2.3, it is possible to classify them. One of the most fundamental distinctions, even from a practical point of view, is between analytical and algebraic reconstruction methods. We can define analytical methods as those aiming at finding an inversion formula for the Radon transform (2D) or X-ray transform (3D), i.e., a direct solution to invert equation 2.3. Any discretization appearing on their implementation is not intrinsic to the nature of these methods, resulting from both the characteristics of the data they will be handling as well as from computational requirements. Nowadays, the most important analytical reconstruction algorithms are the Filtered-Backprojection and the 3DRP¹.

Algebraic methods, on the contrary, are conceived assuming discrete data, and instead of searching for a direct solution of equation 2.3, have an iterative process in their core, which converges to a solution for that equation. Discretization is not, as in the case of analytical methods, a contingency. Among algebraic reconstruction techniques it is also possible to devise another important distinction between statistical (such as ML-EM and OS-SEM) and non-statistical algorithms (as ART). Going further and taking the differences between the objective function (one of the requirements behind a statistical algebraic reconstruction) in each case, we can also distinguish Bayesian algorithms from non-Bayesian algorithms. Latter on this chapter we will go back to the objective function and to the algebraic reconstruction techniques.

Another very important difference is between 2D and 3D reconstruction algorithms. In fact, this is more than a distinction between algorithms, since for the same algorithm it is possible to have both 2D and 3D implementations. A 2D algorithm, or a 2D implementation of an algorithm, is one that assumes 2D data, independent from the fact that these data had been rebinned previously or acquired directly on 2D. A 2D reconstruction assumes data corresponding only to direct plans. 3D data impose no restriction to the data, and this sort of algorithms is able to reconstruct from sinograms (projections) comprising information either for direct plans as well as for oblique planes.

¹ 3DRP stands for 3D Reprojection. However, as it will be referred in section 2.3.3, the whole 3DRP reconstruction algorithm, as it is usually known, does not restrict to a 3D reprojection operation.

In Figure 2.2 we show a general representation of the possibilities for the reconstruction of 2D and 3D data.

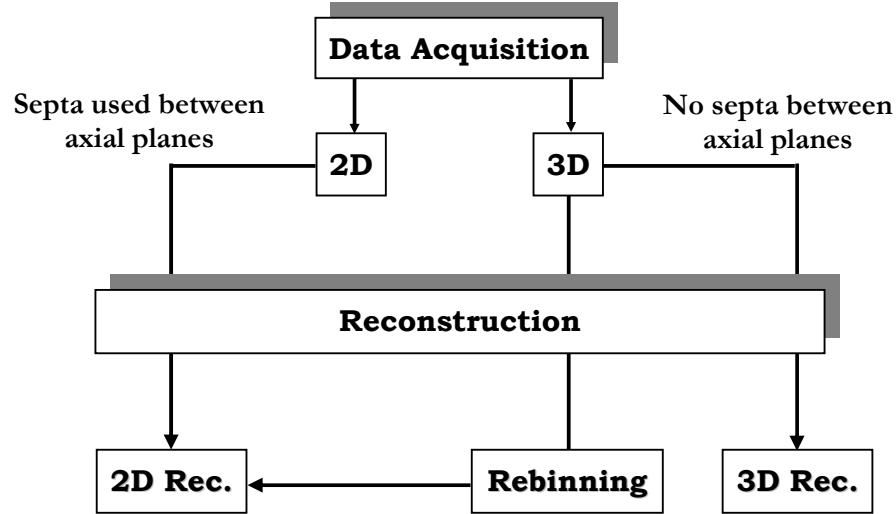


Figure 2.2 – A general representation of the reconstruction strategies for 2D and 3D data.

Lewitt and Matej [Lewitt, 03] adopt an apparently different, but very interesting, strategy to classify reconstruction algorithms in emission tomography. These authors take the fact that the obtained discrete measurements correspond to some integral transformation of a function of continuous spatial variables (the distribution of activity), $f(x, y, z)$. Thus, the process of data collection is represented by a discrete-continuous model (D-C model) relating the discrete data with that function. In addition, they also assume a linear, spatially variant relationship (LSV) between the basic physical component of the data and the function $f(x, y, z)$.

Once established a D-C model for the process of data collection, these authors suggest the existence of three different directions to proceed with the reconstruction:

- i) obtaining a discrete-discrete (D-D) model from the D-C model, representing the unknown function, $f(x, y, z)$, by a finite number of basis functions, each of them multiplied by a coefficient that is initially unknown and determined from the data by a specified computer algorithm;
- ii) obtain a continuous-continuous (C-C) model from the D-C model by interpreting the discrete data as samples of a function of continuous variables in the measurement space;
- iii) proceed with the D-C model and using a set of basis functions in the image space.

In fact, this is an elegant and alternative way to separate algebraic from analytical reconstructions: it is not difficult to identify the basis functions of the D-D model as being the

image voxels and, consequently, this approach with an algebraic reconstruction; analytical reconstruction fit in the C-C model. The direct solution of the fundamental D-C problem without introducing further discretization (D-C model) is reserved for a third category of algorithms not widely used in nuclear medicine. The latter sort of algorithms use basis functions such as “natural pixels” [Buonocore, 81; Hsieh, 96] or “ridge functions” [Kazantsev, 00].

For the sake of clarity, before proceeding with the description of some of the algorithms, it is important to stress the option we took in this chapter to distinguish between analytical and algebraic methods, as the framework for the organization of reconstruction techniques. We believe this is not only a more intrinsic distinction, but also more exclusive in the sense that no algorithm is simultaneously analytical and algebraic, whilst ML-EM and OS-EM, for example, have a 2D and a 3D implementation.

In appendix A.6 we show a diagram with one possible organization for the reconstruction techniques. The following pages are structured based on this diagram.

2.3 – Analytical reconstruction

The distinctive feature of an analytical reconstruction algorithm is the search for a direct solution to invert the Radon (2D) or X-ray (3D) transform (Eq. 2.3) or, in a more practical perspective, the search for an analytical solution to the problem of reconstructing from projections.

A rigorous mathematical analysis of the inversion of the Radon transform poses a bunch of problems whose complexity would better fit in a more mathematical leaning thesis. We are referring, for example, to the question of how to deal with a discrete set of projections rather than an infinite set, as would be required if a rigorous inversion is to be sought; or to the ill posed inverse problem translated in Eq. 2.3 and the related precision of measurements. A more detailed description of these questions could be found in [Deans, 83], as well as many bibliographic references where the subject is exhaustively discussed.

To proceed with this particular text, it is important to note that some algorithms proved to be useful methods for approximating the inversion of the Radon transform [Deans, 83]. These are what we consider under the designation of analytical reconstruction algorithms, and were devised to solve the problem of reconstruction from projections where the projection data constitute a sampling of the Radon transform. This fact, i.e., data as a set of samples of the

Radon transform, is very important since it is on the basis of different methodologies for the inversion of the Radon transform, and, consequently, leads to a further division of analytical techniques into two other categories: direct Fourier methods and frequency space filtering (signal space convolution are also part of this group).

2.3.1 – Direct Fourier Methods

As the name suggests, on direct Fourier methods (DFM) the core of image reconstruction is performed in Fourier space. Because projections are mathematically given by the Radon transform of the image, the central section theorem (cf. appendix A.4) emerges as the key tool for image reconstruction from projections. However, due both to the process of data acquisition and the computational implementation of reconstruction, a discretization of the Radon transform must be used. So, the central idea of performing a 1D Fourier transform of the projections and relating them to the 2D Fourier transform of the image translates, in practice, on the need to take discrete Fourier transforms (direct and inverse) via FFT (Fast Fourier Transform). This is the main problem associated with the DFM, because some sort of interpolation must be used.

The problem could be formulated starting by establishing the equations on a continuous base. As mentioned previously for the 2D case², projections are acquired for different azimuthal angles, ϕ , and for each angle data (which is the Radon transform of the activity distribution) is stored for different radial positions corresponding to different element of detection (bins). So, a polar coordinate system is the natural one to describe acquired data. Moreover, the central section theorem (App. A.4) - relating the 1D Fourier transform of a projection at an angle ϕ with the value along a line through the origin (at the same angle) of the 2D Fourier transform of the activity distribution – underlines the importance of polar coordinates on the link between image and projection. The 1D Fourier transform of a projection in the radial direction is:

$$P(\nu_{x'}, \phi) = \int_{-\infty}^{+\infty} p(x', \phi) \exp(-2i\pi x' \nu_{x'}) dx' \quad \text{Eq. 2.4}$$

The 2D inverse Fourier transform of $f(x, y)$ in Cartesian coordinates is:

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(\nu_x, \nu_y) \exp[i2\pi(\nu_x x + \nu_y y)] d\nu_x d\nu_y \quad \text{Eq. 2.5}$$

² The 2D case is used because the underlying problem is easier to describe and discuss.

If the continuous formulation was to be kept, we would be changing from Cartesian to polar variables in Eq. 2.5 and using the central section theorem to combine this equation with Eq. 2.4, in order to obtain $f(x,y)$ (as will be shown latter on this chapter) via a two-dimensional inverse Fourier transform.

Nevertheless, on a discrete basis, things are not straightforward because we are constrained to use a discrete Fourier transform and, from a computational point of view, this is almost equivalent to use the FFT. Since this algorithm is conceived for a Cartesian grid, Eq. 2.5 is discretized directly for these coordinates. On the contrary, the Fourier transform of data (line integrals) is given on a grid corresponding to polar coordinates (Fig. 2.3). Thus, in order to use the central section theorem, some sort of interpolation is needed to map the polar grid associated with the discretization of Eq. 2.4 onto the Cartesian grid assumed by the FFT. Different interpolation processes result on different algorithms and it was found that the interpolation accuracy is crucial to the quality of images reconstructed using direct Fourier methods [Stark, 81; Matej, 90; Gottlieb, 00; Stearns, 90; Choi, 98].

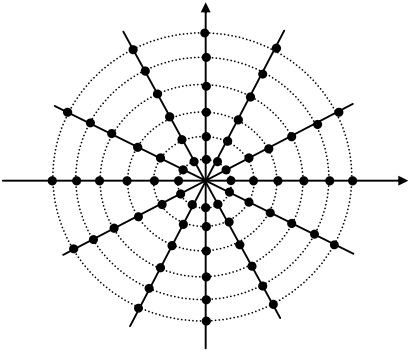


Figure 2.3 – Sampling in the frequency domain.

Based on the previous paragraphs, an algorithm to reconstruct $f(x,y)$ from its projections via DFM can be derived from the following steps [Cherry, 03]:

- i) take projection profiles at N projection angles $\phi_i, i = 1, 2, \dots, N$;
- ii) compute de 1D FFT of each profile;
- iii) insert the values of these FTs at the appropriate coordinate locations in the Fourier space. Values are inserted in polar coordinates along radial lines through the origin (v_x);
- iv) using the values inserted in polar coordinates, map these values onto a Cartesian grid in Fourier space;
- v) use the interpolated values in Fourier space and the inverse FFT to compute the image of the object, $f(x,y)$.

Several types of interpolation (corresponding to step iv in the former sequence) were or have been studied. Among the results described in literature, we retain that the nearest

neighbour and 2-D linear interpolation yield very poor results (serious artifacts) [Natterer, 01; Matej, 90]. Gridding methods [Natterer, 01; O’Sullivan, 85; Jackson, 91] were introduced to obtain optimal inversion schemes of Fourier data and are presently considered the most accurate Fourier reconstruction method [Natterer, 01; Choi, 98].

2.3.2 – Filtered Backprojection

The filtered backprojection (FBP) algorithm is one of the most important reconstruction algorithms. According to Ollinger and Fessler [Ollinger, 97], the first application of FBP in the context of medical imaging can be attributed to Shepp and Logan [Shepp, 74]. Nowadays, is still used in many situations as a gold standard for comparing the quality of reconstructed images. The popularity of FBP is historically due to its relative light computational requirements comparing to other methods such as DFM or iterative algorithms³, when scanty computational power was an almost inevitable limitation. Independent from this very practical aspect, analytical algorithms, and FBP in particular, are linear and, thereby, allow an easier control of most of their well known properties (spatial resolution and noise correlations, for example), something crucial for quantitative data analysis [Defrise, 03].

The filtered backprojection is equivalent to the direct Fourier reconstruction in the limit of continuous sampling [Defrise, 03]. Differences between these two approaches arise when the continuous formulation is replaced by the necessary computational discrete implementation.

According to the order by which filtering and backprojection are used (both are linear and shift-invariant so their order can be interchanged), there are two variations of this method [Brooks, 76; Defrise, 98]: the backprojection-filtering, where data are first backprojected and then filtered; and the filtered backprojection itself, where filtering is applied prior to backprojection. In the following paragraphs we will be discussing the latter approach. Among recent works regarding FBP, we found very useful for the comprehension of FBP, as well as for their relevance in the whole context of image reconstruction, the following: [Defrise, 98; Defrise, 03; Barrett, 04]. [Brooks, 76] and [Kak, 88] are cited very often and can be assumed as classical texts.

To establish the filtered backprojection algorithm we take again Eq. 2.5, i.e., the equation for the 2D inverse Fourier transform of function $f(x,y)$:

³ Iterative algorithms were devised before the introduction of FBP. However, by the time, their computational demands made them not very interesting.

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(v_x, v_y) \exp[i2\pi(v_x x + v_y y)] dv_x dv_y$$

Changing from Cartesian to polar coordinates ($v_x = v \cos \phi$ and $v_y = v \sin \phi$), we have:

$$f(x, y) = \int_0^{2\pi+\infty} \int_0^{2\pi} v F(v \cos \phi, v \sin \phi) \exp[i2\pi v(x \cos \phi + y \sin \phi)] dv d\phi$$

and using the fact that $F(v, \phi) = F(-v, \phi + \pi)$, then:

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{+\infty} |v| F(v \cos \phi, v \sin \phi) \exp(i2\pi v x') dv d\phi \quad \text{Eq. 2.6}$$

where $x' = x \cos \phi + y \sin \phi$. Now, the central section theorem is used (cf. Appendix A4) to introduce in the integrand a known function (projection data). Combining this theorem with Eq. 2.4, we have:

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{+\infty} |v| P(v, \phi) \exp[i2\pi v x'] dv d\phi \quad \text{Eq. 2.7}$$

In the equation above, $v_{x'}$ is identified with v . Up to this point, and once we kept the continuous formulation, the former equations also describe the direct Fourier methods. As we said before, since DFM methods proceed in the Fourier space, the use of the 2D FFT algorithm in a discrete implementation forces an interpolation from polar to Cartesian coordinates. On the contrary, the filtered backprojection splits Eq. 2.7 into two distinct stages: filtering in the Fourier domain, first; backprojection, after.

$$\begin{aligned} 1^{\text{st}}) \text{ Filtering: } \quad \bar{p}(x', \phi) &= \int_{-\infty}^{+\infty} |v| \left(\int_{-\infty}^{+\infty} p(x', \phi) \exp(-i2\pi x' v) dx' \right) \exp(i2\pi x' v) dv \\ &= IFT_{v \rightarrow x'} \{ |v| FT_{x' \rightarrow v} [p(x', \phi)] \} \end{aligned} \quad \text{Eq. 2.8}$$

In the filtering stage, for each azimuthal angle is calculated the 1D Fourier transform of the projections (Radon transform of function $f(x, y)$) for that angle. Then, in the Fourier domain, the signal is multiplied by a high-pass ramp filter, $|v|$. Finally, a 1D inverse Fourier transform is used for going back to the Radon domain.

$$2^{\text{nd}}) \text{ Backprojection: } \quad f(x, y) = \int_0^{\pi} \bar{p}(x \cos \phi + y \sin \phi, \phi) d\phi \quad \text{Eq. 2.9}$$

Eq. 2.9 is the mathematical expression of the backprojection operator. This stage is equivalent to the integration along a sinusoid in the Radon domain.

According to this, the implementation of the FBP algorithm can be based on the following steps [Defrise, 98]:

- i) Fourier transform the projection for a given angle ϕ : $P(\nu_{x'}, \phi) = F\{p(x, \phi)\}$;
- ii) Filter the projection in the frequency space: $\bar{P}(\nu_{x'}, \phi) = |\nu| P(\nu_{x'}, \phi)$;
- iii) inverse Fourier transform the filtered projection: $\bar{p}(x', \phi) = F^{-1}\{\bar{P}(\nu_{x'}, \phi)\}$;
- iv) backproject the filtered projection: $B\{\bar{p}(x', \phi)\}$;
- v) repeat steps 1 - 4 for each ϕ : $0 \leq \phi < \pi$.

Observing the equations above, a first remark should be made: in eq.2.8, the Fourier transform of the projection is multiplied by the ramp filter. However, this would have as consequence the amplification of high frequencies. Since the power spectrum of a typical image usually decreases with increasing frequencies, and noise [Riederer, 78], on the contrary, has likely more significant components in this region of the spectrum [Farquhar, 98], this sort of filtering would translate, in practice, in the amplification of noise. This is usually referred as an evidence of the ill-posed problem in the context of the inversion of the Radon transform: a small perturbation in the signal, due to noise in the measurement, can cause an arbitrarily large error in the reconstructed image [Defrise, 03]. The solution for the reconstruction using FBP is, then, not to use the ramp filter, but to stabilize the process by introducing a low-pass apodizing window. So, Eq. 2.8 is replaced by Eq. 2.10:

$$\bar{p}(x', \phi) = \int_{-\infty}^{+\infty} |\nu| H(\nu) \left(\int_{-\infty}^{+\infty} p(x', \phi) \exp(-i2\pi \cdot x' \nu) dx' \right) \exp(j \cdot 2\pi \cdot x' \nu) d\nu \quad \text{Eq. 2.10}$$

The mathematical definition of some of the most common apodizing windows can be found, for example, in [Herman, 79]

In respect to the discrete implementation of FBP [Defrise, 03] we note that available data are made of N_ϕ projections (corresponding to j azimuthal angles), each one sampled along the radial direction at intervals equal to Δr . This radial sampling constraints the choice of the cut-off frequency of the filters to be used in the filtering step (Eq. 2.9): for a radial sampling distance Δr , the Shannon's theorem states that the maximum frequency that can be recovered without aliasing is $1/2\Delta r$. In addition, the choice of ν_c is also determined by the photon statistics, a factor has a major role in the noise characteristics of data. So, when setting the cut-off frequency there is always a trade-off to be considered: a filter with a cut-off frequency that is too high, favors resolution and contrast but allows noise to degrade image resolution; a filter with a

too low cut-off frequency reduces image noise, but may overly smooth the image, decreases contrast and eventually introduces ring artifacts [Farquhar, 98].

To conclude this section dedicated to FBP, we shortly refer [Wilson, 93] that FBP reconstruction spreads noise from image regions containing high count densities into regions of low count densities, so the noise magnitude (variance) remains relatively constant throughout the image. At low FBP filter cut-off frequencies the noise is correlated at relatively long distances.

2.3.3 – The 3DRP algorithm

The importance of the 3DRP, an analytical algorithm for fully 3D image reconstruction from projections, derives from the specific characteristics of 3D data usually available. Due to the finite axial extension of whole-body PET scanners, some LORs corresponding to oblique planes are missing, i.e., these lines cross the object but just one of their extremities intersects a detector bin. The other extremity would correspond to a point of intersection whose axial coordinate is outside the field of view. So, from a reconstruction perspective, this is equivalent to have truncated projections (for the axial coordinate) and, in consequence, the conditions for the applicability of the 3D central section theorem are not fulfilled.

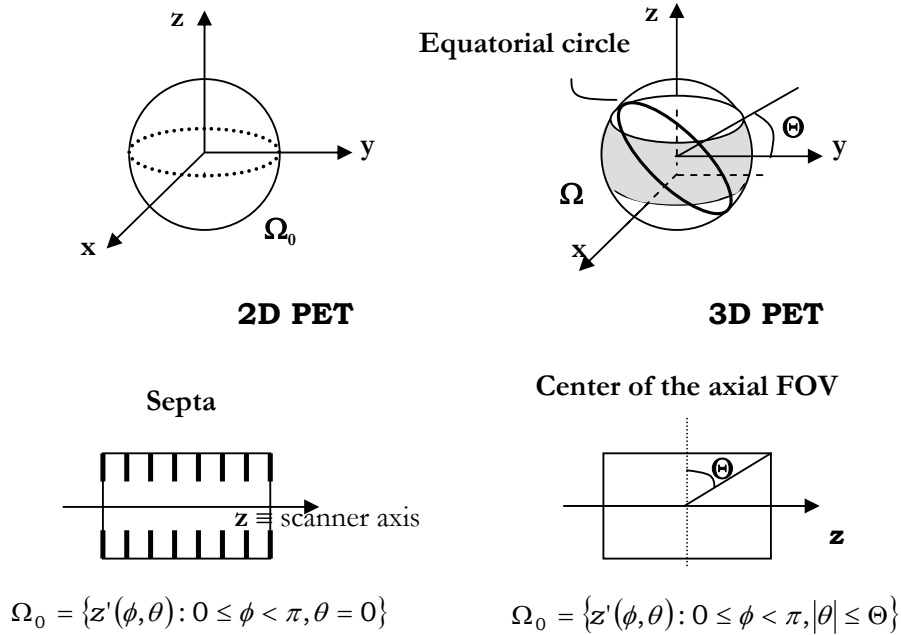


Figure 2.4 – Orlov Spheres for different acquisition modes (2D and 3D) in a conventional cylindrical whole-body PET scanner.

Another important difference between 2D and 3D PET is data redundancy for the latter case. In 3D PET there are additional data corresponding to the oblique planes, that can be used in the reconstruction to improve the SNR by reducing statistical noise [Defrise, 98; Kinahan, 89].

Conversely, we can also ask how many parallel projections provide enough information to reconstruct the 3D image in a stable way. The answer is contained in Orlov's condition [Defrise, 95a]: it is possible to recover $f(x, y, z)$ in a stable way if the set of measured projections, Ω , is intersected by any equatorial circle on the unit Orlov's sphere. This is illustrated in Fig. 2.4, where the orientation of a single 2D projection plane can be described by the location of the end point of the unit vector $\hat{z}'(\phi, \theta)$ (parallel to the LORs corresponding to that projection), on the sphere. For each mode of acquisition (2D and 3D) the resulting set of projections is shown.

Assuming we have complete data, i.e. satisfying Orlov's condition, remains the problem of truncated data for the axial coordinate. 3D truncated PET data results on the direct inapplicability of the 3D central section theorem, and, in consequence, a 3D version of the filtered backprojection (3D FBP) is not directly usable. On the other hand, shift variant characteristics of these data make standard Fourier deconvolution techniques rather inappropriate to solve the problem of reconstruction. In this context, the 3DRP, as a whole, is the reconstruction algorithm that bridges the gap between truncated and shift variant data and the applicability of the central section theorem, via an intermediary reprojection (RP) step.

If data were complete and non-truncated, it would be theoretically possible to reconstruct using the 3D FBP. The 3D FBP [Defrise, 98; Defrise, 02] could be understood as the 3D counterpart of the 2D FBP. An algorithmic implementation of the 3D FBP can be resumed as follows:

- i) calculate the 2D Fourier transform of $p(x, y, \phi, \theta)$, $P(v_x, v_y, \phi, \theta)$;
- ii) multiply by a 2D filter, H_C , and a 2D window, W :

$$\bar{P}(v_{x'}, v_{y'}, \phi, \theta) = P(v_{x'}, v_{y'}, \phi, \theta) H_C(v_{x'}, v_{y'}) W(v_{x'}, v_{y'})$$

- iii) take the inverse 2D Fourier transform to obtain the filtered projection:

$$p(x', y', \phi, \theta) = \int_{-\infty}^{+\infty} dx' \int_{-\infty}^{+\infty} \bar{P}(v_{x'}, v_{y'}, \phi, \theta) \exp[i2\pi(x' v_{x'} + y' v_{y'})] dy'$$

- iv) 3D backproject the filtered projections;
- v) repeat steps i-ii for each ϕ : $0 \leq \phi < \pi$;
- vi) repeat steps i-iii for each θ : $-\Theta \leq \theta < \Theta$.

For the 2D filtering operation in step ii), a factorisable filter is usually adopted [Defrise, 02]. The most often used is the Colsher's filter [Defrise, 98; Colsher, 80].

Let us consider Fig.2.5, where the existence of truncated projections is illustrated. The higher the value of the polar angle, the higher the number of truncated projections. This is synonymous of a drop in the scanner sensitivity when moving from the center of the axial FOV towards the edges. In consequence, applying 3D FBP to truncated projections results in severe artifacts that cannot be corrected by simply dividing each point of the FBP reconstruction by the sensitivity in that point [Defrise, 98].

The 3DRP [Kinahan, 89] is the widespread analytical algorithm for the reconstruction of truncated 3D data. The first step in 3DRP is to complete 2D projections by estimating the truncated projections. This is done by calculating the set of line integrals these projections are made of, taking a preliminary reconstructed image of the object, $f(x,y,z)_{2D}$, obtained from a 2D reconstruction from data in direct planes. This is possible since in normal conditions no direct projections are missing, and a 2D FBP can, hence, be used.

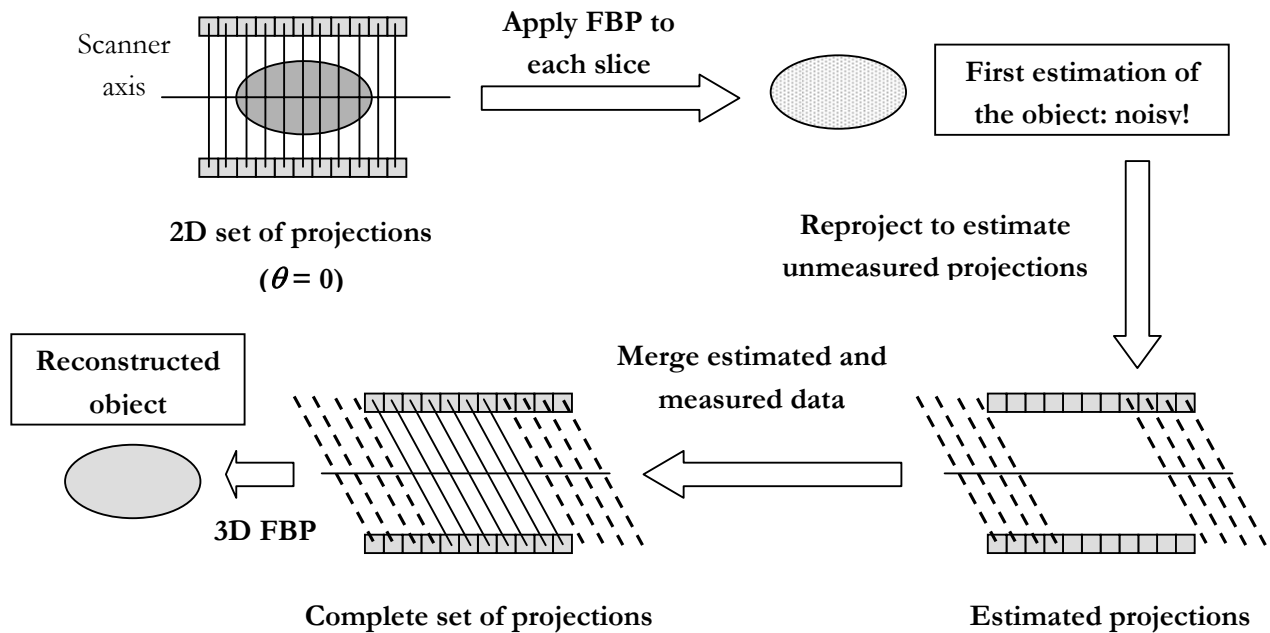


Figure 2.5 – The 3DRP algorithm. The starting point is a 2D set of projections extracted from the acquired data.

After estimating the missing projections, they are stacked with the former acquired projections in order to obtain a set of complete projections. These complete projections are finally used to reconstruct the image via a 3D FBP.

The 3DRP algorithm can be depicted as in Figure 2.5, and the corresponding algorithm stated as follows:

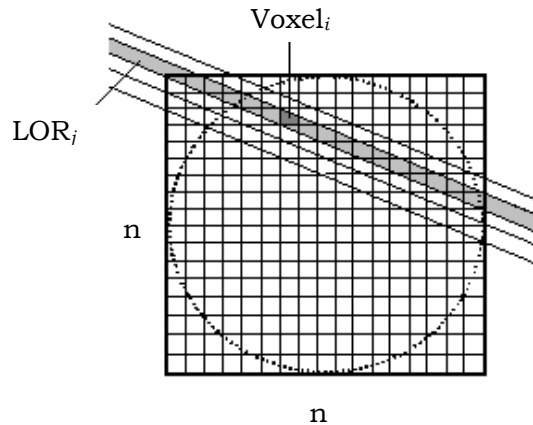
- i) extract the 2D data (direct planes) from the full data set;
- ii) using the former 2D data, reconstruct a first image, $f(x, y, z)_{2D}$, by applying the 2D FBP algorithm to each slice;
- iii) for each projection (ϕ, θ) , with $0 \leq \phi < \pi$ and $|\theta| \leq \Theta$:
 - estimate the truncated projections;
 - merge the estimated and measured data, so a complete set of projections is obtained;
- iv) use the 3D FBP algorithm to reconstruct the image.

2.4 – Algebraic reconstruction

Algebraic techniques are the alternative to analytic approaches, in the context of image reconstruction from projections. As referred at the beginning of the chapter, in spite of being classified as algebraic techniques, this sort of algorithms⁴ are very different from each other. However, a fundamental characteristic is transversal to all of them: they are not conceived for a direct inversion of Eq. 2.3 but, instead, they progressively refine the estimation of the activity distribution, $f(x, y, z)$. In addition, data are assumed to be intrinsically discrete, which is in sharp contrast with the analytic reconstruction techniques. This is so much important as we saw that discretization is on the basis of the main problem related to the algorithms that would be otherwise the natural analytic approach, Direct Fourier Methods, and their replacement in practice by FBP.

Figure 2.6

A schematic representation of the algebraic D-D model.



The D-D (discrete-discrete) model describing the algebraic approach can be depicted as in Figure 2.6. Object discretization is obtained by considering it

⁴ As an introductory reading see, for example, [Vandenberghe, 01; Leahy, 00b].

divided into small volume elements: voxels. Let y_j be the total number of counts in LOR $_j$, m the total number of voxels contributing to this LOR, and λ_i the activity of voxel i . Then:

$$y_j = \sum_{i=1}^m a_{ij} \lambda_i$$

where a_{ij} is the probability of a photon emitted from voxel i being detected in LOR $_j$, i.e., it is a model for the mapping from source to detector. If one has n LORs, the former equation becomes a matricial equation with n incognitas, which is, in principle, invertible. Since the subset of voxels, λ_p , contributing to a bin (LOR) is not necessarily the same for all bins (it gradually changes as one moves from a bin to its neighbor), when establishing the whole set of equations for all the n bins, all the voxels, m , will also be involved in these equations. In the matricial approach, the total number of values a_{ij} are stored in one matrix usually known as system matrix, **a**, whose number of rows (m) and columns (n) equals the number of voxels and bins, respectively. Due to the way the elements of **A** are defined, this matrix represents what is usually called the forward projector.

Moreover, moving from the field of analytical to algebraic techniques is equivalent to step forward from the integral in Eq. 2.3 to a discrete sum: the general goal of an algebraic reconstruction technique is the estimation of the activity of each discrete voxel, from the available discrete acquired data. Mathematically, this could be described by the following equation (the discrete counterpart of Eq. 2.3):

$$\lambda_i = \sum_{j=1}^n (a^{-1})_{ij} y_j \quad \text{Eq. 2.11}$$

where n is the number of LORs to which voxel i contributes, and **a**⁻¹ is the back-projection operator.

Shared by all the algebraic techniques is their iterative nature. In fact, the algorithmic translation of “progressively refining” is an iterative process of estimation-comparison-update (Fig. 2.7), which means that Eq. 2.11 will not be directly solved, but one will be iteratively searching for the best solution within some previously established criteria.

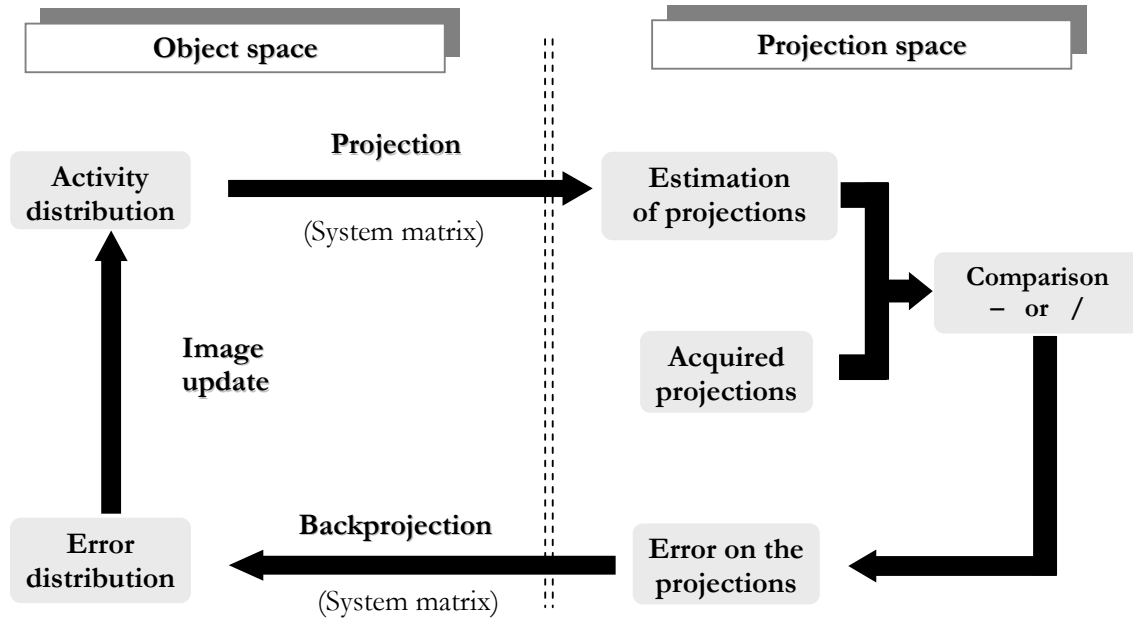


Figure 2.7 – Flow chart of an iterative image reconstruction technique

In general, all techniques developed for 2D PET can, in principle, be applied to 3D data. Nevertheless, their performance with 3D data is expected to be very different from the performance with 2D data: the amount of data in 3D will inevitably slacken the iterative process.

After establishing the fundamental distinction between algebraic and analytical reconstruction as well as stating the general equation that drives the iterative process in the core of algebraic techniques, it is now important to identify the main requirements behind this sort of techniques. They are in the number of five [Fessler, 94; Fessler, 02]:

1. a finite parameterization of the object (image), usually represented as a discrete set of voxels;
2. a model of the measurement, i.e., a system model, \mathbf{a} , that relates the unknown image to the expectation of each detector measurement;
3. a model of the measurement uncertainty, i.e., a model of the probability distribution of each measurement around its expectation value;
4. an objective function. This function gives a measure of how well an image fits the data and how well this image matches the desired image properties (*a priori* constraints);
5. an algorithm, typically iterative, for maximizing the objective function, including specification of the initial estimate and stopping criterion.

For each requirement, different choices have been adopted. The result is a great number of reconstruction techniques, differing, at least, in one of the former components. Thus, a

classification of the algebraic techniques would be, in principle, based on the choices made in relation to these requirements. Unfortunately, this is not always the case, and it is not difficult to get lost in front of the extensive list of exquisite names for all the algebraic techniques. Due to its importance, we devote next section to a more detailed study of this subject.

2.4.1 – Five components of an algebraic reconstruction method

2.4.1.1 – A finite parameterization of the image

This is the distinctive feature of the algebraic methods and the very core of the D-D model. Mathematically, discretization is equivalent to a finite parameterization of the object, $f(\vec{r})$, achieved using a set of basis functions, $b_j(\vec{r})$:

$$f(\vec{r}) \approx \tilde{f}(\vec{r}) = \sum_{k=1}^r x_k b_k(\vec{r})$$

Basis functions can be classified into a few categories, such as localized or global and linear or nonlinear. An extensive list of these functions is available in [Fessler, 02]. Fourier series are an example of a linear, global set of basis functions. However, we are especially concerned with linear and localized basis functions, the group accounting for both voxels and blobs (or Keiser-Bessel window functions).

When introducing the discretization inherent to the algebraic techniques, we have anticipated the importance of voxels, associating that discretization with a division of the object into small volume elements called voxels. In fact, voxels are the *de facto* standard for parameterizing images, having the very important property of minimal support i.e., the existence of no overlap: the voxel basis function has a value equal to the unity inside a small cube and zero outside [Fessler, 94].

In contrast with the FBP method, which is, in principle, able to reconstruct images with arbitrarily fine pixel grids, for statistical iterative reconstruction methods the size of the voxels is a matter of consequence: if these methods are unregularized, voxel sizes that are too small can lead to over-parameterization and numerical instability; conversely, too large voxels can produce mismatch and loss of image features [Fessler, 94].

Blobs [Lewitt, 92; Matej, 96] are alternative smoothed basis functions, linear and localized. They are a two-parameter family of functions, with spherical symmetry and bell-shaped radial profiles.

2.4.1.2 – System matrix

As said before, the element a_{ij} of the system matrix, \mathbf{a} , denotes the probability that an annihilation in the j th voxel is detected by the i th detector pair. To efficiently represent the system matrix, and in particular to reduce the storage size, this matrix is usually factorized into the following components [Qi, 98a; Qi, 98b; Leahy, 00a]:

$$\mathbf{a} = \mathbf{a}_{\text{det.sen}} \cdot \mathbf{a}_{\text{det.blur}} \cdot \mathbf{a}_{\text{attn}} \cdot \mathbf{a}_{\text{geom}} \cdot \mathbf{a}_{\text{positron}} \quad \text{Eq. 2.12}$$

where:

- $\mathbf{a}_{\text{positron}}$: includes the effect of the distance traveled by the positron, before annihilation with an electron. This distance is dependent from the specific isotope and the density of the surround tissue.
- $\mathbf{a}_{\text{det.sen}}$: is a diagonal matrix containing the detection efficiency of each detector pair. For this factor contribute the intrinsic sensitivities of individual crystals, the relative position of the crystal within a detector block, and geometric factors related to the distance of the detector pair from the centre of the FOV. All of these factors are measured through calibration. System dead-time can also be included in this matrix. One possible expression for the determination of this matrix components can be found in [Qi, 98b].
- $\mathbf{a}_{\text{det.blur}}$: this factor acts as a local blurring function applied to the sinogram. Blurring occurs for three primary reasons: the not exact co-linearity of photons; the scattering of photons from one crystal to another, resulting is a mis-positioning of the detected photons; and the fact that the crystal surface is not always orthogonal to the direction of arrival, so that a photon may penetrate through one or more crystals before being stopped. Monte Carlo techniques are often used to establish the contribution of each of the former effects to this matrix.
- \mathbf{a}_{attn} : it accounts for the effect of attenuation. Since the probability of attenuation is approximately constant for all photon pairs emitted along a certain LOR, attenuation factors can be represented by a diagonal matrix.
- \mathbf{a}_{geom} : contains the geometrical mapping between the source and data, i.e., between voxels and bins. There are various methods [Zhuang, 94] for establishing the values of each of the elements, a_{ij} , of this matrix. One of them is known as the Siddon's algorithm [Siddon, 85] and was devised for X-ray CT. According to this algorithm, the value for a_{ij} is proportional to the length of intersection between LOR j and voxel i . This method has

two main drawbacks [Fessler, 02]: in general, it does not preserve counts and leads to artifacts. Since this method is centered on the ray (LOR), in the sense that intersected voxels are determined based on ray direction, it is usually known as a ray-driven method. From this point of view, in opposition to ray-driven are voxel-driven methods [Peters, 81]. In that case, a voxel, k , is taken, and for a direction defined by the azimuthal angle ϕ , an imaginary line crossing the center of this voxel is drawn. Once it is likely that this line is not coincident with any of the LORs along which data were acquired, for the calculation of a_{ik} there will be considered N LORs parallel and close to this imaginary line. The distance between each one of these N LORs and the initial imaginary line crossing the centre of voxel k could be used to establish a_{ik} , $i = 1, \dots, N$.

An alternative approach to Siddon's algorithm adopts the volume of intersection between LOR j and voxel i . It demands a greater computational effort, but it is more accurate than the previous one since it can exactly preserve the total number of counts (all the areas can be considered) [Fessler, 02]. A third alternative is to use the solid angle spanned from the voxel k to the faces of the detector pair i , for establishing matrix element a_{kj} [Chen, 91; Qi, 98a]. Recently, De Man and Basu [DeMan, 04] introduced the distance-driven approach.

Involved in the forward and backprojection, \mathbf{a}_{geom} will be part of a huge number of multiplication operations. Hence, the way this matrix is represented will be critical to the efficiency of the reconstruction algorithm. Given the number of LORs and voxels taking part in a 3D acquisition, a raw storage of \mathbf{a}_{geom} would require an enormous and prohibitive storage space. However, a close inspection of this matrix reveals two important features: this matrix is very sparse and exhibits a high degree of symmetry. The use of symmetries is a way to avoid redundancies and makes storage affordable. It is possible to identify the following symmetries [Qi, 98a; Johnson, 95; Chen, 91]: in-plane rotation symmetries, resulting from rotating the projection rays by $\phi = 90^\circ, 180^\circ$ and 270° and a 45° reflection symmetry, result in a reduction factor of eight; axial reflection symmetry (reflection of a chord with respect to plane $z = 0$) provides an additional factor of two for ring differences other than zero; parallel axial chord symmetries are justified under the assumption that the axial point spread function is invariant to axial position. In that case, an additional reduction factor of $(N_R - s)$ is gained, where N_R is the number of detector rings and s is the ring separation of rays in the sinogram. To render concrete the reduction achievable using the former

symmetries, let us consider a scanner with 18 rings and a maximum ring separation allowed of 11 rings. The reduction factor, R_f , is equal to 188:

$$R_f = \frac{((7+8+9+10+11+12+13+14+15+16+17) \times 2 + 18) \times 8}{12} = 188$$

The accuracy on establishing the elements of each one of the former matrices will provide us with a realistic system matrix as a whole, modeling, as good as possible, the acquisition and detection process. This, on his side, determines the accuracy of the reconstruction. Nevertheless, the better achievable model for the system matrix could be unaffordable from a computational point of view, thus giving room for approximate models whose trade-off between accuracy and computation demands is more favorable.

Another possibility for determining the elements of the system matrix is to use a Monte Carlo simulation. This is discussed, for example, by Veklerov [Veklerov, 88] and Rafecas [Rafecas, 04]. The main idea is to directly incorporate all physical effects into the elements of the system matrix.

2.4.1.3 – A model of the measurement uncertainty

The choice of a model for the measurement of uncertainty, or, in other words, for the noise, is on the basis of one of the most clear distinction between algebraic techniques. Even the absence of such a model is, by itself, a distinctive feature. In fact, as shown in Appendix A6, if no model is assumed for the noise, an algebraic reconstruction technique can be classified as *non-statistical*, in opposition to *statistical iterative reconstruction techniques* (SIR), a group including all the algebraic techniques incorporating any statistical model for the data. ART (Algebraic Reconstruction Technique) is an example of a non-statistical algorithm. Among SIR methods, differences rely on the statistical model adopted.

Before proceeding with the discussion, let us link this model for the noise with the two previous components of the algebraic techniques: the finite parameterization of the object and the system matrix. Using voxels and the system matrix, as well as the main assumption that acquired data are discrete, one can establish a deterministic model [Fessler, 02]:

$$y_i \approx g_i([a\lambda]_i) \tag{Eq. 2.13}$$

where g_i is a function accounting for the noise. If r_i is the mean number of random counts for the i th bin (detector unit), a formulation of g_i can be, for example: $g_i(l) = l + r_i$. We stress that

Eq. 2.13 expresses an approximation but not an equality, which is justified by the existence of noise. If noise was absent, then r_i would be equal to zero and the use of an accurate model for the measurement would render possible the equality. As this is not the case, and noise is inevitable, the approximation is kept and its degree of usefulness derived from the statistical model for the measurement. So, quoting J. Fessler [Fessler, 02], we can assume that this statistical model is concerned with the approximation in the equation above. Moreover, it also becomes clear why we refer sometimes to a statistical model for the measurement and others to a statistical model for the noise: this model deals with the noise affecting the measurement. Going further, a third expression is usually referred: a statistical model for the data. Strictly regarding Eq. 2.13, one is tempted to use indistinctly either this expression or one of the previous two. However, as seen before, there are differences between measured data and data used on the reconstruction. Due to the importance of these differences and their direct implication on the distinction between SIR methods, from now on we adopt for the component being studied the designation “statistical model for the data”.

Once established a first division among algebraic techniques (statistical versus non-statistical methods), let us concentrate on the group of SIR methods, i.e. those including a statistical model for the data. The Poisson distribution is the model usually assumed for the number of photons reaching a detector bin within a specific interval of time. So, some algorithms were devised assuming a Poisson distribution as the statistical model for the data. This is the case of the ML-EM (Maximum Likelihood – Expectation Maximization) and OSEM (Ordered Subsets – Expectation Maximization), two very important algorithms in the context of the present thesis that will be detailed latter.

However, if corrections are to be applied to the data prior to reconstruction, the Poisson model assumption becomes less adequate for described corrected data. To be more precise, corrected data are no more Poisson data, but Poisson based algorithms can still be used along with these data resulting in reconstructed images whose quality and truthfulness make them useful for clinical purposes⁵. This is the sense for the expression “less adequate”: it works, but it is not accurate.

From the paragraph above, two main questions remain opened: why corrected data are no more Poisson data, and, in face of this, which model should be used or, alternatively, how, if possible, could data be given Poisson characteristics and what does this mean. For corrected data, in the present context, we understand data resulting from corrections for physical effects

⁵ Otherwise, once acquired data is corrected for physical effects in almost all the cases, conventional ML-EM reconstructions would be rejected.

applied upon the prompts (assumed as the acquired data) from a 3D acquisition. If we use a notation similar to those introduced in section 1.4 (except for the dead time correction which can be included under the normalization), correcting the trues, p_i^{3D} , for randoms, scattered and attenuation, after normalization the result can be mathematically expressed as in equation 2.14:

$$y_i^{3D} = N_i \cdot \mathcal{A}_i (p_i^{3D} - r_i - s_i) \quad \text{Eq. 2.14}$$

The expression for the mean value expected on bin pair i is:

$$\langle y_i^{3D} \rangle = N_i \cdot \mathcal{A}_i (\langle p_i^{3D} \rangle - \langle r_i \rangle - \langle s_i \rangle) \quad \text{Eq. 2.15}$$

while the variance for the same bin pair is given by:

$$\sigma^2(y_i^{3D}) = N_i^2 \cdot \mathcal{A}_i^2 (\langle p_i^{3D} \rangle + \langle r_i \rangle) \quad \text{Eq. 2.16}$$

since both the prompts and the randoms are modeled by a Poisson distribution and the variance for the scattered photons is assumed to be equal to zero. Comparing equations 2.15 and 2.16 we verify that the mean value and the variance for the variable y_i are not equal. So, this variable is not a Poisson variable, justifying the claim made earlier that corrected data are not Poisson data.

The second question to be addressed is how to handle non-Poisson data. To overcome this problem, two main approaches have been used: re-shape data in order to give them characteristics close to those required by the Poisson model; or, alternatively, adopt another statistical model, for example, the Gaussian model. The first approach will generally involve the use of a weight, so that weighted data show a variance equal to the mean value, the most important characteristics of the Poisson model. This doesn't mean, in fact, that weighted data becomes Poisson data, because nothing is done in respect to high-level moments of the distribution. However, the matching of the first two moments could be understood as a first order approximation to the Poisson model, allowing us to keep using this model for the data. Reconstruction methods based on this approach are usually known as weighted reconstructions.

In alternative to the use of Poisson model, Fessler [Fessler, 94] suggests the Gaussian model as being appropriate for describing data, and purposes the use of a weighted least-squares similarity measure. This is another important cleavage between SIR methods: those assuming a

Poisson model for the data could be gathered into a group; those adopting the Gaussian model in another group. The later usually have the expression “*least-squares*” in their designation.

2.4.1.4 – Objective function

The expression of the objective function can account for two different terms: one that gives a measure of how well the image fits the data, and another to control how this image matches some desired properties. In fact, the former term is always present in the objective function of all SIR methods. To be more precise, to verify how the image fits the data is equivalent to take one previous estimation of the image (assumed voxelized), and use both the system matrix and the statistical model for the data to obtain an estimation of the expected values in each detector bin, under these assumptions. These are the values to be compared with the acquired data and the result of this comparison reflects the fitting between image and data, the extension of which is measured by the likelihood function. The likelihood (cf. Appendix A.5), in the present context, is the joint probability of obtaining a data vector identical to the measured vector (acquired data), if the measurement process (system matrix and statistical model for the data) is applied to a given image. Denoting the object by λ and the vector with the measured data by \mathbf{Y} , the likelihood term of the objective function can be described by $L(\mathbf{Y}, \lambda)$.

Assuming a Poisson model for the data, each element of vector \mathbf{Y} , y_i is a Poisson distributed random variable whose expected value is $\sum_{j=1}^m a_{ij} \lambda_j$, where λ_j is the mean activity of voxel j , a_{ij} is the element of the system matrix linking voxel j and bin i , and m is the number of voxels contributing to bin i . Thus:

$$y_i \sim P\left(\sum_{j=1}^m a_{ij} \lambda_j\right)$$

Under this model, the probability of detecting y_i photons on bin i is given by (cf. Appendix A.6):

$$p(y_i) = \frac{\exp\left(-\sum_{j=1}^m a_{ij} \lambda_j\right) \left(\sum_{j=1}^m a_{ij} \lambda_j\right)^{y_i}}{y_i!}$$

and the joint probability of observing all the values of vector \mathbf{Y} , i.e., the likelihood function (the joint probability of observing the actual detector count data over all possible densities, λ) is:

$$L(\mathbf{Y}, \boldsymbol{\lambda}) = \prod_{i=1}^n \frac{\exp\left(-\sum_{j=1}^m a_{ij} \lambda_j\right) \left(\sum_{j=1}^m a_{ij} \lambda_j\right)^{y_i}}{y_i!} \quad \text{Eq. 2.17}$$

The corresponding log-likelihood function (for \mathbf{g} , a vector of Poisson distributed variables) is given by [Lange, 84; DePierro, 93; Rockmore, 76]:

$$\log L(\mathbf{Y}, \boldsymbol{\lambda}) = \sum_{i=1}^n \left(-\sum_{j=1}^m a_{ij} \lambda_j + y_i \ln \left(\sum_{j=1}^m a_{ij} \lambda_j \right) - \ln(y_i!) \right) \quad \text{Eq. 2.18}$$

Some authors [Fessler, 94] argue that objective functions based solely on the measurement statistics (Poisson or Gaussian) perform poorly due to the ill-conditioned nature of tomographic reconstruction. In particular, this sort of unregularized methods, i.e., methods whose objective function comprises just the likelihood term, produce increasingly noisy images as the number of iteration also increases. So, the same authors point the need for regularization methods containing a term to impose a smoothness constraint on the reconstructed image. This is the justification for the additional term on the objective function, the one that controls how well the image matches some desired properties. Explicit regularization procedures include the introduction of a prior distribution through a Bayesian approach. Among priors that are mentioned in the literature one finds Gibbs distributions, Gauss-Markov priors, Hube priors, non-convex functions, Markov Random Fields and other. If we designate this regularization term by $R(\boldsymbol{\lambda})$, then the general expression for the whole objective function could be as follows:

$$\Phi(\mathbf{Y}, \boldsymbol{\lambda}) = L(\mathbf{Y}, \boldsymbol{\lambda}) - \beta R(\boldsymbol{\lambda}) \quad \text{Eq. 2.19}$$

where β is a parameter that controls the balance between the data fitting criterion and the image property criterion, that are usually conflicting goals [Fessler, 94]. From a practical perspective, this translates on tradeoff between spatial resolution and noise on the reconstructed image [Lewitt, 03; Qi, 99].

As shown in Appendix A.6, the existence of both terms or just the likelihood term, in the objective function, is on the basis of another dichotomy this time inside SIR methods. If just the likelihood term is present, these methods are referred as non-Bayesian reconstruction methods. On the contrary, when the objective function includes in addition to the likelihood the

regularized term, the reconstruction is said to be Bayesian. So, whenever a reconstruction method is said to be regularized, or penalized, it can also be classified as Bayesian.

2.4.1.5 – Numerical algorithm

The numerical algorithm is the most visible face of the reconstruction method. It is often abusively identified with the whole reconstruction procedure, while, as we have been stating, it is in fact just one of its parameters. This is, likely, a result from the practice: the computational implementation of a statistical reconstruction is almost always synonymous of coding or adapting a numerical algorithm while the choices for the other parameters had been implicitly assumed.

A numerical algorithm is designed to find the image vector for which the objective function is maximized (or minimized). In most of the cases, this algorithm is an iterative one that produces successive estimates of the set of coefficients of the basis functions or, in another way, assuming a voxelized image, successive estimates of each voxel's value. The algorithm is conceived so that the sequence of estimates converge to the aforementioned solution maximizing (or minimizing) the objective function.

Inside the algorithm is comprised an estimator: the tool behind the capacity of maximize or minimize the objective function. This is better explained using the example of the ML-EM – an acronym for Maximum Likelihood-Expectation Maximization. As the name suggest, the objective function to be maximize is the Likelihood, while the tool used for its maximization - the estimator - is the Expectation. In [Vardi, 98] there is a discussion relating the maximum likelihood estimate and the minimum for the Kullback-Leibler information divergence function (assumed as a distance function), which is relevant because the idea of associate the maximum of the objective function with the minimum of a distance is very appealing and intuitive.

The Expectation Maximization is itself an algorithm [Dempster, 77]. It lies in the core of the iterative computation of the maximum-likelihood estimate when the observation can be viewed as incomplete data, and its name results from the fact of being possible to identify two main steps inside this algorithm: an expectation step (E-step) and a maximization step (M-step).

For the application of the EM algorithm to the emission tomography, let us consider a vector, \mathbf{Y} , with the observed (acquired) data. Since it is assumed that detected photons are emitted from a discretized (voxelized) object and one are intending to reconstruct is the activity distribution of this discretized object, then vector \mathbf{Y} is a vector on incomplete data: there is a superposition of the Poisson stream of photons [Green, 90] and it is not possible to observe the

number of photons recorded at bin i and emitted from voxel j . What is missing (thus the incompleteness) is a label stating the source of each photon in the object.

The EM postulates a vector of complete data [Dempster, 77; Lange, 84; Hebert, 89], \mathbf{X} , such that \mathbf{Y} is a function, $h(\mathbf{X})$, of \mathbf{X} . In the present context, the elements of \mathbf{X} are x_{ij} - the number of emissions from voxel j that are detected at bin i - and $y_i = \sum_j x_{ij}$. The complete data are well

modeled as independent Poisson random variables with means equal to $a_{ij}\lambda_j$, where a_{ij} is the element of the system matrix for the probability that a photon emitted from voxel j will be detected at bin i ; and λ_j is the mean total number of photons emitted from the same voxel. It is also important to note that the relation between the complete and incomplete data must be a many-to-one mapping from X to Y , the two sample spaces, i.e., given a realization \hat{x} , only one particular realization \hat{y} has a nonzero probability of having occurred; given a realization \hat{y} , there is a set of \hat{x} with a nonzero probability of having occurred. So, the art in using the EM algorithm [Lange, 84] relies in the option to choose an appropriate complete data specification, \mathbf{X} . However, although many possible ways of embedding \mathbf{Y} in a large sample space are possible, often physical considerations suggest a natural definition of \mathbf{X} , as is clearly the case in stated complete/incomplete data formulation for the case of emission tomography.

The two steps of the EM algorithm can be resumed as follows [Dempster, 77; Lange, 84]:

- E-step: estimate the complete-data sufficient statistics, \mathbf{X} , by means of the conditional expectation: $\hat{x}^{(p)} = E[\mathbf{X} | \mathbf{Y}, \boldsymbol{\lambda}^{(p)}]$;
- M-step: the former conditional expectation is maximized with respect to $\boldsymbol{\lambda}$, so $\boldsymbol{\lambda}^{(p+1)}$ is determined as the solution of the equation: $E[\mathbf{X} | \boldsymbol{\lambda}] = \hat{x}^{(p)}$

where $\boldsymbol{\lambda}^{(p)}$ is a vector with the parameter's values after p iterations and $\hat{x}^{(p)}$ is one realization of \mathbf{X} .

If the objective function is not the Likelihood but, instead, a quadratic distance to be minimized, then the standard numerical algorithm will be the conjugate gradient or another closely related with this one. Nevertheless, Kaufman [Kaufman, 93] proposes the use of the Expectation Maximization with a least squares approach - resulting in the EM-LS algorithm. The perspective behind this work points for the possibility to use the same estimator, with the necessary adaptations, independent from the objective function that has been assumed.

Ideally, the objective function would determine the statistical properties of the estimator. In practice, however, the convergence characteristics of the algorithm that maximize the objective may also influence those properties (if the algorithm only finds local extrema of the objective function, for example, then the estimator is inefficient) [Fessler, 94]. The same convergence characteristics may result, and in practice do often result, in the need to stop the iterative process before the sequence of estimates has converged. This is the case, for example, when the convergence is too slow.

As a final remark, we note that as the objective function becomes more complex it becomes more difficult to derive a special-purpose algorithm.

2.4.2 – Algebraic Reconstruction Technique (ART)

Algebraic Reconstruction Technique (ART) [Herman, 80; Gordon, 74; Herman, 73] is an example of a non-statistical reconstruction method. It can be resumed as follows: a projection, corresponding to a set of LORs with a certain orientation, is selected. As seen, each LOR intersects a certain number of voxels, whose contribution to that LOR was previously established and is stored in the system matrix (a_{ij}). Based on these contributions, a forward projection is done and the value for the bin corresponding to the integration along the LOR is estimated. Then, the estimated and the acquired values for that particular bin are compared by the means of a subtraction, and a correction term is obtained. Finally, the latter term is backprojected and used for updating the values of the voxels intersecting the current LOR. The process continues in a similar way using in each cycle a different projection, i.e. a different set of LORs with an angular orientation different from all the ones used previously. When all the projections had been used once, a single iteration is said to be complete.

Since ART is an iterative process, the steps described above are repeated in subsequent iterations, with the constraint that all the projections must be used once, and only once, per iteration. When some error or convergence-rate threshold is reached, the process is stopped. In a mathematical formulation, the equation that drives the ART is the following [Herman, 93]:

$$\lambda_j^{(k)} = \lambda_j^{(k-1)} + b \frac{y_i - \sum_{v=1}^m a_{iv} \lambda_v^{(k-1)}}{\sum_{v=1}^m (a_{iv})^2} a_{ij} \quad \text{Eq. 2.20}$$

where $\lambda_j^{(k)}$ and $\lambda_j^{(k-1)}$ are the values estimated for voxel j on iteration k and $k+1$, respectively; y_i is the measured value for bin i ; a_{ij} is the element of the system matrix storing the probability that

and annihilation in the l th voxel is detected by the i th detector pair; and b is a relaxation parameter [Herman, 80]. Inspecting Eq. 2.20 we observe that the numerator of the fraction translates the comparison, in the form of a subtraction, between acquired data, y_b and estimated data, $\sum_{l=1}^m a_{il} \cdot \lambda_l^{(k-1)}$. For the estimation of the value expected in detector bin i , all the m voxels contributing to the corresponding LOR are considered, and the system matrix is used as the “look-up table” where each contribution is quantified. Voxels’ values obtained in the previous iteration are taken as the starting point for that estimation. The denominator in Eq. 2.20 is a normalization factor; the fraction, it self, is backprojected and the result used to update the values in each voxel.

An alternative approach known as Simultaneous ART (SART) [Mueller, 97] updates a pixel by the weighted contribution of all the LORs, y_b that are part of the current projection set.

Two very important aspects related the ART are the ordering in which the data are accessed, i.e. the order in which the projections are selected, and the relaxation parameter.

In terms of projections’ order, it is desirable that they are accessed in such a way that subsequently projections are the most uncorrelated as possible, which means that consecutively applied projections must have significantly different angular orientations. In fact, if subsequently projections were chosen at similar angular orientations, there would be a tendency to bias the reconstruction with respect to that particular angle, without introducing much new information each time Eq. 2.20 were to be used. As a consequence, converge rate would be reduced and the solution could be biased [Mueller, 97]. In order to minimize correlation in projection access, the same authors postulate that projections should be arranged according to the following:

- a series of subsequently applied projections should be evenly distributed across a wide angular range;
- at no time should any angular range be covered more densely than others.

Many different alternatives had been appointed for ordering the projections in such a way to maximize, in terms of this parameter, the quality of reconstructed images. In [Mueller, 97] some of these alternatives are referred and the results compared.

Concerning the relaxation parameter, it is advocated in [Herman, 93] that there is no single choice for the best value of this parameter, which depends on the medical purpose of the reconstruction, the method used for data collection and the number of iterations to be used. So, it is concluded, the relaxation parameter should be selected by means of experimentation. In [Andersen, 89] it is stated that its values should be in the range between 0.0 and 2.0.

2.4.3 – Maximum Likelihood – Expectation Maximization (ML-EM)

The Maximum Likelihood – Expectation Maximization (ML-EM) was introduced by Carson and Lange [Lange, 84] and Shepp and Vardi [Shepp, 82]. It is, probably, the most important statistical reconstruction algorithm. This assertion can be justified for two main reasons: first, in consequence of the extensive number of studies devoted to it, some of the properties of this algorithm, namely, its convergence and noise characteristics, have been established and used either to justify its use (or the subsequent OS-EM) as it was originally conceived or to suggest more or less closely related algorithms base on the strategies behind ML-EM, such as, for example, generalized EM algorithms or other Bayesian. That is, many reconstruction schemes were derived having the ML-EM as a reference; second, it is a fact that ML-EM was, and OS-EM is nowadays, a widespread algorithm in clinical routine.

The ML-EM is an algorithm conceived for maximizing the likelihood – the objective function – using the algorithm of the expectation maximization. Since the likelihood is the single term present in the objective function, it is a non-Bayesian reconstruction.

Let us start by retake the expression for the log-likelihood:

$$\log L(\mathbf{Y}, \boldsymbol{\lambda}) = \sum_{i=1}^n \left(- \sum_{j=1}^m a_{ij} \lambda_j + y_i \ln \left(\sum_{j=1}^m a_{ij} \lambda_j \right) - \ln(y_i!) \right) \quad \text{Eq. 2.21}$$

where, we recall, y_i is a random Poisson distributed variable describing the number of photons detected at bin i ; λ_j is the mean number of photons emitted from voxel j , and a_{ij} is the probability that a photon emitted from voxel j will be detected at bin i . In face of this, the aim of the ML-EM is to determine an estimate $\hat{\lambda}$ of $\boldsymbol{\lambda}$ which maximizes the log-likelihood.

The first thing to be analyzed is the possibility of obtaining a maximum for the log-likelihood. The first partial derivative of the log-likelihood is:

$$\frac{\partial}{\partial \lambda_l} \log L(\mathbf{Y}, \boldsymbol{\lambda}) = - \sum_{i=1}^n a_{il} + \sum_{i=1}^n \frac{y_i}{\sum_{k=1}^m a_{ik} \lambda_k} a_{il}$$

while the second partial derivative is given by:

$$\frac{\partial^2}{\partial \lambda_k \partial \lambda_l} \log L(\mathbf{Y}, \boldsymbol{\lambda}) = - \sum_{i=1}^n y_i \frac{a_{ik} \cdot a_{il}}{\left(\sum_{k=1}^m a_{ik} \lambda_k \right)^2}$$

and from these equations it was shown that $\log L(\tilde{f}, g)$ is concave and that the Khun-Tucker conditions given by:

$$0 = \lambda_l \frac{\partial \log L(\mathbf{Y}, \boldsymbol{\lambda})}{\partial \lambda_l} = -\lambda_l \sum_{i=1}^n a_{il} + \lambda_l \sum_{i=1}^n \frac{y_i}{\sum_{k=1}^m a_{ik} \lambda_k} a_{il} \quad \text{and} \quad \frac{\partial \log L(\mathbf{Y}, \boldsymbol{\lambda})}{\partial \lambda_l} \leq 0, \text{ if } \lambda_l = 0$$

determine a maximizer for $\log L(\mathbf{Y}, \boldsymbol{\lambda})$ such that $\lambda \geq 0$ [Kaufman, 87; Vardi, 85; Shepp, 82].

Once established the convergence of the ML-EM, for describing the two steps of this algorithm we need to consider the aforementioned complete data (cf. section 2.4.1.5).

The first stage - E-step – involves taking the conditional expectation of the complete data, \mathbf{X} , with respect to \mathbf{Y} and the current vector of parameter estimates, $\boldsymbol{\lambda}^{(n)}$. For the complete data, the log-likelihood function is expressed as (the two subscripts associated with variable x underline the completeness of data):

$$\log L(\mathbf{X}, \boldsymbol{\lambda}) = \sum_{i=1}^n \left(-\sum_{j=1}^m a_{ij} \lambda_j + x_{ij} \ln \left(\sum_{j=1}^m a_{ij} \lambda_j \right) - \ln(x_{ij}!) \right)$$

and, so [Lange, 84]:

$$E[\log L(\mathbf{X}, \boldsymbol{\lambda}) | \mathbf{Y}, \boldsymbol{\lambda}^{(n)}] = \sum_{i=1}^n \left(\sum_{j=1}^m -a_{ij} \lambda_j + N_{ij} \ln(a_{ij} \lambda_j) \right) + R$$

where $N_{ij} = E(x_{ij} | y_i, \boldsymbol{\lambda}^{(n)}) = \frac{a_{ij} \lambda_j^{(n)} y_i}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)}}$ and R does not depend on the new $\boldsymbol{\lambda}$.

For the M-step, the partial derivatives of the former expression are taken, and the result set equal to zero. In consequence, it is obtained the mathematical expression that drives the iterative ML-EM algorithm⁶ [Kaufman, 87; Vardi, 85; Shepp, 82; Lange, 84]:

$$\lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{\sum_{i=1}^n a_{ij}} \sum_{i=1}^n \frac{a_{ij} y_i}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)}} \quad \text{Eq. 2.22}$$

⁶ In [Hsiao, 04] it is stated that there are five different ways to derive the EM-ML. However, if that is the case (the confirmation of this is beyond the scope of this thesis), the derivation we shortly describe is probably the most often found in literature.

where $\lambda_j^{(n+1)}$ is the j th component of the vector with the new parameters. Under this equation, the non-negativity constraints, $\lambda_j^{(n)} \geq 0$, are always satisfied.

Due to its importance, Eq. 2.22 deserves a close inspection: the factor $\frac{y_i}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)}}$ is the ratio between measured counts (acquired data) and the estimate for the mean number of counts detected at bin i . Thus, the main sum, $\sum_{i=1}^n \frac{a_{ij} y_i}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)}}$, is the backprojection of the former ratio

for voxel j . In a less mathematical way, the whole iterative equation can be interpreted as follows [Bryant, 02]:

$$\text{Img}^{(k+1)} = \text{Img}^{(k)} \times \text{Normalized Backproj}(\text{Measured Proj} / \text{EstimatedProj}^{(k)})$$

or, in other words, using a system model and the current estimate of the object (image) available for each iteration, it is possible to estimate a new set of projections. This new estimated projections are then compared with the measured projections (acquired data) and the ratio between estimated and acquired projections is used to update the estimate of the object, which will be used in the subsequent iteration as the starting point. The whole process corresponds to one iteration, and is repeated a number of times equal to the desired number of iterations.

The proof that the EM iterations converge to the appropriate maximum likelihood estimate regardless of their initial value is given in [Lange, 84; Vardi, 85]. This is the monotonicity property of the algorithm. However, it was observed that reconstructed images become noisier as the algorithm converges towards the maximum-likelihood estimate [Wilson, 94; Snyder, 85]. As it is argued in [Snyder, 85], this is not a result of using the EM algorithm to maximize the likelihood, but something that would occur with any algorithm to produce the maximum-likelihood estimate. In face of this fact, two types of solutions had been proposed: the use of sieves to stabilize the estimation procedure [Snyder, 85; Llacer, 93]; and the termination of the iteration process before the onset of deterioration, based on some stopping criteria [Hebert, 90a; Lion, 91; Llacer, 89; Veklerov, 87].

The existence of distortions, near the edges, that become larger as the number of iterations increases is reported by Snyder et al. [Snyder, 87]. In the same paper the authors also

argue for the intrinsic character of this effect which, as for the increase of noise, is fundamental to reconstructions based on the maximum-likelihood.

Another important remark that should be clarified is the following: in [Shepp, 82], the factor $\sum_{k=1}^m a_{ik}$, corresponding to the normalization of the backprojection, doesn't appear outside the main sum, which means that its value is equal to one. Such assumption is equivalent to consider that all the photons emitted from voxel j are detected at some bin i .

Using a notation based on matrices and vectors, Barrett [Barrett, 94] re-writes Eq. 2.22 as:

$$\hat{\lambda}^{(n+1)} = \frac{\hat{\lambda}^{(n)}}{\mathbf{s}} \left\{ \mathbf{A}^T \left[\frac{\mathbf{Y}}{\mathbf{A}\hat{\lambda}^{(n)}} \right] \right\}$$

where: $\hat{\lambda}^{(n+1)}$ is a vector containing the estimates for the parameters after $n+1$ iterations; \mathbf{s} is a vector with dimensions $(N \times 1)$ and component $s_n = \sum_m A_{mn}$; \mathbf{A} and \mathbf{Y} are for the usual system matrix and vector with the measured data, respectively. In the same paper an equation is derived for describing the behavior of the ML-EM algorithm with noise-free data. It is also predicted that the variance at any point in the image is given by:

$$\text{var} \left\{ \hat{\lambda}_n^{(k)} \mid \lambda \right\} = \left[a_n^{(k)} \right]^2 \left[\mathbf{K}_\epsilon^{(k)} \right]_{nn} \quad \text{Eq. 2.23}$$

where $E \left[\mathbf{z}^{(k)} \mid \lambda \right] = \ln(\mathbf{a}^{(k)})$, $\mathbf{z}^{(k)} = \ln(\hat{\lambda}^{(k)})$ and $\mathbf{K}_\epsilon^{(k)}$ is a term related with the covariance for the noise. The interpretation of the equation above is the following: the variance at any point contains a factor that is proportional to the square on the ensemble mean image at that point, which results in bright regions with a large variance and regions with fewer counts with a smaller variance. However, it is not correct to infer that the variance is proportional to the square of the mean, since the factor $\mathbf{K}_\epsilon^{(k)}$ depends on the object in a complicated way [Barrett, 94].

At this point, some confusion can arise since previous studies based on ML-EM reconstructed images, such as for example [Llacer, 93; Holte, 90; Lion, 91], also suggested that noise in regions of low counts is lower than in regions of high counts, but pointed to a monotonically increasing dependence of the noise variance on the mean (or noise-free) image pixel value. Under this framework, Wilson *et al* [Wilson, 94] used Monte Carlo simulations to verify the relation between the variance in one point of the image and the mean image value at the same point. The results confirm that to bright points correspond larger variances, but the function describing the relation is not obvious (due to the presence of noise covariance factor):

deviations from linearity were found by the authors. Nevertheless, it is predicted that the variance image would resemble the mean image, something that is confirmed by Nuyts [Nuyts, 02]. So, the theoretical quadratic dependency present in Eq 2.23 is masked by the covariance factor, leading to a whole function that, as shown, is neither quadratic nor linear.

Once established this characteristic of the noise, it is also stated [Llacer, 93; Holte, 90] that ML-EM images provide better quantitation than FBP in regions of low intensity in the presence of regions of high intensity in the same image [Llacer, 93]. Noise in ML-EM is much more local than in FBP where the variance tends to be uniform over the reconstructed image.

Nuyts [Nuyts, 02] use numerical experiments to study the variance of post-smoothed ML-EM images. It was found that smoothing is more effective in reducing the variance in high count regions than in low count regions, and, as a result, the Fisher information matrix can be used to produce a measure of the variance in post-smoothed in ML-EM images.

In terms of spatial resolution, a disadvantage of ML-EM is the non-uniform recovery of the spatial resolution, which is not uniform over the image plane and depends on the source distribution when the iteration process is stopped to avoid image deterioration due to noise increasing [Tanaka, 92; Lion, 93]. This is because low-frequency components of the image tend to be recovered earlier in the iteration than high-frequencies [Lion, 93; Pan, 91]. However, it is interesting to note that Pan [Pan, 91] as well as Ranganath [Ranganath, 88] use this putative disadvantage as the basis for a multigrid reconstruction method whose purpose is to accelerate the convergence rate of the ML-EM. The idea is to use first a coarse-grid iteration to reconstruct low-frequency components and provide the result to subsequent fine-grid iteration for an efficient reconstruction of high-frequency components.

In [Lion, 93] the resolution behaviour in images reconstructed by ML-EM is compared to those resulting from a FBP reconstruction.

Hwang [Hwang, 06] describes an accelerated version of the ML-EM, using the whole data set for each update, that – the authors claim – produces almost the same images, noise properties and log-likelihood values as regular ML-EM, but is three times faster. The acceleration is achieved by increasing the step size of the ML-EM, a multiplicative term whose identity becomes clear when the conventional ML-EM is written in its additive form [Hwang, 06]:

$$\lambda_j^{(n+1)} = \lambda_j^{(n)} + \frac{\lambda_j^{(n)}}{\sum_{i=1}^n a_{ij}} \sum_{i=1}^n \frac{a_{ij}}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)}} \left(y_i - \sum_{k=1}^m a_{ik} \lambda_k^{(n)} \right)$$

The step size is the $\lambda_j^{(n)} / \sum_{i=1}^n a_{ij}$ factor in the equation above (which reveals the ML-EM as a process in which the solution is updated by adding the weighted sum of error terms). The increase in the step size is obtained by raising the update factor in the multiplicative (conventional) form of ML-EM to a power b .

2.4.4 – Ordered Subsets – Expectation Maximization (OSEM)

One of the greatest disadvantages of the standard ML-EM is the rate of convergence, which is rather low in face of the practical constraints associated with the clinical routine. Put it in another way, in order to obtain images with an acceptable liability for diagnosis purposes, it is necessary to iterate many times. Thus, the need for a strategy to accelerate the ML-EM algorithm has soon become an evidence: some authors re-wrote the ML-EM equation [Lewitt, 86; Kaufman, 87; Rajeevan, 92]; others [Chen, 91; CruzRivera, 95; Jones, 99] adopted the parallelization as the strategy for speed up the ML-EM algorithm. In the following paragraphs we will be discussing the Ordered Subsets – Expectation Maximization (OSEM), an algorithm resulting from a modification of the ML-EM and, in practice, considered as an accelerated version of this algorithm. It was devised by Hudson in 1994 [Hudson, 94].

It should be noted, however, that Holte [Holte, 90] had already suggested that a major improvement of ML-EM could be obtained by the used of projection subsets for each iteration. In this study, the authors compare (even if using a very simple FOM such as the percent error) this strategy and the regular ML-EM, founding similar results. In the conclusion of Holte's paper, it is referred that a "*major speed improvement is achieved through the use of projection subsets*".

The main idea of OS-EM is to group projections into subsets and to use them in an ordered sequence (Fig. 2.8). In contrast to ML-EM, where an iteration implies handling all the acquired projections data as a whole and their use to update the estimate of activity just one time, one OSEM iteration is defined as a single pass through all the subsets of projections, with the additional and distinctive feature that each subset uses the current estimate to initialize the algorithm with that data subset. That way, the convergence is accelerated by a factor proportional to the number of subsets [Hudson, 94]. Further iterations can be performed by passing through the same ordered subsets, using as starting point the reconstruction provided by the previous iteration.

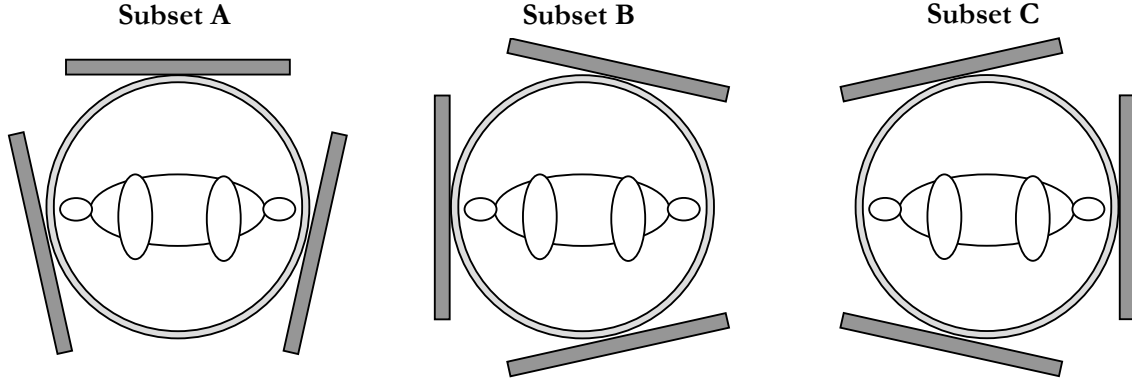


Figure 2.8 – An example of a subset arrangement used by OSEM. Among all collected projections around the scanner (not represented), inside the same iteration for each new update is used just one subset (A, B or C). The order by which subsets are used is relevant and not necessarily A-B-C.

Since the algorithm used for each data subset is the standard ML-EM, from a practical point of view⁷ it is usually assumed that OS-EM is an accelerated version of ML-EM. In [Meikle, 94] a simulation study is described where OSEM is compared to ML-EM (for whole-body PET and using the mean squared error⁸) under this perspective.

Let S_1, S_2, \dots, S_L be the chosen projections subsets in the selected order. For better describing the OSEM algorithm we will adopt a notation similar to those used in Eq. 2.22, but, due to the existence of subsets, an additional index will be included to indicate the subject that is being handled at each step of the algorithm. So, inside a single iteration, $\lambda^{(n,l)}$ will be the image of the activity distribution for iteration n and subset l . Since the latter index just makes sense inside the subsets loop, outside this loop it is omitted.

Considering $\lambda^{(0)}$ as the starting image, OSEM can be resumed as follows (adapted from [Hudson, 94]):

- i) $n = 0$: $\lambda^{(0)}$ is initialized positive;
- ii) repeat until $\lambda^{(n)}$ fits an established criteria \rightarrow Iterative process
 - (considering the existence of subsets) ---
 - a) $\lambda^{(n,l)} = \lambda^{(n)}$

⁷ We emphasize the “practical point of view”, because from a theoretical perspective things are not so simple and such a statement can hardly be sustained. There is the problem of convergence, for example.

⁸ As defined, the mean squared error indicates the point at which the pointwise accuracy of the reconstruction is greatest.

$$\begin{aligned}
& \text{b) for subsets } l = 1, \dots, t \\
& \quad \bullet \text{ project: } \mu_i^{(n,l)} = \sum_{k=1}^m a_{ik} \lambda_k^{(n,l)}, \text{ for detector bins } i \in S_l \\
& \quad \bullet \text{ backproject: } \lambda_j^{(n,l+1)} = \frac{\lambda_j^{(n,l)}}{\sum_{i \in S_l} a_{ij}} \sum_{i \in S_l} \frac{a_{ij} y_i}{\mu_i^{(n,l)}}
\end{aligned} \tag{Eq. 2.24}$$

If the divisor is zero, set $\lambda_j^{(n,l+1)} = \lambda_j^{(n,l)}$

$$\text{c) } \lambda^{(n+1)} = \lambda^{(n,t+1)}, \text{ i.e., whenever all subsets have been used, terminate current iteration.}$$

It is important to note that if a single subset includes all projections, then OSEM becomes the standard ML-EM.

Hudson [Hudson, 94] states that the order of projections is arbitrary, though it may be advantageous for the quality of reconstruction to choose a special order. For that purpose, they suggest the introduction of new information as quickly as possible by choosing first the projection corresponding to the direction of greatest variability in the image. For providing a maximum of new information each time the algorithm uses a different subset, projections should be ordered so as to be maximally separated [Hutton, 97].

Another import point is the way projections are grouped into subsets: these should be selected so that each voxel activity contributes equally to any subset, or, putting it in a different way, the sum of counts for all projections forming a subset should be equal for all subsets [Hudson, 94]. When this condition is verified, subsets are said to be balanced.

In [Hudson, 94], at the end of the paper, is presented a “proof of convergence of OSEM with subset balance and exact projections”. However, it is also stated that not only this proof of convergence is not valid for noisy data as well as except with noise-free data, OSEM converges to a non-ML solution. According to Byrne [Byrne, 98], when subset balance is missing, OSEM is observed to have a limit cycle consisting of as many distinct images as there are subsets, even when there is no noise. How distinct these images are depends on the balance of the subsets.

The subject of how to group projection into subsets, i.e., how to reduce OSEM’s dependence on subset selection is the main goal of Huang, in [Huang, 99]. The investigation is centered on the oscillation between sub-iterations when inconsistent subsets, due to noise, are used for image updating: the adjustment of pixel values for one subset could be in conflict with the adjustment on a previous or subsequent subset, because of noise. In consequence, each image pixel can go through ups and downs in the iteration process, affecting its convergence to a

stable value. Huang proposes two methods to reduce image oscillation: pre-processing the sinogram to make it satisfy a sinogram consistency condition; and take the average of the ups and downs of the pixels adjustments among subsets, using the average as the correct value. In its conclusion, apart from referring that a combined application of these two methods can give better results than the use of just of them alone, it is noted that the image quality achieved by the “converged reconstruction” (Huang’s terminology) is not necessarily better than that obtained with regularized EM or OSEM algorithms [Huang, 99].

In spite of the unproved formal convergence of OSEM, some results are mentioned in the literature, using different figures-of-merit for describing the algorithm’s performance. Hutton *et al* note that the degree of acceleration may provide a solution (which appears noisy) after a single iteration, since the algorithm has already reached a point equivalent to many EM iterations [Hutton, 97]. In the same paper, two interesting figures are shown, where the chi-squared error and the mean square difference between reconstructed image and original object are plotted against the number of iterations. From the observation of these figures, the authors state that as the number of iterations increase, the likelihood continues to increase, which is supported by the increasing close fit between estimated and true projections. However, the mean square error between reconstructed and true activity improves in early iterations, but then increases due to increasing noise [Hutton, 97]. In relation to this, the authors go one step further suggesting that the visual optimal image typically occurs at an iteration number slightly higher than that corresponding to the minimum mean square error⁹.

An interesting study, and probably one of the first analyzing the performance of OSEM, is described in [Meikle, 94]. There, one can find a discussion about the behaviour of this algorithm with the number of iterations, the choice of the number of subsets, as well as a comparison with both ML-EM and FBP¹⁰.

The problem of OS-EM convergence is addressed in the work described by Hsiao *et al*, where a convergent¹¹ OSEM-type algorithm (E-COSEM-ML – Enhanced Complete Data OSEM-ML) is purposed, aiming to be fast and allowing the avoidance of the user-specified

⁹ This statement should be taken carefully, since the visual optimal image depends on the type of study and task under consideration. However, we believe it deserves mention because raises the point that the optimum for a visual task-oriented inspection is not necessarily coincident with those resulting from a formal perspective where a well established and “physicist-friendly” FOM (such as the mean squared error) is used.

¹⁰ The final paragraph of this paper’s *Discussion* contains the following important observation: “It should be noted that SNR (...) is a mathematical entity and, while tumour detectability can be inferred from this figure of merit, it does not take into account the transfer of information from the image to the observer.”

¹¹ An outline of the proof of convergence of this algorithm is provided as an appendix of the paper by Hsiao *et al*.

relaxation schedules associates with convergent ML based methods [Hsiao, 04]. The core of the derivation of this algorithm is a different complete data objective function, the use of which results in the convergent algorithm designated just by COSEM. Since the authors found that this is a slower algorithm than regular OSEM, they devised E-COSEM as a linear combination of OSEM and COSEM. So, E-COSEM translates into a trade-off between faster OSEM and slower but convergent COSEM, with a relaxation parameter, α (ranging between 0 and 1), expressing the degree of the trade-off. Mathematically, under E-COSEM the update of the activity distribution can be expressed as follows [Hsiao, 04]:

$$\lambda_j^{(k,l)} = \alpha^{(k,l)} \tilde{\lambda}_j^{(k,l)} + (1 - \alpha^{(k,l)}) \tilde{\lambda}_j^{(k,l)}$$

where $\tilde{\lambda}_j^{(k,l)}$ is the OSEM update and $\tilde{\lambda}_j^{(k,l)}$ is the COSEM update. The relaxation parameter is an iteration dependent parameter automatically chosen (not user-specified) to guarantee convergence while keeping the speed-up characteristic of OSEM.

For accelerating the OSEM algorithm itself, some efforts have been made in order to develop parallel processing techniques. Jones [Jones, 03b] describes two parallel algorithms for fully 3D OSEM implemented on a cluster consisting of 9 nodes and 18 processors. Another parallelizing approach is suggested by Vollmar [Vollmar, 02], for both 2D and 3D OSEM reconstructions.

2.4.4.1 – Block-iterative methods

If a Poisson model for the noise is assumed, block-iterative methods are closely related with OSEM and are often included under the same classification. However, since OSEM is well established in clinical routine and unequivocally rooted in the paper by Hudson and Larkin [Hudson, 94], we prefer to separate this algorithm from its relatives. In practice, if a distinction is to be made between block-iterative methods and OSEM, it will rely into two main things: the statistical model (either non-Poisson or inexistent) and/or the claimed convergence of the block-iterative method (assumed as an advantage over OSEM).

As Byrne [Byrne, 98] formulates the framework, block-iterative (or ordered subset) methods, in which a single block or subset is processed at each step, are between simultaneous methods, which use all the data at each step (such as ML-EM), and sequential or row-action methods, that use a single data value at each step.

Byrne [Byrne, 96; Byrne, 98] indicates an equation for driving an iterative process designated as Block-Iterative EMMML (BI-EMML) that uses subsets and is claimed to converge

for all choices of subsets provided that, for each subset S_j , $\sum_{i \in S_j} a_{ij} \leq 1$. The equation, letting the update index, $k+1$, to designate just the k -th update resulting from a certain subset and independent from the fact that a new iteration is beginning or not, is the following:

$$\lambda_j^{(k+1)} = \left(1 - \sum_{i \in S_j} a_{ij} \right) \lambda_j^k + \lambda_j^k \sum_{i \in S_j} \frac{a_{ij} y_i}{\mu_i^k}$$

which reduces to the ML-EM equation when there is just one subset. However, even when subset balance holds, the former equation does not reduce to OSEM [Byrne, 98]. In consequence, to get this equivalence, Byrne introduced the Rescaled BI-EMML (RBI-EMML) given by [Byrne, 98]:

$$\lambda_j^{(k+1)} = \left(1 - m_n^{-1} \sum_{i \in S_j} a_{ij} \right) \lambda_j^k + m_n^{-1} \sum_{i \in S_j} a_{ij} \frac{\lambda_j^k y_i}{\mu_i^k}$$

where $m_n = \max_j \left\{ \sum_{i \in S_j} a_{ij} \right\}$. A proof of convergence of this method is presented as an

appendix to the same paper. A detailed study of the noise properties of block-iterative methods (in particular RBI-EMML) can be found in [Soares, 00; Soares, 05].

2.4.4.2 – Row-Action methods

The Row-Action Maximum Likelihood Algorithm (RAMLA) [Browne, 96; Daube-Witherspoon, 01] is another alternative to the ML-EM for maximizing the Poisson likelihood, aiming to speed up the rate of convergence. It differs from OS-EM because the reconstructed image is updated for each projection view (row of the system matrix), in a controlled way using a relaxation parameter. It is claimed that RAMLA is faster than MLEM and converges to the maximum likelihood solution [Daube-Witherspoon, 01]. In this same paper, the authors refer that the use of an appropriate relaxation parameter grants RAMLA with a stable performance and an independence of the starting point. The use of a controlled relaxation parameter for reconstructing the image in such a way that noise propagation from projections is substantially independent of the access order of subsets in each cycle is described in [Tanaka, 03]. This variation of RAMLA is called DRAMA.

The relation between RAMLA and OS-EM is explored in [Browne, 96].

2.4.5 – Bayesian Reconstructions

The keyword behind Bayesian reconstruction is regularization. This is the general approach for dealing with the instability associated with image processing, in general, and statistical image reconstruction, in particular. Regularization of an ill-posed problem and its Bayesian interpretation is discussed in [Demoment, 89].

Bayesian reconstructions do also assume a statistical model for the data, but the objective function includes two terms: one for the likelihood (as in the case of ML-EM and OS-EM) and the other for a *prior*. Green [Green, 90] formulates the paradigm of a Bayesian reconstruction as involving two probability models:

- the first describes the manner in which the detected counts are generated by the tomographic transformation and other physical circumstances. In particular, this model accounts for the Poisson nature of data;
- the second, is a probability distribution on the space of true patterns of isotope concentration and provides a means of quantitatively express prior information about knowledge of some sort of patterns expected even before data are collected.

In the same paper, the author stresses the close link between this Bayesian paradigm and the principle of penalized likelihood.

If each voxel is given the subscript $j, j = 1, 2, \dots, m$ (the total number of voxels), then λ_j is the isotope concentration in voxel j , and the whole array of voxels, $\{\lambda_j\}$ can be denoted by $\boldsymbol{\lambda}$. Similarly, using the subscript i to index each bin, then y_i is the number of counts in bin i and \mathbf{y} , $\mathbf{y} = \{y_i\}$, the complete set of collected data. Under these models, the reconstruction of the isotope concentration will be accomplished by considering the posterior distribution that follows Bayes' theorem [Casella, 90], $p(\boldsymbol{\lambda} | \mathbf{y})$, from the two components: the data model¹², $p(\mathbf{y} | \boldsymbol{\lambda})$, and the prior model, $p(\boldsymbol{\lambda})$, i.e., the probability distribution reflecting previous knowledge about isotope concentration patterns [Green, 90]. Mathematically:

$$\ln p(\boldsymbol{\lambda} | \mathbf{y}) = \ln p(\mathbf{y} | \boldsymbol{\lambda}) + \ln p(\boldsymbol{\lambda}) - \ln p(\mathbf{y}) \quad \text{Eq. 2.25}$$

where $p(\mathbf{y})$ is a constant.

There are many reconstruction schemes that can be classified as Bayesian. A characteristic that may obviously contribute to differences between these techniques, whilst not unique, is the nature of the prior.

¹² Usually assumed as being a Poisson model.

In this same paper by Green, [Green, 90], is described a type of Bayesian reconstruction which is an extension of the ML-EM reconstruction, based on the Bayes' theorem. We will use the algorithm described in this paper as an example for the discussion of the main features of a Bayesian reconstruction. The first thing to be considered is the prior and its relation to some knowledge *a priori*. As advocated by Green and many other researchers in the field of emission tomography, long-range, high-level structures in the object may not be important, so the prior distribution could be such that only pairwise interactions, among pairs of pixels that are neighbors, are allowed. This is the reason for the widespread adoption of Gibbs priors: they allow the capture of local properties of the object as well as result in computationally efficient update strategies. The general expression of a Gibbs distribution is:

$$p(\boldsymbol{\lambda}) \propto \exp[-\beta U(\boldsymbol{\lambda})]$$

$U(\boldsymbol{\lambda})$ is known as the Gibbs energy function, which is defined as a sum of potentials. The energy function should reflect two conflicting properties: the local smoothness of images within the same organ or tissue, and the sharp variations in tracer uptake between different structures. Different forms of this energy function lead to differences in the Bayesian reconstruction algorithm. In the case of [Green, 90], this prior translates in the following probability distribution:

$$p(\boldsymbol{\lambda}) \propto \exp[-\beta U(\boldsymbol{\lambda})] = \exp\left(-\beta \sum_{s,r} w_{sr} \phi\left(\frac{\lambda_s - \lambda_r}{\delta}\right)\right)$$

where β and δ are parameters and w_{sr} is a weight coding the strength of neighborhood between pixels s and r : $w_{sr} = 1$ if s and r are orthogonal nearest neighbors, $\sqrt{1/2}$ for diagonal neighbors, and 0 otherwise. The function ϕ is nonnegative, symmetric about 0, and monotonically increasing for positive values of its argument. Choosing the more appropriated function ϕ depends on the knowledge for the pattern of the isotope concentration. This is, we recall, just an example of a prior, but useful for the illustration of some considerations behind its definition.

The goal of this sort of reconstruction algorithms is to chose $\boldsymbol{\lambda}$ in order to maximize the posterior probability $\log p(\boldsymbol{\lambda}|\mathbf{y})$. Thus, using Eq. 2.25 and the expression of a Poisson distribution, the Bayesian formulation results in:

$$\ln p(\boldsymbol{\lambda}|\mathbf{y}) = \ln p(\mathbf{y}|\boldsymbol{\lambda}) + \ln p(\boldsymbol{\lambda}) + \text{const} = \sum_t \left(y_t \log \left(\sum_s a_{ts} x_s \right) - \sum_s a_{ts} x_s \right) - \beta V(\boldsymbol{\lambda}) + \text{const}$$

a_{ij} are for the system matrix elements, and $U(\lambda)$ is the energy function described above.

Due to the huge number of Bayesian reconstruction algorithms - making use of different priors, data models and strategies to overcome some other practical problems (such as how to handle negative counts) – in the following paragraphs we will be shortly referring just a few of them that are often cited in the literature devoted to this subject. However, in any case this means that these are more important than others.

Green [Green, 90] notes that a direct extension of the EM algorithm to maximize $\log p(\mathbf{x}|\mathbf{y})$ can still be appropriated, but the M step becomes impractical, except in the degenerated case where the pixels are independent under the prior. In alternative, a *one-step-late* (OSL) approximation is suggested and its rate of convergence (whilst not proved) is compared with that for the standard unregularized ML-EM. The mathematical equation of the OSL algorithm is the following:

$$\lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{\sum_{i=1}^n a_{ij} + \beta \frac{\partial}{\partial \lambda_j} V(\lambda_j^{(n)})} \sum_{i=1}^n \frac{a_{ij} y_i}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)}} \quad \text{Eq. 2.26}$$

where $\beta \frac{\partial}{\partial \lambda_j} V(\lambda_j^{(n)})$ is the derivative of the energy function (prior term). If we compare this equation with equation 2.22 for the ordinary ML-EM algorithm, we realize that the only difference between these two equations lies in the existence of the prior term in the former equation. So, when β is equal to zero, i.e., when the weight associated with the prior term vanishes, the OSL algorithm becomes the ML-EM. Lange [Lange, 90] describes a modified version of the OSL algorithm, suggests some energy functions and gives a detailed study of the convergence of this algorithm.

The quadratic prior is discussed in [Bryant, 02].

Hebert and Leahy [Hebert, 89] developed a generalized expectation maximization (GEM) Bayesian algorithm based upon the Poisson model and a prior in the form of Gibbs function. Development of this algorithm follows that of the regular ML-EM but, for the M-step, a form of coordinate gradient ascent is derived. In addition, as the prior tends toward a uniform distribution, the GEM algorithm reduces to the ML-EM.

Unlike the GEM, the SAGE (*Space-Alternating Generalized EM*) algorithm [Fessler, 95] updates image voxels sequentially using a sequence of small “hidden” data spaces, rather than simultaneously using one complete-data space. Because the new hidden data spaces are less informative than the conventional EM complete-data space for Poisson data, the SAGE

algorithm yields improvements in convergence rate while maintaining the monotonic properties of the EM algorithm.

Fessler [Fessler, 94] suggests a “*penalized weighted least-squares*” (PWLS) reconstruction, under which a Gaussian model for the data is assumed and, in consequence, a least-squares term is included in objective function. This is an alternative to the more usual combination of a Poisson model for the data and the use of likelihood as the objective function. In addition to the least-squares, the objective function also includes a regularizing penalty term, $R(\lambda)$:

$$\Phi(\lambda) = \frac{1}{2}(\mathbf{y} - \mathbf{A}\lambda)^T \mathbf{E}^{-1}(\mathbf{y} - \mathbf{A}\lambda) + \beta R(\lambda)$$

where \mathbf{E}^{-1} is a diagonal matrix whose i th entrance is equal to σ_i^2 : an estimate of the variance of the i th precorrected measurement y_i . The non-negative successive over-relaxation (SOR) algorithm [Sauer, 93] - which relies on updates of single pixel values rather than the entire image, at each iteration - is adopted by Fessler to minimize the objective function, and is reported a rapid convergence to the global minimum of the PWLS objective. The whole procedure is designated by PWLS+SOR.

Levitan [Levitan, 87] and Herman [Herman, 91] develop an exact EM algorithm for MAP optimization with a multivariate Gaussian *a priori* probability distribution. Mumcuoğlu [Mumcuoğlu, 94; Mumcuoğlu, 96] describes a penalized preconditioned conjugate gradient method for both emission and transmission reconstruction, using Markov random field priors and letting the preconditioner to be changed at each iteration.

In [Alenius, 97] it is investigated a type of Bayesian OSL reconstruction method which utilizes a median root prior (MRP), as the authors recognize that median filters are good in terms of noise reduction and, at the same time, preserve edges. To implement this method, the authors replace the derivative of the energy function in Eq. 2.25 by a non-linear function: the median. Hsiao *et al* [Hsiao, 03] claim that it is impossible to associate an objective function to the MRP algorithm and, thus, formulate a median prior that can be used within a MAP context.

The inclusion of anatomical *a priori* information is the scope of an extensive list of papers. The utility of anatomical priors depends on the degree of spatial correlation of anatomical structure with radionuclide distribution: if the correlation is good the anatomic information is expected to serve as a useful prior. Bowsher *et al* [Bowsher, 96] present a method for simultaneously segmenting and reconstructing emission images and for incorporating anatomical *a priori* information. This approach models images as consisting of regions (whose number is estimates) such that all voxels within a given region have similar mean activities; random variables of the model include region mean activities as well as the mean activities and

region classification of individual voxels. Gindi *et al* [Gindi, 93] reconstruct images with the aid of a prior derived from registered anatomical MR images. The prior information is obtained from the anatomical image by an ‘edge map’ reflecting the confidence that a correlated functional edge exists in each location. The authors stress the fact that each value of the edge map should reflect the particular anatomical boundary under consideration. Lipinski *et al* implemented two approaches (a Markov-GEM and a Gauss-EM algorithm) that use the anatomical information obtained from magnetic resonance (MR) images to model a prior distribution. In the same paper is investigated the impact of misleading MR information. Baete *et al* [Baete, 04a] describe an anatomy-based MAP reconstruction algorithm (A-MAP) that includes anatomical information derived from MR imaging data as well as the approximation of a tissue composition model as a priori knowledge. Applications of this algorithm are demonstrated in epilepsy [Baete, 04a] and for partial volume correction in brain FDG-PET [Baete, 04b]. Comtat *et al* [Comtat, 02] investigate the potential gains in image SNR by incorporating the anatomical information derived from aligned CT images obtained from a PET/CT scanner.

Two disadvantages of penalized algorithms are identified by Fessler and Rogers [Fessler, 96]: the absence of a method for choosing the value of the regularization parameter, with the additional complication that for a fixed β , the reconstructed spatial resolution varies between subjects and even within the same subject; and the object-dependent nonuniform spatial resolution (even for spatial-invariant tomographic systems), which is a consequence of the nonuniform variance of Poisson noise [Fessler, 96].

Concerning the first problem, the authors refer the development a method for normalizing the penalty function such that the object-dependent component of β is nearly eliminated, allowing the building of an object-independent table relating that parameter to the spatial resolution (FWHM) for a given tomographic system. Thus, it is possible [Fessler, 96] to select β to achieve a consistent specified resolution within planes, between planes and between subjects. Since the prior is data dependent, it can no longer be regarded as a Bayesian prior [Nyys, 03] and, in consequence, it is referred as a penalty term. In relation to the nonuniform spatial resolution, the authors examine the spatial properties of penalized-likelihood image reconstruction methods via the local impulse response, and propose a modified space-variant penalty that yields images with an increased spatial uniformity over conventional space-invariant regularization. Stayman and Fessler [Stayman, 00] update the method for designing a shift-variant penalty function that attempts to provide uniform resolution properties, for an idealized PET system whose geometric response is space-invariant. In [Stayman, 04] the same penalty function

is applied to space-variant systems. Another method for obtaining a symmetric and shift-invariant point spread function is discussed in [Nyts, 03].

Qi and Leahy [Qi, 99] derive approximate equations for the local impulse response contrast recovery coefficient (CRC) and variance of each voxel in images reconstructed using MAP estimators. These expressions are then applied on the selection of the smoothing parameter, β , in such a way the authors claim to be more appropriate for lesion detection tasks.

In fact, the selection of the regularization parameter is a key problem associated with Bayesian methods. Leahy [Leahy, 00a] separates the approaches for choosing β into two broad classes: treating β as a regularization parameter and applying techniques such as generalized cross validation and the L-curve; and using theoretical estimation approaches such as the maximum likelihood. In the same paper is given an extensive list of these approaches.

Finally, a short mention to ordered subsets-type algorithms for maximizing the penalized likelihood; for example: the Block Sequential Regularized Expectation Maximization (BSREM) [DePiero, 01], and the Relaxed OS Separable Paraboloidal Surrogates (OS-SPS).

2.5 – Rebinning algorithms

A rebinning algorithm can be defined as an algorithm that sorts 3D data into 2D data. As 3D data contains information for oblique planes in addition to transaxial planes (N^2 sinograms, for a scanner with N rings), the amount of information available can be unaffordable for the reconstruction algorithm, in terms of time required for this task. So, whenever the reconstruction time is a serious constraint, the use of a rebinning algorithm is a solution to be considered.

This solution, nevertheless, should be considered carefully. In fact, since it sorts 3D into 2D data ($2N - 1$ transaxial sinograms), there is always the possibility that part of the information contained in the 3D data set is lost; and, because the redundancy of 3D data can be used to improve the SNR, the final quality of the reconstructed image can be affected. How far information is lost in the rebinning process and how deeply it decreases the image quality depends on the rebinning algorithm: the SSRB (Single-Slice Rebinning algorithm), for example, has an accuracy very limited in space, so it can considerably contribute to a degradation of the reconstructed image; the Fourier Rebinning (FORE), on its hand, can produce similar results to

those obtained with fully 3D reconstruction methods [Liu, 01]. As in many other fields, the more accurate is the rebinning algorithm the more complex is its implementation.

In sum: the rebinning decomposes the 3D reconstruction problem into a set of 2D independent reconstructions. If the rebinning algorithm is efficient, the reconstruction becomes fast as in 2D, while retaining the increased sensitivity of the 3D acquisition [Defrise, 97].

2.5.1 – FORE (Fourier Rebinning)

The Fourier Rebinning (FORE) is probably the most important rebinning algorithm for PET, since it retains a remarkable accuracy in practice while its implementation and use does not burden the reconstruction process. In the context of the present thesis it is also very important, since it is used along with 2D reconstruction algorithms as part of the research we have done.

Two different approaches can be adopted for the derivation of FORE. In one of them [Defrise, 97], FORE is present as an approximate version of an exact rebinning algorithm; in the other [Defrise, 95b], the most intuitive, it is a natural application of the frequency-distance relation¹³. The latter approach will be discussed in the following paragraphs.

Let us consider Fig. 1.11 and a cylindrical scanner with length L and radius R . In 3D, the integral of $f(x, y, z)$ – defined in a cylinder of radius $R_\Omega < R$ – along a LOR is given by [Defrise, 95b]:

$$s(x', \phi, z_\Omega, \Delta) = \int_{-\infty}^{+\infty} f(x' \cos \phi + y' u_x, x' \sin \phi + y' u_y, z_\Omega + y' u_z) dy'$$

where \vec{u} is a unit vector along the line of integration, $\vec{u} = \frac{-\sin \phi, \cos \phi, \frac{\Delta}{2\sqrt{R^2 - x'^2}}}{\sqrt{1 + \frac{\Delta^2}{4(R^2 - x'^2)}}}$, and $\Delta =$

$z_\Omega - z_\phi$. Using a slightly different parametrization:

$$p(x', \phi, z_\Omega, \delta) = \frac{s(x', \phi, z_\Omega, \Delta = 2\delta\sqrt{R^2 - x'^2})}{\sqrt{1 + \delta^2}} = \int_{-\infty}^{+\infty} f(x' \cos \phi - y' \sin \phi, x' \sin \phi + y' \cos \phi, z_\Omega + y' \delta) dy'$$

¹³ The frequency-distance relation was introduced by Edholm *et al.*: “Novel properties of the Fourier decomposition of the sinogram”, *Int. Workshop on Physics and Engineering of Computerized Multidimensional Imaging and Processing, Proc. Of the SPIE*, 671, 8 – 18, 1986.

where δ is the tangent of the polar angle, θ , and can be called the ring difference. This parametrization requires a 1D interpolation across different oblique sinograms [Defrise, 95b], but once this interpolation is done, it is a very convenient parametrization.

The continuous Fourier transform, for each pair (z, δ) of a 3D sinogram with respect to the variables x' and ϕ is:

$$P(w, k, z, \delta) = \int_0^{2\pi} d\phi \int_{-R\Omega}^{+R\Omega} \exp(-ik\phi - iw x') p(x', \phi, z, \delta) dx'$$

Using the symmetry $p(x', \phi + \pi, z, \delta) = p(-x', \phi + \pi, z, -\delta)$ and combining the two equations above we get:

$$P(w, k, z, \delta) = \iint_{\Omega} dx dy \int_0^{2\pi} d\phi \exp[-ik\phi - iw(x \cos \phi + y \sin \phi)] f(x, y, z + \delta(-x \sin \phi + y \cos \phi))$$

Eq. 2.27

Eq. 2.27 plays an important role in the establishment of the frequency-distance relation. In fact, considering the phase in the equation above, it is possible to observe that when k and w are large, this phase varies very rapidly with ϕ . Therefore, since this is equivalent to an exponential oscillating rapidly between positive and negative values, when f is smooth the contribution to the integral over ϕ is negligible. So, contributions to the integral come from the neighborhood of values of ϕ where the phase is stationary, i.e.:

$$\frac{\partial}{\partial \phi} (k\phi + w(x \cos \phi + y \sin \phi)) = k + w(-x \sin \phi + y \cos \phi) = 0 \quad \text{Eq. 2.28}$$

When $|k| \leq |w\sqrt{x^2 + y^2}|$, equation 2.28 has two solutions in the interval $[0, 2\pi[$ and both solutions correspond to the same y' (distance along the line of integration) [Defrise, 95b; Xia, 95]. In other word, the frequency-distance relation can be stated as follows [Defrise, 95b; Alessio, 06]: the 3D sinogram at frequency component (k, w) receives contributions primarily from sources located at distance $y' = -k/w$. So, different distances along the line of integration contribute to different locations in the frequency space.

The Fourier rebinning approximation, which is the basis for the FORE algorithm and relies on the frequency-distance relation, relates the 2D frequency transform of an oblique sinogram, $P(w, k, z, \delta)$, with the 2D transform of a direct sinogram corresponding to a slice shifted

axially by $\Delta z = -k\delta/w$. Its mathematical expression is stated in equation 2.29 [Defrise, 95b; Defrise, 97]:

$$P(w, k, z, \delta) \approx P\left(w, k, z - \frac{k\delta}{w}, 0\right) \quad \text{Eq. 2.29}$$

The implementation of the FORE algorithm involves the following steps [Defrise, 95b; Defrise, 97]:

- for each pair (z, δ) , take the 2D Fourier transform of the 3D sinograms: $P(w, k, z, \delta)$ and $P(w, k, z, -\delta)$
- Based on Eq. 2.28, estimate each transaxial rebinned sinogram:

$$P_{2D}(w, k, z) = \frac{\sum_{x, \delta} P(w, k, x, \delta) W(z - x + \delta k / w)}{\sum_{x, \delta} W(z - x + \delta k / w)}$$

where $W(z)$ is a weighting for selecting, for each frequency (k, w) , the 3D sinograms that contribute to the slice z . This is an interpolation in z (FORE does not include any interpolation in k or w) where interpolation strategies can be used;

- for each transaxial slice z , take the inverse 2D Fourier transform of $P_{2D}(w, k, z)$;
- reconstruct each transaxial slice separately, using a 2D reconstruction algorithm.

There is, nevertheless, an important constraint associated with the use of FORE. In fact, FORE is based on the stationary-phase approximation, which is an asymptotic formula accurate just for large values of k and w [Defrise, 95b]. So, when analysing the whole (k, w) plane, three different regions can be considered [Defrise, 97] (Fig. 2.9) in terms of rebinning approaches.

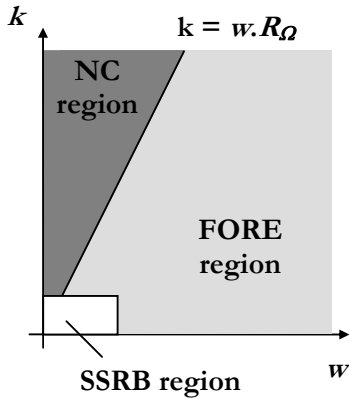


Figure 2.9

The three different regions of the (k, w) space, in terms of rebinning approach: SSRB region, FORE region, and NC (Non-Consistency region).

The FORE region is the high-frequency region, where the stationary-phase approximation is valid and the FORE algorithm is accurate. In the Non-Consistency region, $|k/w| \geq R_\Omega$ (the radius of the FOV). So, $P_{2D}(w, k, z) = 0$. Finally, in the low-frequency region, the Fourier Rebinning is not applicable (stationary-phase approximation not valid) and another algorithm should be considered for data rebinning. This is called the SSRB because the Single Slice Rebinning (discussed in the following section) algorithm's premises are valid and, therefore, it can be adopted. However, as noted by Matej [Matej, 96], an abrupt change from FORE to SSRB is probably not the best strategy. Fig. 2.9 should be understood as an illustration of the region where these two important rebinning algorithms are valid from a theoretical point of view (and not where they are strictly used).

The performance of FORE is less accurate at larger maximum ring differences [Defrise, 97]. In practice, it is accurate when the axial aperture, θ_{max} , is smaller than about 20° [Defrise, 03]. The performance of FORE for a wide range of axial acceptance angles is investigated in [Matej, 96].

2.5.2 – Other rebinning techniques

The most simple rebinning algorithm is the SSRB (Single Slice Rebinning) [Daube-Witherspoon, 87], which was introduced before FORE. Its fundamental assumption is that each LOR only traverses a single transaxial plane within the support of the tracer distribution. If that is the case, then each oblique LOR can be converted into a LOR belonging to the transaxial plane halfway to the planes containing the extremities of the original LOR. This is illustrated in Fig. 2.10.

Mathematically, the SSRB algorithm can be expressed as follows [Defrise, 03]:

$$s_{ssrb}(x', \phi, z) = \frac{1}{2\theta_{max}(x', z)} \int_{-\theta_{max}}^{\theta_{max}} s(x', \phi, \zeta = z, \theta) d\theta$$

where θ_{max} is the maximum axial aperture for an LOR at distance x' from the axis in slice z .

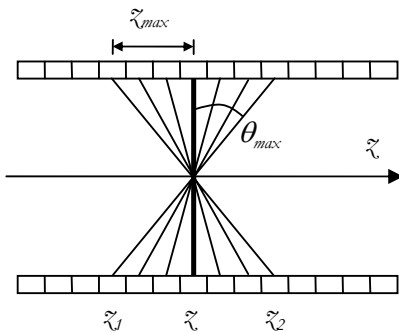


Fig. 2.10

SSRB: an illustration of the set of oblique LORs transformed into a single transaxial LOR.

An inspection of the assumption SSRB is based on restricts the use of this algorithm for activity distributions that are spanned within a short radial distance from the scanner axis, as well as for LORs corresponding to small values of θ . For realistic distributions these conditions can not be met and the accuracy of this algorithm is very limited. The main advantage of SSRB is its simplicity.

In strong contrast with SSRB are the *exact* rebinning methods, whose name derives from the fact that they do not consider the approximation FORE is based on. The study of this sort of algorithms is beyond the scope of this thesis, so we are just referring some of them.

Liu [Liu, 99] introduces the FOREPROJ, a projection algorithm that allows the calculation of 3D attenuation correction factors directly from a 2D transmission scan. In the same paper, FOREX, an exact rebinning algorithm, is also discussed. Another exact rebinning algorithm for 3D PET, using the John's equation – FORE-J – is introduced by Defrise in [Defrise, 99].

3 – The weighted OSEM approach and the adopted implementation

The work which is on the basis of the present thesis has been centered in the use of OSEM for tumor detection.

This is a chapter bridging the hitherto theoretical part of the thesis, with the following chapters, where the experimental results are shown and discussed. We start by discussing the weighted OSEM approach and the main reasons for the fundamental role this strategy plays for handling PET corrected data. After that, we proceed with NEC-OSEM, which allow us to accurately model the statistical properties of corrected data. Finally, we discuss some practical subjects, such as the effect of FORE on the variance, as well as some of the most important features of the used OSEM implementation.

As said before, OSEM is an algebraic, non-bayesian, statistical reconstruction algorithm which, in practice, can be considered as an accelerated version of ML-EM, and, therefore, a very appealing alternative to the latter algorithm. In short, behind the adoption of OSEM as the working algorithm, it is possible to identify three key fundamental decisions:

- the use of a statistical, algebraic, reconstruction algorithm, instead of an analytical;
- the choice for a non-bayesian approach, avoiding the practical difficulties usually associated with the implementation of Bayesian algorithms as well as the control and optimal setting of some of their intrinsic parameters;
- the assumption of a Poisson model for describing data (instead, for example, of a Gaussian), which is just one way to proceed with the investigation, but the way the group of persons taking part in the present work has been following for a few years.

However, even with the computer power available nowadays, OSEM is still slower than FBP when reconstructing images from clinical data¹. Therefore, to make OSEM affordable for clinical routine, in special when 3D reconstructions are required, the time consumption of

¹ On an empirical basis one can roughly say that one OSEM iteration takes as much time as two FBP reconstructions.

OSEM must be reduced. The widespread solution is to rebin data into 2D, using FORE, and, after that, to take these data and perform a 2D OSEM reconstruction.

From a theoretical point of view this is not so straightforward, since the Poisson model on the basis of OSEM must be taken into account. In fact, prior to reconstruction, data are often² corrected for effects such as the attenuation, scattered and random photons, as well as for normalization. In that case, we have shown (section 2.4.1.3) that the relation between corrected data, y_i^{3D} , and the prompts, p_i^{3D} , can be expressed by the following equation (Eq. 2.14): $y_i^{3D} = N_i \cdot \mathcal{A}_i(p_i^{3D} - r_i - s_i)$. In addition, this equation led us to derive the mean value, $\langle y_i^{3D} \rangle$, and the variance, $\sigma^2(y_i^{3D})$, for bin par i , estimated from corrected data:

- $\langle y_i^{3D} \rangle = N_i \cdot \mathcal{A}_i(\langle p_i^{3D} \rangle - \langle r_i \rangle - \langle s_i \rangle)$
- $\sigma^2(y_i^{3D}) = N_i^2 \cdot \mathcal{A}_i^2(\langle p_i^{3D} \rangle + \langle r_i \rangle)$

Since these two moments for the variable y_i are not equal, it is no longer possible to admit that corrected data are Poisson distributed. This is central in the context of the present work: in one hand, OSEM assumes Poisson data; on the other, corrected data are not Poisson distributed. This means that the major premise of OSEM is not verified for corrected data and, therefore, the use of OSEM with these data is abusive (nevertheless, with satisfying results!). The solution for such contradiction can essentially be one of three: to simply ignore it and proceed with data reconstruction using regular OSEM³; not to correct data before reconstruction and include corrections in the iterative equation (this is the strategy in the core of the Ordinary-Poisson OSEM3D [Comtat, 04], but it is incompatible with the use FORE); or, third option, to devise and use a weighted version of OSEM for the reconstruction of corrected data, whose main goal is to partially restore the Poisson characteristics of the data. This last approach is discussed in the following section.

3.1 – Weighted OSEM

The weighted OSEM [Michel, 99] was conceived for reconstructing non-Poisson corrected data. In the original weighted approach, corrected data are scaled by a multiplicative

² When data are to be Fourier rebinned, there is no alternative and should always be first corrected for such effects [Liu, 01].

³ We recall: it works, but with suboptimal results.

weighting factor, W , to partially restore a Poisson-like distribution, i.e., to grant data, as completely as possible, the fundamental feature of a Poisson distribution: mean equal to variance. The iterative equation driving a pure multiplicative weighted OSEM algorithm is the following:

$$\lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{\sum_i w_i a_{ij}} \sum_i a_{ij} \frac{w_i \cdot \max\{y_i, 0\}}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)}} \quad \text{Eq. 3.1}$$

where λ_j is the value on image voxel j , y_i is the input sinogram with corrected data, a_{ij} is the system matrix element for voxel j and sinogram element i , and w_i is the weighting factor. The zero-thresholding, $\max\{y_i, 0\}$, is needed to avoid negative values that OSEM is not able to handle. Comparing to Eq. 2.22 (for regular OSEM, without weighting), the difference relies in the presence of this latter factor, w_i .

In addition, the expression of the multiplicative weight is also the key difference between different weighted OSEM schemes: the extension to which weighted data are given Poisson-like characteristics depends fundamentally on this factor. The most common approaches are:

- AWOSEM [Hebert, 90b]: nowadays routinely used for clinical studies, it just includes the Attenuation Correction Factors (ACFs) in the weighting.

Therefore: $w_i = 1/A_i$, where A_i are the ACFs;

- ANWOSEM [Michel, 99]: it can be thought as an extension of the AWOSEM, where both the attenuation correction and the normalization factors, N_p , are taken into account in the definition of the weight:

$$w_i = 1/(A_i \cdot N_i).$$

Although attenuation correction is responsible for the major deviation from the Poisson model in whole-body PET imaging [Comtat, 98], weighting data according to one of the approaches referred above, does not result in Poisson-like data: it gives data properties closer to Poisson, as expected, but the fully required equivalence between the two first data moments is not met. So, these two weighted versions (AWOSEM and ANWOSEM) still don't handle Poisson data, as assumed by the OSEM algorithm.

In order to achieve this goal, two other weighted versions were devised: NECOSEM, with capital importance for the present work and discussed separately in the following section; and the Shifted-Poisson approach (SPOSEM), a fully 3D weighted reconstruction⁴.

The distinctive feature of the SPOSEM3D [Yavuz, 97] (fully 3D) approach is the existence of an additive weight, in addition to the multiplicative factor that is present in all the OSEM versions we have been reporting. It was introduced mainly to overcome the consequences of on-line random subtraction – through the use of the delayed window technique referred in Chapter 1 -, in most clinical systems, a procedure that blocks the access to raw data, i.e., the prompts. This has two important consequences: first, while both the prompts and the delayed are Poisson variables, the difference is not⁵; second, negative values can appear in the subtracted data. But, since OSEM can not handle negative data, one is obligated to set to zero all the negative values (zero-thresholding), which may lead to an overestimation at the center of the FOV [Liu, 01]. Therefore, the additive weight not only intrinsically reduces the existence of negative values⁶ as well as, reconstructing the variable $v = t + 2\langle d \rangle$, recovers the reconstruction of a Poisson variable: for $v = p - d + 2\langle d \rangle$, $\langle v \rangle = \langle p \rangle + \langle d \rangle$ and $\sigma^2(v) = \sigma^2(t) = \langle p \rangle + \langle d \rangle$, i.e., these two moments are equal.

The driving iterative equation of SPOSEM3D is:

$$\lambda_j^{(n+1)} = \frac{\lambda_j^{(n)}}{\sum_i w_i a_{ij}} \sum_i a_{ij} \frac{\max\{t_i + 2\langle d_i \rangle, 0\}}{\sum_{k=1}^m a_{ik} \lambda_k^{(n)} + \frac{2\langle d_i \rangle + s_i}{w_i}} \quad \text{Eq. 3.2}$$

where t_i are the true coincidences, and d_i and s_i the correction factors respectively for randoms and scattered coincidences, assuming online random subtraction with a delayed coincidence window.

⁴ All the weighted OSEM versions referred above (AWOSEM, ANWOSEM, NECOSEM, and SPOSEM) were used in practice for image reconstruction, in the context of the experimental work the present thesis is devoted to.

⁵ In fact, for $t = p - d$ (t : trues, p : prompts, and d : delayed), the mean and the variance of this variable are $\langle t \rangle = \langle p \rangle - \langle d \rangle$ and $\sigma^2(t) = \langle p \rangle + \langle d \rangle$, respectively.

⁶ Their existence is still possible, but much less likely.

3.1.1 – NEC transform and NEC-OSEM

The NEC-OSEM approach makes use of the NEC transformation methods proposed by J. Nuyts *et al* [Nuyts, 01], to transform the sinogram values to their noise equivalent counts, such that the two statistical first moments of the transformed data become equal.

As Nuyts refers, the NEC transformation is not unique. So, three different strategies can be used:

- the NEC-scaling, where the scaling factor, w , is defined by: $w = \frac{\langle y \rangle}{\sigma^2(y)}$;
- the NEC-shifting, that uses, instead, an offset factor, g , defined by: $g = \sigma^2(y) - \langle y \rangle$;
- a combination of the former two, known as the general NEC-method, that is also useful for avoiding bias due to the zero thresholding of negative values resulting from data corrections. In this case, the shift parameter is $a_i = \frac{\langle x_i \rangle + g_i}{\sigma^2(x_i)}$, where g_i is a function designed to ensure that $x_i + g_i > 0$ [Nuyts, 01].

In the context of the present work, we will detail the NEC-scaling approach, since it was the one adopted for both simulation and clinical data studies.

It is important, therefore, at first, to clearly understand why this sort of weighting is designated as NEC weighting. If we consider a Poisson variable, $L: L \cap P(\lambda)$, for that variable: $\langle l \rangle = \sigma^2(l) = \lambda$. So, its signal-to-noise ratio is given by: $\text{SNR}_l = \frac{\langle l \rangle}{\sigma(l)} = \sqrt{\lambda}$. Since, by definition, the NEC of x (another variable, not Poisson distributed) is equal to the mean of the Poisson distribution with the same SNR as X [Nuyts, 01], it follows that the NEC of x equals the square of its SNR. Schematically:

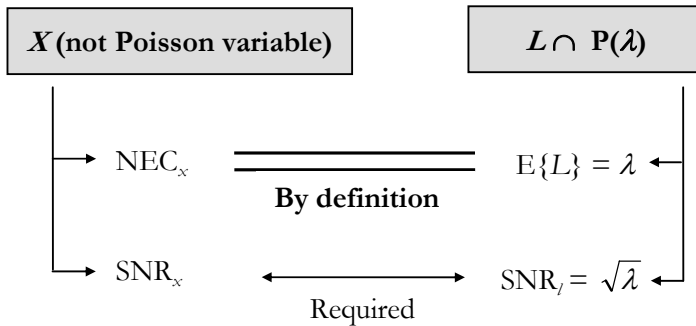


Figure 3.1

Schematics of the NEC definition

So, if, by definition, $\text{NEC}_x = \lambda = (\text{SNR}_x)^2$, then, from the requirement that the SNR of the two variables must be equal:

$$\text{NEC}_x = (\text{SNR})^2 = (\text{SNR}_y)^2 = \frac{\langle y \rangle^2}{\sigma^2(y)} \quad \text{Eq. 3.3}$$

Weighting data using the NEC scaling approach we have been describing corresponds to multiply corrected data, y , by the weighting factor $w = \frac{\langle y \rangle}{\sigma^2(y)}$:

$$z_i = w_i y_i = y_i \frac{\langle y_i \rangle}{\sigma^2(y)} \approx \frac{\langle y_i \rangle^2}{\sigma^2(y_i)} = (\text{NEC})_{z_i}$$

From the expressions above we easily understand the fact that weighted data, z , are approximately equal to the NEC of corrected data, y , is the reason for calling this sort of weighting, “NEC weighting”.

For verifying that NEC weighting restores the Poisson characteristic of data, let us consider the mean value and the variance of the NEC weighted data:

- $\langle z_i \rangle = \langle y_i \rangle \cdot w_i = \frac{\langle y_i \rangle^2}{\sigma^2(y_i)}$
- $\sigma^2(z_i) = \sigma^2(y_i) \cdot w_i^2 = \sigma^2(y_i) \cdot \frac{\langle y_i \rangle^2}{\sigma^4(y_i)} = \frac{\langle y_i \rangle^2}{\sigma^2(y_i)}$

From the comparison of the two equations above, we conclude that these two moments are equal, and, therefore, that NEC scaling transforms data into a Poisson-like distribution.

Nevertheless, two important things must be taken into account. The first is the fact that zero thresholding is still needed for granting non-negative values to the reconstruction process.

The second, and most important, is the expression of the weighting factor itself, $w = \frac{\langle y \rangle}{\sigma^2(y)}$.

In fact, to use this approach one needs to know the variance associated with each bin of the corrected sinogram. Since knowing the variance exactly is impossible (except for the case of simulated data), one must devise a procedure for estimating the value of the variance. And, as will be discussed latter on this thesis, this is a very important limiting factor, intrinsic to the NEC-OSEM approach. In addition, if 3D data are to be rebinned into 2D data, the effects of the rebinning algorithm on the data variance must also be included on this estimation.

3.2 – The effects of FORE on the data variance

The FORE algorithm was the rebinning algorithm used in all the studies performed for the research work we are discussing in the present thesis.

As said previously, FORE represents a good compromise between accuracy and implementation feasibility for the rebinning of 3D into 2D data. Nevertheless, if used to provide input for a NEC-OSEM⁷ reconstruction, it is necessary to understand, and model, the effects of FORE on the variance of the rebinned data, because the weighting factor – whose definition requires a value for the variance – is applied to the corrected and rebinned data. In fact:

$$y_i^{2D} = \text{FORE}\{N_i \cdot A_i(p_i^{3D} - r_i - s_i)\} \quad \text{Eq. 3.4}$$

A simple and reasonably accurate model of the effect of FORE on the variance of the corrected data was developed by Comtat *et al* [Comtat, 98]. In the frequency space, the Fourier rebinning can be mathematically resumed as follows⁸:

$$Y^{2D}(w, k, \xi) = \sum_{\xi', \delta} G(w, k, \xi, \xi', \delta) Y^{3D}(w, k, \xi', \delta)$$

where ξ' is the mid-slice axial position (the others are the usually defined variables). G are weighting factor for taking into account effects such as linear interpolation, zeroing the inconsistent regions, and special handling of the low frequency terms [Comtat, 98; Defrise, 97]. Taking the inverse 2D Fourier transform, $y_{2D}(x', \phi, \xi)$:

$$y^{2D}(x', \phi, \xi) = \sum_{x'', \phi'} \sum_{\xi', \delta} a(x' - x'', \phi - \phi', \xi, \xi', \delta) \cdot y^{3D}(x', \phi', \xi', \delta)$$

The $a(x', \phi, \xi, \xi', \delta)$ terms can be calculated for each pair (ξ', δ) by applying the FORE algorithm to a 3D impulse sinogram: 1 at $x' = \phi = 0$ for the pair (ξ', δ) and zero elsewhere [Comtat, 98]. So, assuming uncorrelated data, the variance of the rebinned data can be expressed as:

$$\sigma^2(y^{2D}(x', \phi, \xi)) = \sum_{x'', \phi'} \sum_{\xi', \delta} a^2(x' - x'', \phi - \phi', \xi, \xi', \delta) \cdot \sigma^2(y^{3D}(x', \phi', \xi', \delta))$$

⁷ For the rest of this thesis, except if explicitly referred, NEC-OSEM is synonymous of the NEC-scaling (multiplicative factor) OSEM approach.

⁸ We adopt here the letter y , instead of the literature widespread p , to avoid confusion with designation of the prompts. y refers to corrected data: prompts after corrections for randoms, scattered, normalization and attenuation.

In face of this convolution equation, Comtat *et al.* proceed by simplifying the calculations, making the following assumptions:

- first: for each slice, z , the variance of all 3D sinograms that contribute to z is independent of the ring difference, δ . This is equivalent to make a single slice rebinning (SSRB) of the variance;
- second: $a(x', \phi, z, z', \delta)$ is highly local in x' and ϕ .

Under these two approximations, the variance of the rebinned sinogram can be simplified to [Comtat, 98]:

$$\sigma^2(y^{2D}(x', \phi, z)) = K(z) V_{SSRB}(x', \phi, z)$$

where $V_{SSRB}(x', \phi, z)$ is the result of applying the SSRB to the variance sinogram (for corrected data), and $K(z)$ are slice variance reduction factors:

$$K(z) = \sum_{x', \phi} k(x', \phi, z) = \sum_{x', \phi} \sum_{z', \delta} a^2(x', \phi, z, z', \delta)$$

These factors are very local and depend only on the axial position of the rebinned sinogram. In sum, when data are rebinned in accordance to Eq. 3.4, the corresponding effect on the variance can be stated as follows:

$$\sigma^2(y_i^{2D}) = K(z) SSRB \left\{ N_i^2 \cdot A_i^2 \left(\langle p_i^{3D} \rangle + \langle r_i \rangle \right) \right\} \quad \text{Eq. 3.5}$$

once more, assuming online random subtraction with a delayed coincidence window.

3.3 – The adopted OSEM implementation

In practice, two different implementations of OSEM were used: one for 2D reconstruction and the other for 3D. However, from a conceptual point of view, the 2D version can be understood as a particular case of the 3D one. The main difference between the two versions relies on the need to handle (3D), or not (2D), sinograms with data for oblique plans, i.e., to handle the information in theta (azimuthal angle).

In face of this, we will describe the structure and the most relevant features of the 3D OSEM implementation adopted for image reconstruction in our case. Schematically, the OSEM algorithm can be depicted as in Fig. 3.2.

The main goal of OSEM (as of any other reconstruction algorithm) is to take as input the corrected data and generate the reconstructed image (output). To succeed, the OSEM core – the very reconstruction procedures – requires defining the values of some parameters it will be using. Some of them are definable by the user in the reconstruction stage; the others can be defined before the acquisition, or derived from the scanner characteristics.

Among the former – user definable in the reconstruction stage - we have:

- the number of iterations;
- the number of subsets, which is a particular feature of OSEM and the main difference from ML-EM. As referred before, the number of subsets determines the degree of the acceleration introduced by OSEM when compared to ML-EM (OSEM with a single subset including all the projections). However, each subset must be balanced, which limits, in practice, the maximum number of subsets;
- the dimensions of the output reconstructed image. In particular, the number of pixels in each of the two coordinates (x and y) characterizing the stack of 2D reconstructed images along the axial coordinate. An usual definition is requiring a stack of 2D images, 128×128 pixels each.

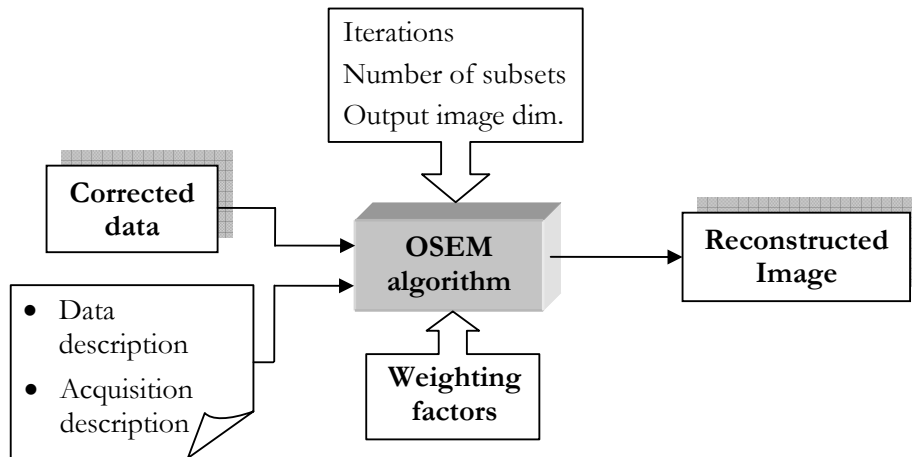


Figure 3.2 – A schematic description of the OSEM core's interface.

In addition to this sort of input, there are other parameters that depend from the acquisition conditions as well as from the scanner model. The most fundamental is to indicate

whether data have been acquired in 2D or 3D mode. The scanner, itself, imposes constraints to the values of some variables: the number of rings, the maximum ring difference, the bin size in both coordinates (transaxial and axial), the value of the span, and the number of views. In the adopted reconstruction package there is a file containing a structure with the description of these and other parameters associated with the scanner model. In another file, the values for this set of parameters are defined for different scanner models.

A data description is contained in the header of the data file. The size of data, and their structure, depends strongly from the acquisition process. For example: the number of segments data are organized in, which is very relevant for the reconstruction process, results from the combining of the number of rings, the ring difference and the value of the span. Data description should also include information about the data type (integer, float or other), their organization in sinogram or projection format, and the segment structure, i.e., in case of 3D data, how data are organized for different values of theta.

The weighting factors are contained in a file which OSEM uses through a flag in the command line. This means that they must be calculated prior to the reconstruction, as expected. However, data weighting⁹ is performed inside the OSEM algorithm, but before the reconstruction core.

```

FOR (iter = 0; iter < iterations; iter++) {
  FOR (subset = 0; subset < number_of_subsets; subset++) {
    FOR (segment = 0; segment < number_of_segments; segment++) {
      Sub-iteration
      Estimate projections;
      Calculate the ratio between measured and estimated projections;
      Backproject the ratio between measured and estimated projections;
      Update the object using normalization and weighting factors;
    }
  }
}

```

This OSEM core is spread among different files and is the computational implementation of Eq. 3.1. It can be summarized as follows, where: “iterations” is the number of iterations defined by the user; “subsets”, the number of subsets also defined by the user; and “number_of_segments”, the number of segments that results from the number of rings, the maximum ring difference and the span. It is a value easily drawn from the Michelogram (cf. Ap.

⁹ What we have mathematically expressed by $\tilde{z}_i = w_i y_i$

A8) corresponding to the particular acquisition to be considered. Since there are positive and negative segments and all should be used, “number_of_segments” is equal to the absolute value of the higher segment, multiplied by two, plus one – for including segment zero (the one corresponding to a 2D acquisition).

Deserving mention is also the strategy devised to handle the azimuthal views (projections) for each subset. In fact, as referred in chapter two, subsets should be balanced, something that is achieved, in practice, by ordering projections so that each time the algorithm uses a different subset, projections should be separated the maximum. In the present case, this done by pairing the subset number (*subset*) with the segment number (*segment*)¹⁰ is a convenient way. For a reconstruction using 144 views, 8 subsets and 5 segments (-2, -1, 0, 1, 2, nevertheless designated, respectively, by a different order number: 3, 1, 0, 2, 4), the pairing of these two variables, for each iteration, is described in Table 3.1.

Sub-iteration	segment	subset	Sub-iteration	segment	subset
0	0	0	20	0	2
1	1	4	21	1	6
2	2	1	22	2	3
3	3	5	23	3	7
4	4	2	24	4	0
5	0	6	25	0	4
6	1	3	26	1	1
7	2	7	27	2	5
8	3	0	28	3	2
9	4	4	29	4	6
10	0	1	30	0	3
11	1	5	31	1	7
12	2	2	32	2	0
13	3	6	33	3	4
14	4	3	34	4	1
15	0	7	35	0	5
16	1	0	36	1	2
17	2	4	37	2	6
18	3	1	38	3	3
19	4	5	39	4	7

Table 3.1 – The order by which the pairs of subsets and segments are handled, in order to grant OSEM balanced projections. This is for 8 subsets and 5 segments (-2, -1, 0, 1, 2 ordered as 3, 1, 0, 2, 4), inside a single iteration.

¹⁰ In fact, the program uses auxiliary variables derived from these two, but this is a detail not relevant for the present discussion.

Since all the segments and subsets must be used inside a single iteration and for each pair (**segment**, **subset**) the object activity is updated, the handling of the corresponding data by the OSEM algorithm is referred as an OSEM sub-iteration. For the case referred, one single OSEM iteration would include 40 subiterations.

In respect to the 144 azimuthal views, they are divided among the 8 subsets, resulting in 18 views per subset. So, the subset number establishes the first view for that subset. The other 17 are determined by successively adding 8 to the first view. For example, subset 4 contains views number 4, 12, 20, 28, 36, 44, 52, 60, 68, 76, 84, 92, 100, 108, 116, 124, 132, 140.

4 – Numerical observer studies comparing different OSEM approaches for 3D whole-body PET

In this chapter we will describe simulation studies whose goal was to compare the performance of 2D and 3D OSEM implementations. They were done prior to the clinical images studies (the other main framework), where some of the results obtained here were directly used.

Since the comparison of an algorithm's performance needs a previous definition of the clinical task that images are going to be used for, this is the first thing to clearly establish: in the present case, the performance was investigated in terms of tumor detectability (and also lesion-to-tissue contrast) in images from 3D whole-body PET. The general aim was, therefore, to verify if an accurate statistical modeling in OSEM improves tumor detectability. In particular, we aimed to clarify two main things: how does FORE+AWOSEM compare with FORE+NECOSEM, investigating if the theoretical improvement behind FORE+NECOSEM – the accurate statistical modeling - results in a practical advantage over FORE+AWOSEM; and how does fully 3D reconstructions compare with FORE followed by a 2D reconstruction.

The most important tool adopted for comparing the performance of the different OSEM implementations was the 3D Non-Prewhitening Matched Filter (NPWMF) numerical observer. In fact, since it was unaffordable to perform this kind of studies with human observers, due to the larger amount of data and images, we had to replace them by an observer whose performance for similar studies and similar conditions mimics the performance of humans. Lartizien [*Lartizien, 04*] has shown in a previous study, for similar imaging conditions, that the NPWMF detectability index allows good correlation with human observer detectability. Therefore, this numerical observer was chosen for the tumor detection task in 3D whole-body PET images.

In addition, the tumor-to-background contrast and the noise were also used in order to compare performances.

The strategy to implement an accurate statistical modeling in OSEM was to use the NEC transform (cf. section 3.1.1). As referred, the NEC scaling (to correct data by using the NEC

weighting factors) restores the Poisson-like characteristics of data. This confers NECOSEM a theoretical advantage over other weighted OSEM versions, such as the AWOSEM or the ANWOSEM, because one major premise of OSEM is almost fulfilled: the requirement for Poisson data. In both AWOSEM and ANWOSEM, on the contrary, even considering that the attenuation correction is the responsible for the major deviation from the Poisson model in whole-body imaging [Comtat, 98] and that such effect is intrinsically accounted for in these algorithms, data are no more Poisson-like. That is the reason why NECOSEM (2D and 3D) is compared to the other weighted versions.

To control many different parameters (such as the original tumor-to-background contrast or the noise level) and to have the possibility to calculate the NEC weights for every data bin using directly Eq. 3.3 (one needs to know exactly what are the values for the mean and variance associated with each bin) it is necessary to proceed with simulation studies. In fact, in clinical data neither it is possible to exactly calculate the NEC weights nor the contrast or noise are controllable.

These simulation studies are the subject of the present chapter. We will describe the analytical simulator and the simulation conditions that we used to generate input data (sinograms and/or projections). The mathematical anthropomorphic phantom for which data were generated is also present.

4.1 – Data simulation

4.1.1 – The anthropomorphic phantom

The anthropomorphic phantom adopted for the generation of simulated data is based on the 3D Mathematic Cardiac Torso (MCAT) phantom [Tsui, 93], with the addition of the head, arms, and bladder objects in order to reproduce the skeleton and the main organs in the abdomen and torso. The development of the particular implementation used in our work is due to Carole Lartizien [Lartizien, 01]. It is conceived as the superposition of many different ellipsoids, each one of them “filled” with a value. With this superposition scheme, each organ gains its own contrast to the background as well as to the other organs.

Once the organs are defined, it is possible to simulate tumors therein, by adding spheres whose locations lie inside the chosen organ. As an example, Fig. 4.1 shows the aforementioned

phantom where it is possible to observe one tumor in the left lung, one in the right lung and another in the liver.

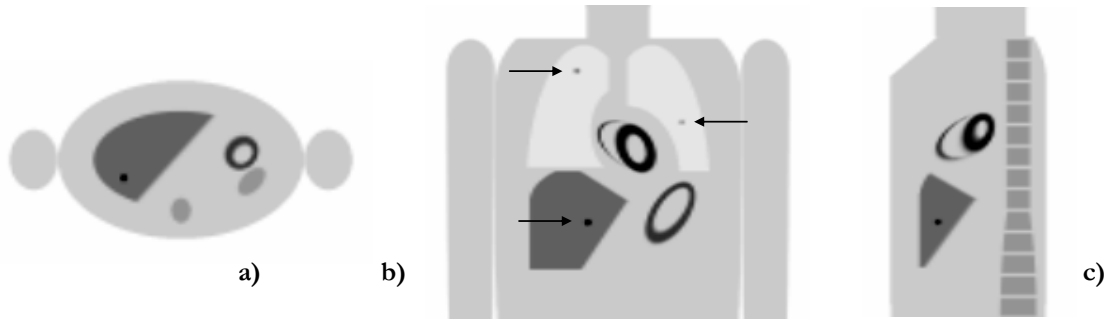


Figure 4.1 – Axial (a), coronal (b) and sagittal views of the anthropomorphic phantom used for the simulations. Three tumors are visible and indicated by arrows: one the left lung, one in the right lung, and another in the liver.

4.1.2 – The analytical simulator (ASIM)

The ASIM analytical (non Monte-Carlo¹) simulator [Comtat, 99] was used for generating multiple realizations of 3D whole-body datasets. The principle of ASIM is to first analytically calculate noiseless transmission and attenuated emission sinograms on the basis of geometric specifications of the emission and attenuation objects (i.e., the anthropomorphic phantom), the geometry of the considered scanner, the position of the objects in the FOV, and the number of bed positions (acquisitions) in case of multi-bed whole-body acquisitions [Comtat, 99; Lartizien, 03]. In addition, sinograms of noiseless random and scattered coincidences are also generated, with the assumption, however, that the fraction of scattered coincidences in the transmission scan is not significant and that there is no contamination from the emission objects to the transmission sinogram. The developers of the simulator also stress that the goal is not to accurately simulate the detection of randoms and scattered, but rather to simulate their effect on the emission and transmission scans [Comtat, 99]. For the calculation of these randoms and scattered sinograms is used a model where the total activity in each slice is assumed to be concentrated along the scanner axis, and instead of the attenuation medium specified by the phantom, it is used an uniform cylinder.

¹ The amount of data required for this study, namely the need of many replicas, would be not feasible with Monte Carlo techniques.

If a mathematical expression is to be given to the sinograms for scattered and random coincidences, it includes two different components, one for axial and other for radial directions. The radial profiles of the sinograms containing the randoms (emission and transmission) and the scattered (emission) are not calculated, but given as an input to the simulations. These profiles, are typically estimated by very high-count scans of a cylinder with an uniform activity and are used regardless of the geometry of the simulated phantom [Comtat, 99].

On the contrary, axial profiles are calculated due to the greater dependence on source distribution. For random coincidences, the axial profile is established by the single photon axial flux on the detector pair in coincidence, considering only non-scattered photons. The estimation of the axial profile for scattered photons (in the emission sinogram) assumes that only one of the two emitted photons undergoes a single Compton interaction. However, this is a more complicated process than in the case of randoms, and involves a Monte Carlo integration technique based on the Klein-Nishina Compton scatter cross-section that includes both the energy resolution and the lower energy threshold of the detectors [Comtat, 99].

In sum: ASIM is able to simulate five fundamental sinograms – the true coincidences and the randoms for both emission and transmission scans; and the scattered for the emission scan. These sinograms are firstly generated without noise. In order to produce a realistic simulation, ASIM is also able to add Poisson noise to the aforementioned sinograms. The process is the following: first, the simulator is given the total number of true (T_E, T_T), random (R_E, R_T), and scattered coincidences (S_E) for some range of bed positions (as input information). With this information plus the half-life of the isotope, the scan duration, and the scan start type, the simulator computes global scale factors: $\alpha_{t_E}, \alpha_{r_E}, \alpha_{s_E}, \alpha_{t_T}$, and α_{r_T} . The values for the noiseless sinogram elements (t_E, r_E, s_E, t_T , and r_T) are computed based on the specified geometries of the phantom and the scanner, the bed position, and, for multi-bed acquisitions, the number of bed positions and bed overlap. The global scale factors are then used in the process of adding pseudo-random noise to the noiseless sinograms, in accordance to the following equations [Comtat, 99]:

$$\begin{aligned}\tilde{p}_E &= P[\alpha_{t_E} \cdot t_E + \alpha_{r_E} \cdot r_E + \alpha_{s_E} \cdot s_E] \\ \tilde{p}_T &= P[\alpha_{t_T} \cdot t_T + \alpha_{r_T} \cdot r_T]\end{aligned}$$

where \tilde{p} are the noisy prompt coincidences and $P[m]$ is a Poisson pseudo-random realization for the mean m .

ASIM also accounts for many other different important effects that occur in real clinical whole-body PET imaging, such as random and scattered coincidences arising from the activity outside the FOV, detector efficiencies, activity decay between bed positions, and the spatially

variant detector response function. The simulator was validated by comparing means and variances measured from multiple realizations of simulated and measured studies based on identical phantoms, with several activity levels inside and outside the FOV [Lartizien, 03].

4.1.3 – Simulation conditions

For the simulation of tumors, 1 cm diameter [Lartizien, 01] spheres were generated and added to the anthropomorphic phantom depicted in Fig. 4.1. It was established, taking into account the dimensions of the organs and what seemed to be reasonable for this purpose, that these tumors would be in the number of 86: 34 in the lungs (distributed among the left and right lung), 22 in the liver, and 30 locations in the soft tissues. With the constraint of the organ they were to be added to, each tumor location was randomly generated².

However, to avoid an image visually saturated with lesions as well as to properly use the numerical observer (no pile up of the tails of the profile involved in the observer's expression, cf. A.7), the tumors were not included all in one single phantom, but equitably distributed among 6 replicas of the same anthropomorphic phantom.

As referred, from a structural point of view, the phantom is the result of a superposition of different objects. So, setting the activity of each object properly it is possible to define the activity of every organ in the phantom: the final activity of one organ will be the sum of the activities of all the objects that are superposed in order to shape the organ. Using this property, one can also establish the activity ratio between the lesion and the surrounding tissue.

In the studies we are describing, a first activity ratio, $AR_{50\%}$, was chosen such that the fraction of lesions found by a human observer was about 50%, for an average patient size, as described by Lartizien [Lartizien, 03] for a similar study. In parallel, another study was performed for a lower activity ratio, AR_{low} , just to investigate if any relevant difference is found between the reference activity ratio, $AR_{50\%}$ and the lower one. Therefore, it was used an activity ratio one unit below, in each organ, than the corresponding reference ratio, a value adopted with no other objective than setting a lower activity ratio. The lesion to tissue activity ratios are listed in Table 4.1.

² It was observed in a preliminary test, that increasing the number of tumors per organ while keeping their locations randomly distributed (within each particular organ) would result (especially in the liver and the lungs) in tails overlapping when using the NPWMF observer.

Location of tumor	AR _{50%}	AR _{low}
Soft tissues	7.5	6.5
Liver	3.0	2.0
Lungs	6.5	5.5

Table 4.1 - Lesion to tissue activity ratios³

For each replica of the anthropomorphic phantom, 25 independent noisy realizations were generated, using the geometry of the EXACT ECAT HR+ scanner. Four bed positions were simulated, and the following conditions were assumed:

- maximum ring difference: 22
- span: 9
- bed overlap: corresponding to 5 axial planes.

The third bed position was chosen in order to establish the number of counts in each bed position⁴. For this bed positions ASIM was given the following input: a total of 37×10^6 trues, 90×10^6 random coincidences, and 45×10^6 scattered coincidences. As referred in the previous section, these values are used in the process of adding Poisson noise to the sinograms.

A flowchart of the file generation chain for each replica of the anthropomorphic phantom is represented in Fig. 4.2. Shadow boxes correspond to files used as input in the reconstruction stage (which is not included in the flowchart). Programs' names are in courier font.

The first step in the simulation is the generation of the noiseless emission sinogram, which is the most fundamental file on the basis of the rest of the chain, from the definition of the anthropomorphic phantom. This is main task of ASIM, that uses all the input information about the scanner, the simulation conditions and other relevant intrinsic user defined parameters. In parallel, a voxelized image on the phantom is created for visual inspection.

Another fundamental file is the one containing the Attenuation Correction Factors. It is also generated by ASIM, on the basis of the phantom description and detailed input information. The use of these ACFs as weighting factors in the AWOSEM reconstruction requires an assemble bed operation, because the reconstruction of a single sinogram is done after assembling the different bed positions.

³ These values are defined exactly as ratios and not as a contrast, i.e., the ratios are not subtracted by one.

⁴ ASIM takes one bed as reference and calculates the values for the others.

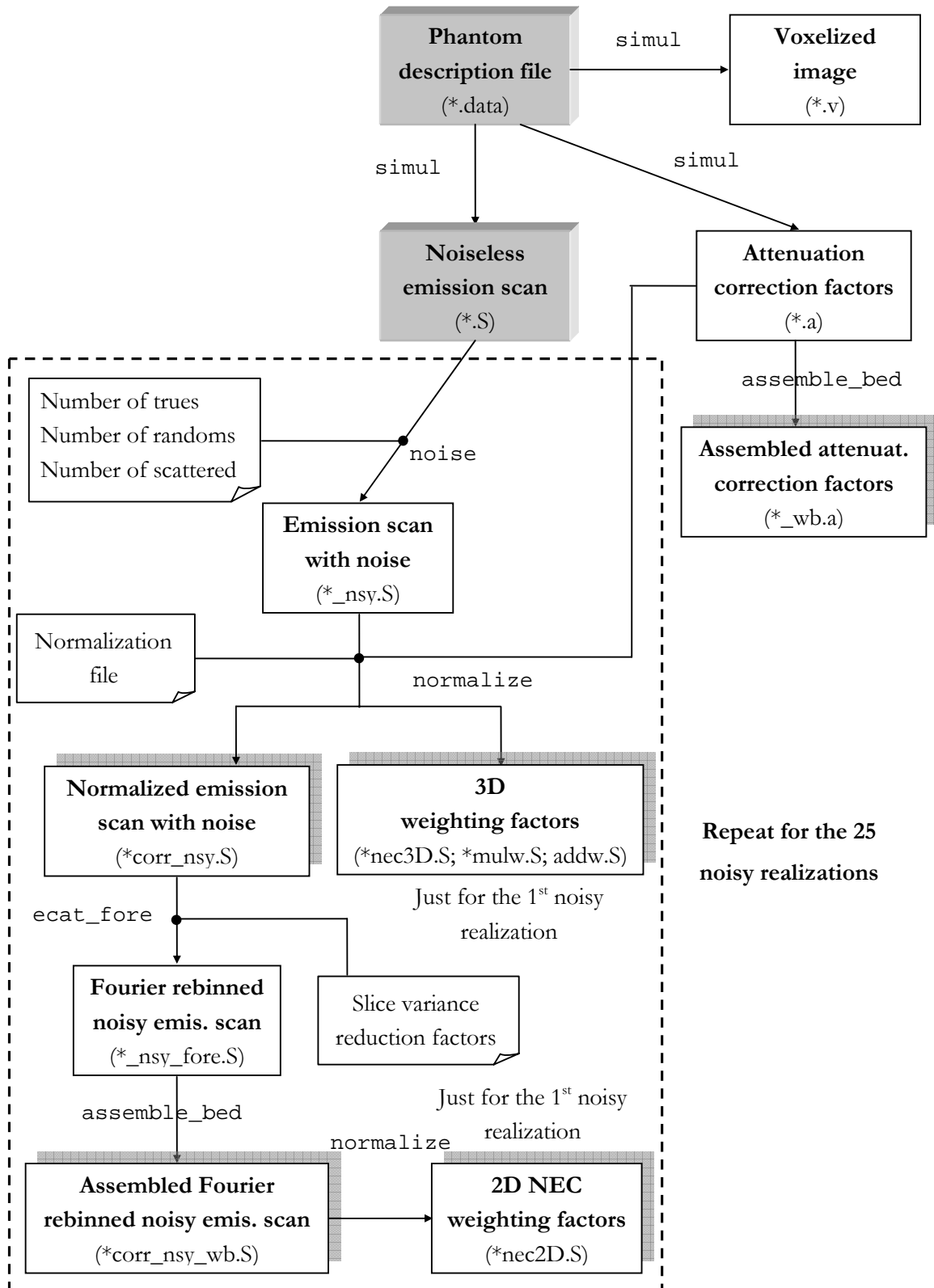


Figure 4.2 – A flowchart of the file generation chain for each replica of the anthropomorphic phantom (reconstruction stage not included).

In addition to the fundamental noiseless emission sinogram and ACFs' file, there is a third sinogram playing a fundamental role in all the simulation process: it is the normalization sinogram. Once more, with the correct input data, ASIM is able to generate such a file⁵ and, after, to use it for normalizing sinograms as well to take part in the creation of the multiplicative and additive 3D weighting factors.

With these three major files (noiseless sinogram, ACFs' file, and normalization file), ASIM enters the loop on the noisy realizations, 25 times. However, the script driving this loop is structured in such a way that some files are created just once, i.e, for the first noisy realization. Once they are created, they can be used in the remaining 24, because the information in there is common to all the replicas (for example, the multiplicative and additive weights, as well as the NEC weight).

For all the noisy realizations except the first, each time the script reenters the loop, the way it proceeds is straightforward: the number of true, random, and scattered events is used to add pseudo-random noise. Then, the noisy sinogram is normalized, corrected for attenuation, scatters and randoms, (using the normalization file) and, since we aim also to reconstruct in 2D, the corrected noisy sinogram is Fourier rebinned and their corresponding bed positions assembled into a single axially extended file, containing information for all the transaxial planes to be reconstructed. This latter sinogram and the 3D normalized noisy emission scan are used as input data for the reconstruction⁶.

In addition to these files, the first noisy realization was used for generating other important sinograms, required by the weighted OSEM approaches adopted for image reconstruction, in particular by NECOSEM.

Since the multiplicative weighting factor taking part in NECOSEM is defined as a ratio between the mean and the variance, the exact value of this factor can only be determined for one sinogram bin if the mean and the variance are known for that individual bin. This is a key point in all the research work done for this thesis, because it establishes a deep distinction between simulation studies and clinical data handling for reconstruction: while within a simulation is possible to keep knowledge about the mean and the variance used for adding pseudo random noise to each individual bin, in a sinogram from a real clinical scan there is a single value per bin – the counts recorded for the corresponding LOR. So, for real clinical data, in contrast to simulation studies, there is no possibility to accurately calculate the NEC weighting factors.

⁵ Something that it is not represented in the flowchart of Fig. 4.2.

⁶ It is not possible to assemble bed for 3D sinograms.

The requirement of keeping knowledge about the mean and the variance for all the individual bins in the emission sinogram is fulfilled by storing in the sinogram file corresponding to the first noisy realization, not only the simulated counts (as for the other 24 noisy realizations or for real clinical data), but also two other sinograms: one containing the mean values, the other the variance for each individual bin. In practice, these three sinograms (emission, mean, and variance) are organized inside the same sinogram file by resorting to the expedient of using three different gates: gate 1 for the emission sinogram, gate 2 for the variance sinogram, and gate 3 for the mean values sinogram. Because all the 25 noisy realizations for one replica of the anthropomorphic phantom are based on the same mean values and variances, it is enough to store the two latter sinograms only once.

After normalizing the noisy emission scan, an additional operation is done for the first noisy realization: the calculation of the 3D NEC weighting factors. They are:

- the 3D NEC weights: due to the computational implementation of the algorithm, the NEC weights are, in fact, the reciprocals of the NEC factors. So, the variance is divided by the mean value, or, considering the sinogram file structure, values inside gate 2 are divided by those in gate 3 (cf. App. A8);
- the multiplicative weights: to be used in ANWOSEM and, therefore, mathematically given by $w_i = 1/(A_i \cdot N_i)$, where A_i are the ACFs and N_i the normalization factors (previously simulated by ASIM);
- the additive weights: established in accordance to Eq. 3.2 (SPOSEM3D driving equation).

A comparison between the radial profiles of the multiplicative and the NEC 3D weighting factors is graphically represented in Figure 4.3. This corresponds to the simulated weights obtained from one of the simulated replicas⁷ of the anthropomorphic phantom, for the second bed position and axial plane 42 (in 63) within segment 0 (polar angle equal to zero). The profile has 288 values, since this is the number of radial bins.

With the same purpose – calculation of NEC weighting factors –, in parallel to the Fourier rebinning of the noisy emission sinogram (gate 1), gate 3 of the same file, i.e., the mean values sinogram, is also Fourier rebinned. However, the effect of FORE on the data variance, as referred in section 3.2, can be modeled by single-slice rebinning (SSRB) the variance, and multiply the result by the slice variance reduction factors, $K(z)$ (that depend only on the axial position of the transaxial plane). Therefore, for the variance data (gate 2), the Fourier rebinning

⁷ These weights do not depend on the replica.

operation is replaced by a SSRB, followed by the multiplication of the result by the slice variance reduction factors. The values for these factors⁸, one per transaxial plane, are contained in one simple dedicated text file (as indicated in Fig. 4.2).

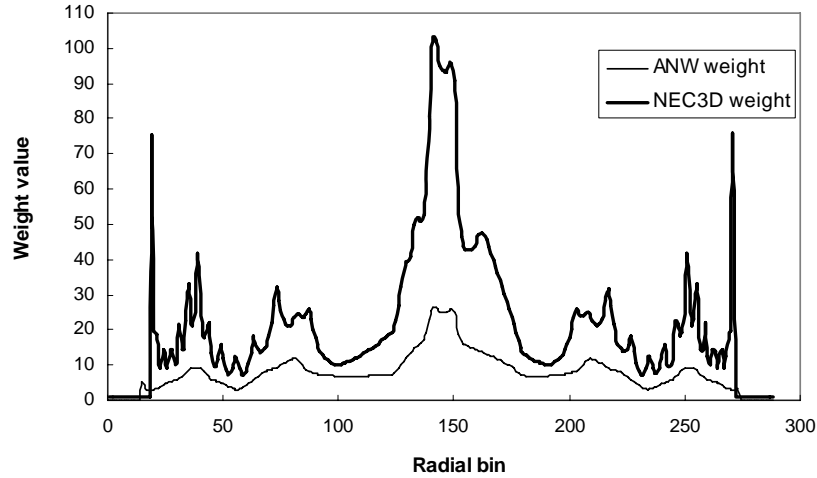


Figure 4.3 – Radial profiles of the multiplicative (ANW) and NEC 3D weighing factors. Profiles are for the second bed position, axial plane 42 of segment 0 and azimuthal angle equal to 0, for one replica of the anthropomorphic phantom.

After the FORE stage, data are assembled in terms of bed positions, and once this is done, the 2D NEC weighting factors are finally determined by just dividing, for each sinogram bin, the mean value by the corresponding variance, i.e., dividing gate 3 by gate 2 values⁹.

4.1.3.1 – Calculation of the slice variance reduction factors

The calculation of the slice variance reduction factors, $K(z)$, is done only once, since all the simulations are for the same scanner and, therefore, the normalization file is also the same. As referred, these factors depend only on the axial position of the transaxial plane.

A uniform cylinder was considered for the simulation on the basis of this calculation, because it is a very simple object with a geometry equal to the scanner's geometry. After generating the noiseless emission sinogram, 50 noisy realizations are created by adding pseudo-random noise to the former sinogram. Then, all these noisy realizations are Fourier rebinned. Once this task is completed, the variance across the 50 noisy realizations is determined for each bin in the sinogram, and the sum of all the variance values (one per bin) in each transaxial plane

⁸ To be more exact, their reciprocals.

⁹ All programs keep track and update the values of the mean (gate 3) and variance (gate 2).

is recorded in a variable named “ $var_fore(z)$ ”. So, this variable is, in fact, a vector with 63 elements (one per transaxial plane).

In parallel, the variance used for adding Poisson noise to the original noiseless emission scan is stored and subsequently rebinned using the SSRB algorithm. The values in each transaxial plane of this rebinned variance sinogram are summed and the resulting value attributed to another variable, “ $var_srb(z)$ ”.

With the values contained in “ $var_fore(z)$ ” and “ $var_srb(z)$ ”, the slice variance reduction factors are mathematically established according to the following equation:

$$K(z) = \frac{var_fore(z)}{var_srb(z)}$$

The reciprocals of these factors are graphically represented in Figure 4.4.

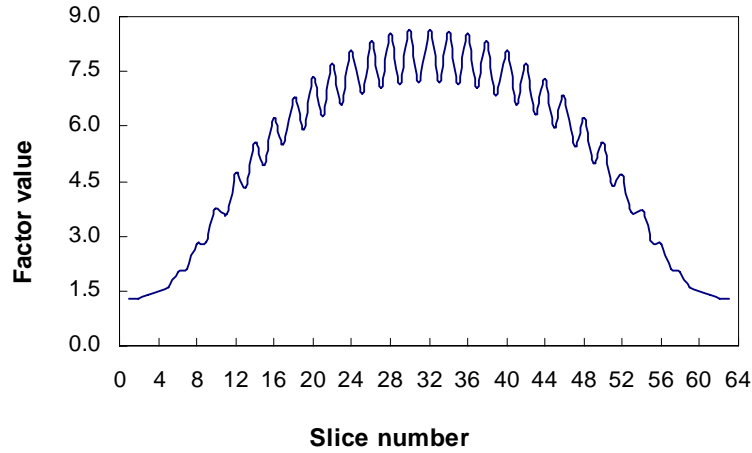


Figure 4.4 – The reciprocal values of the slice variance reduction factors, $K(z)$.

4.2 – Image reconstruction

The simulated data were reconstructed with five different weighted versions of the OSEM algorithm: FORE+AWOSEM, FORE+NECOSEM, ANWOSEM3D, NECOSEM3D, and SPOSEM3D. Figure 4.5 illustrates how the output of the simulation is used in the reconstruction stage.

With the exception of SPOSEM3D, all approaches use just a multiplicative weighting factor and can be mathematically expressed as in Eq. 3.1. In SPOSEM3D (Eq 3.2), there is also an additive weight, that introduces the additional need of an estimation of the mean for the

randoms, $\langle d_i \rangle$. The multiplicative factor, as in the case of ANWOSEM, is given by $w_i = 1/(A_i \cdot N_i)$.

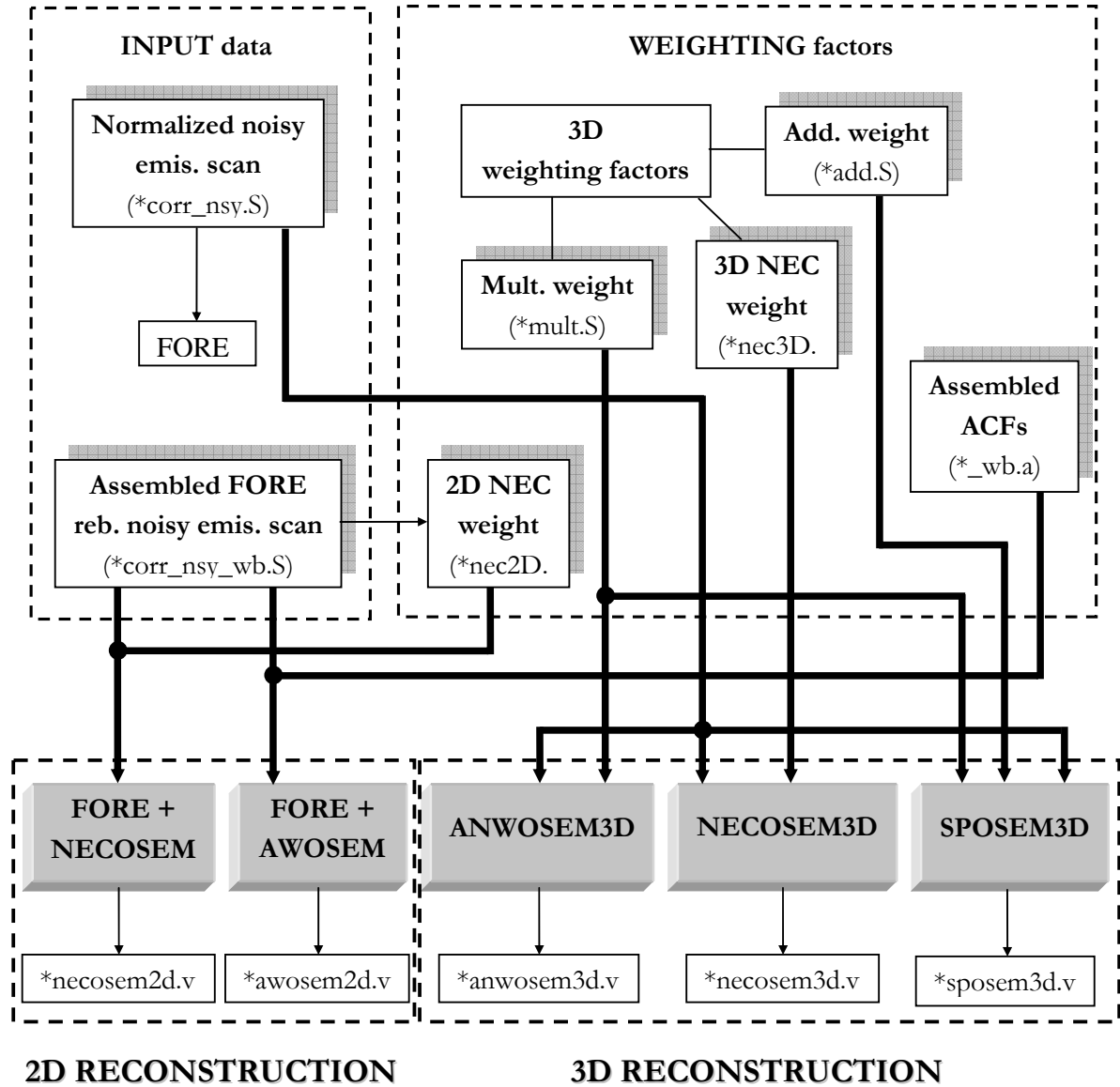


Figure 4.5 – A schematic representation of the five different weighted versions of OSEM used for reconstructing simulated data.

The implementation of NECOSEM3D, combined with the way the NEC weights were established (variance over mean value), requires caution in the use of these weights. In fact, for bins corresponding to LORs that do not cross the object, the mean value (gate 3) will be equal to zero. Therefore, without any procedure to avoid this situation, the computer will be required to

perform a division by a zero value. The strategy adopted to overcome this problem was to define the object boundaries by using the attenuation map: sinogram bins where the ACFs are equal to 1, correspond to LORs that are not attenuated and, thus, do not cross the object.

However, an additional problem was found for bins in the borders of the object: the mean values for such bins are very small and the resulting NEC weights would be extremely high, in sharp contrast to the situation outside the object where the NEC weights were forced to be equal to 1. So, to avoid sharp transitions in the NEC weighting factors, instead of exactly 1, a threshold slightly higher than 1 was used in the attenuation map to discard a region over the borders where the NEC weights would be extremely high due to the rather low mean values on the emission sinogram. In sum, the NEC weighting was expressed as:

- for LORs outside the object or crossing the borders: $w_{NEC3D} = 1$ (Fig. 4.3);
- for LORs crossing the object: $w_{NEC3D} = \frac{\sigma^2(y)}{\langle y \rangle}$.

The threshold used within the attenuation map to distinguish these two situations – attenuation superior to 10% - was established empirically, after testing many different values. This fact, however, raises a few questions that will be addressed later on this chapter.

In all OSEM reconstructions, eight subsets were used. This value was adopted mainly based on practice, since it is usually assumed that the result of an OSEM reconstruction after a certain number of iterations is equivalent to the result of a ML-EM reconstruction after iterating a number of times equal to the product of the OSEM iterations by the number of subsets. Under such a premise, four iterations of these weighted OSEM versions (except for SPOSEM3D) would be directly comparable to 32 iterations of the ML-EM algorithm – a relation people usually find useful in practice.

The maximum number of iterations was limited to 12, except for the SPOSEM3D algorithm, where a ceiling of 20 iterations was defined. This is because the additive term relents the convergence rate of SPOSEM and, therefore, it should be allowed to iterate more.

A 3D Gaussian post-smoothing filter was applied to the reconstructed images in order to improve the tumor detectability.¹⁰

¹⁰ In a previous study, for similar conditions, Lartizien et al. [Lartizien, 00] used a post-smoothing filter for maximizing the SNR on reconstructed images.

4.3 – Image quality assessment

In medical imaging, the definition of image quality is intrinsically dependent of the clinical procedure. In the present case, the goal, as said in the introduction of this chapter, is to detect tumors in whole-body PET imaging. Therefore, from this point of view, image quality will be directly dependent on the improvement on the ability of the human observer (physician, in most of the cases) to perform the detection task. Since the dimension of the sample (number of images to be analyzed) is enormous, relying on human observers becomes impracticable. The 3D Non-Prewhitening Matched Filter (NPWMF) numerical observer was chosen.

An image contrast versus image noise study was also performed in order to complete the information from the numerical observer study.

4.3.1 – NPWMF observer study

The NPWMF observer was used since it has been shown in a previous study for similar imaging conditions that the NPWMF detectability index allows good correlation with human observer detectability [Lartizien, 04]. In fact, even if the tasks under comparison were different (detection and localization for the case of the human observer whereas signal known exactly for the numerical observer), both observers gave similar ranking orders for the reconstruction approaches under study.

The NPWMF numerical observer uses as a decision variable, d_{NPW} , the inner product of the image with the noise-free volumetric target (tumor) profile in the image, and the corresponding detectability index is given by (cf. Appendix A.7):

$$d_{NPW} = \frac{\langle \mathbf{s} \cdot \mathbf{g}_1 \rangle - \langle \mathbf{s} \cdot \mathbf{g}_0 \rangle}{\sqrt{\frac{1}{2} \cdot (\sigma^2(\mathbf{s} \cdot \mathbf{g}_1) + \sigma^2(\mathbf{s} \cdot \mathbf{g}_0))}} \quad \text{Eq. 4.1}$$

where \mathbf{g}_0 and \mathbf{g}_1 are respectively the vectors of the image voxel values of the signal-absent (no lesion) and signal-present (with lesion) class, \mathbf{s} the vector of the expected signal profile, $\langle \cdot \rangle$ and $\sigma^2(\cdot)$ the average and the variance across the realizations.

Lartizien et al. [Lartizien, 04] have also studied the minimum number of realizations that would lead to a stable estimate of the variance terms in Eq. 4.1, concluding that the variance

estimates stabilizes after approximately 25 realizations. So, this is the justification for the number of realizations adopted for our own study.

As required by the NMWMF index, in addition to the noisy, signal-present, emission sinograms whose simulation is described in section 4.1.3, two other sets of sinograms had to be generated:

- noisy emission sinograms of the same anthropomorphic phantom, without any signal present, i.e., with no tumors added. The same number of noisy realizations, 25, was simulated, under the same simulation conditions;
- noiseless emission sinograms of the tumors alone, using, again, the same simulation conditions. Since there is no noise, these sinograms needed to be generated only once for each of the 6 tumor sets associated with the 6 replicas of the anthropomorphic phantom.

All of these additional sinograms were reconstructed with the same OSEM algorithms used for reconstructing former noisy emission sinograms of the phantom with tumors. The same number of iterations was also used and, when calculating the NPWMF index for a certain iteration number, the image vectors and the template (expected signal profile) in Eq. 4.1 were based on reconstructed images after that particular number of iterations. For example, if the NPWMF is calculated for iteration number 4, this means that all the reconstructed images, i.e., anthropomorphic phantom with tumors, anthropomorphic phantom without tumors, and tumors alone, are those obtained after 4 iterations of the OSEM algorithm under study.

The standard deviation associated with the definition of the NPWMF index (Figures 4.11 – 4.15) can be estimated based on the error propagation formula [Bevington, 94]:

$$\sigma^2(d_{NPW}) = \left(\frac{\partial d_{NPW}}{\partial l_1} \right)^2 \sigma^2(l_1) + \left(\frac{\partial d_{NPW}}{\partial l_0} \right)^2 \sigma^2(l_0) + \left(\frac{\partial d_{NPW}}{\partial \sigma(l_1)} \right)^2 \sigma^2(\sigma(l_1)) + \left(\frac{\partial d_{NPW}}{\partial \sigma(l_0)} \right)^2 \sigma^2(\sigma(l_0))$$

$$\text{where } l_1 = \langle \mathbf{s} \cdot \mathbf{g}_1 \rangle, \quad l_0 = \langle \mathbf{s} \cdot \mathbf{g}_0 \rangle, \quad \sigma(l_1) = \frac{\sigma(l_1)}{\sqrt{N_1}}, \quad \sigma(l_0) = \frac{\sigma(l_0)}{\sqrt{N_0}}, \quad \sigma(\sigma(l_1)) = \frac{\sigma(l_1)}{\sqrt{N_1 - 1}},$$

$$\sigma(\sigma(l_0)) = \frac{\sigma(l_0)}{\sqrt{N_0 - 1}}. \quad \text{Combining these expression, the final result for } \sigma^2(d_{NPW}) \text{ is given by}$$

equation 4.2:

$$\sigma^2(d_{NPW}) = \frac{2}{\sigma^2(l_1) + \sigma^2(l_0)} \left[\frac{\sigma^2(l_0)}{N_0} + \frac{\sigma^2(l_1)}{N_1} + \frac{(d_{NPW})^2}{2(\sigma^2(l_1) + \sigma^2(l_0))} \left(\frac{\sigma^2(l_0)}{N_0 - 1} + \frac{\sigma^2(l_1)}{N_1 - 1} \right) \right] \quad \text{Eq. 4.2}$$

4.3.2 – Study of contrast versus noise

For each independent realization, i , and reconstruction method, j , the activity $a_{ij,k}$ of the k lesion was defined as the average value of all image voxels within 10 mm spherical region-of-interest (ROI) centered on the true lesion location. The mean value $\langle a_{j,k} \rangle$ and the standard deviation $\sigma(a_{j,k})$ of the lesion activity was estimated across all realizations $\{a_{i,j,k} | i = 1, \dots, 25\}$. Noise was defined as $\sigma(a_{j,k})$ and the contrast, $C_{j,k}$, as $C_{j,k} = \frac{\langle a_{j,k} \rangle}{\langle \tilde{a}_{j,k} \rangle} - 1$, where $\langle \tilde{a}_{j,k} \rangle$ is the average across the 25 independent noisy realizations of the phantom without tumors.

4.3.3 – Methodology to fix the amount of post-smoothing and the number of iterations.

To select the appropriate amount of post-smoothing we studied the variation of the NPWMF index as a function of the FWHM of the post-smoothing filter, for all the OSEM approaches under study. This was done for tumors in the liver, lungs, and soft tissues, as well as for the two activity ratios. The optimum value for the FWHM was chosen such as to optimize the NPWMF indices.

For the number of iterations, we proceeded differently: we have, first of all, investigated how does the NPWMF index vary with the number of iterations. Since, as it will be shown latter on this chapter, a monotonic variation was found, i.e., there is a lack of an optimum value for the number of iterations, some other criterion had to be identified in order to define the number of iterations. This criterion involved taking as reference the tumor-to-background contrast level in reconstructed images for FORE+AWOSEM (since it is used in clinical practice) and find, for the other algorithms, the number of iterations resulting in images with an equivalent contrast for the corresponding organ. As it is possible to fit this criterion for a different number of FORE+AWOSEM iterations, we fixed in 4 the number of iterations for this reconstruction scheme, which is a number usually adopted in clinical routine, and used the aforementioned criterion to establish the number for the other algorithms. Whenever it was interesting to observe the results for another number of iterations, the same procedure was used, taking as reference 8 iterations of the FORE+AWOSEM.

4.4 – Results

A different number of results were obtained, concerning the different parameters under investigation. These results will be grouped and introduced in different sub-sections of the present section.

We will start by showing how the NPWMF index varies with the number of iterations, and use this knowledge to investigate the post-smoothing filter's FWHM that maximizes the NPWMF detectability index. These can be considered preliminary results upon which relies the most fundamental comparison between the performance of the different weighted OSEM versions. Such a comparison is the core of the present study, and detailed results will be present, in special those involving the NPWMF numerical observer. As complementary information, the most important results from the study of contrast versus noise are also described.

The discussion of all the results is postponed to section 4.5.

4.4.1 – Preliminary results for fixing the amount of post-smoothing and the number of iterations

The first parameter to be fixed was the number of iterations. As referred in section 4.3.3, we started by plotting the variation of the NPWMF index, averaged over all tumors in each region, as a function of the number of iterations, for the two activity ratios and using a post-smoothing filter with a FWHM equal to 5.5 mm. These plots are in Figure 4.6. The observed monotonic variation – lack of an optimum value - justifies the requirement of another criterion (as done).

Based in results reported later on this section, suggesting equivalent values for the tumor-to-background contrasts on the liver and lungs, on images after 4 iterations of FORE+AWOSEM, FORE+NECOSEM, ANWOSEM and NECOSEM3D, for studying the amount of post-smoothing we used this number of iterations for all of these strictly weighted OSEM versions. Due to the slower convergence of SPOSEM3D, as a consequence of the additive term on the equation (Eq. 3.2), we used, instead, 15 iterations of this algorithm. In Figure 4.7, the variation of the NPWMF index averaged over all the tumors in each organ (lungs, liver, and soft tissues), is plotted against the FWHM of the post-smoothing filter, for the two activity ratios, $A_{50\%}$ and A_{low} , after 4 iterations of OSEM (except for the case of SPOSEM3D where 15 iterations were used).

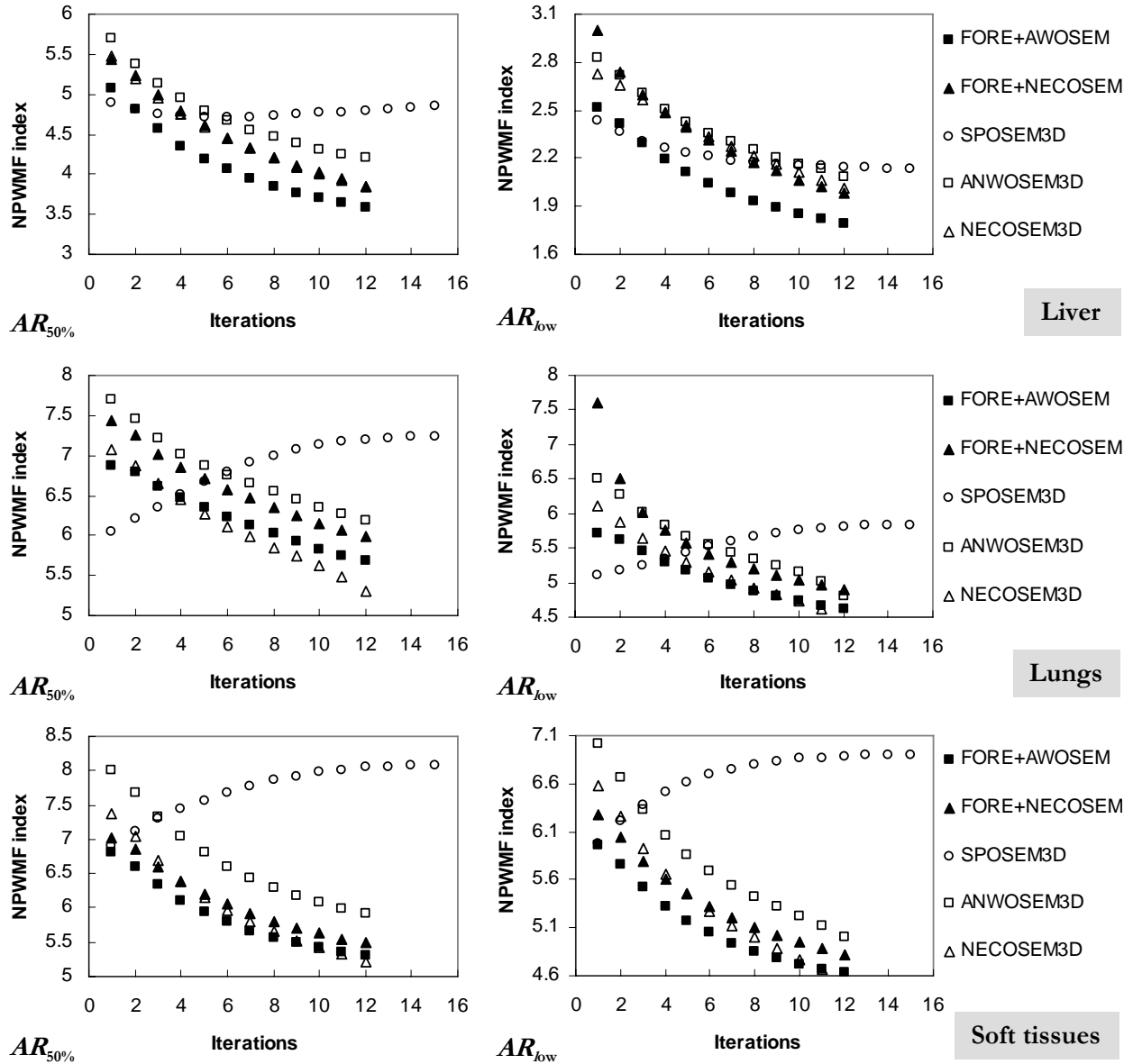


Figure 4.6 – Variation with number of iterations of the NPWMF index averaged over all tumors in the liver (top), lungs, and soft tissues (bottom row). Left column: $AR_{50\%}$; right column: AR_{low} . Post-reconstruction smoothing filter FWHM = 5.5 mm.

Observing the curves, we verify that there is no absolute optimum value for all the organs, activity ratios and OSEM versions. However, a FWHM equal to 5.5 mm corresponds either to the optimum value (in most of the cases), or is close to this value¹¹. Using images after

¹¹ Exceptions are the NECOSEM3D, whose maximum value is, in general, slightly higher, and the situation in the liver, for the SPOSEM3D (after 15 iterations), where it seems there is no optimum value.

8 iterations¹² (figures not shown), we found similar trends. So, henceforth this value will be used for the FWHM of the post-smoothing filter.

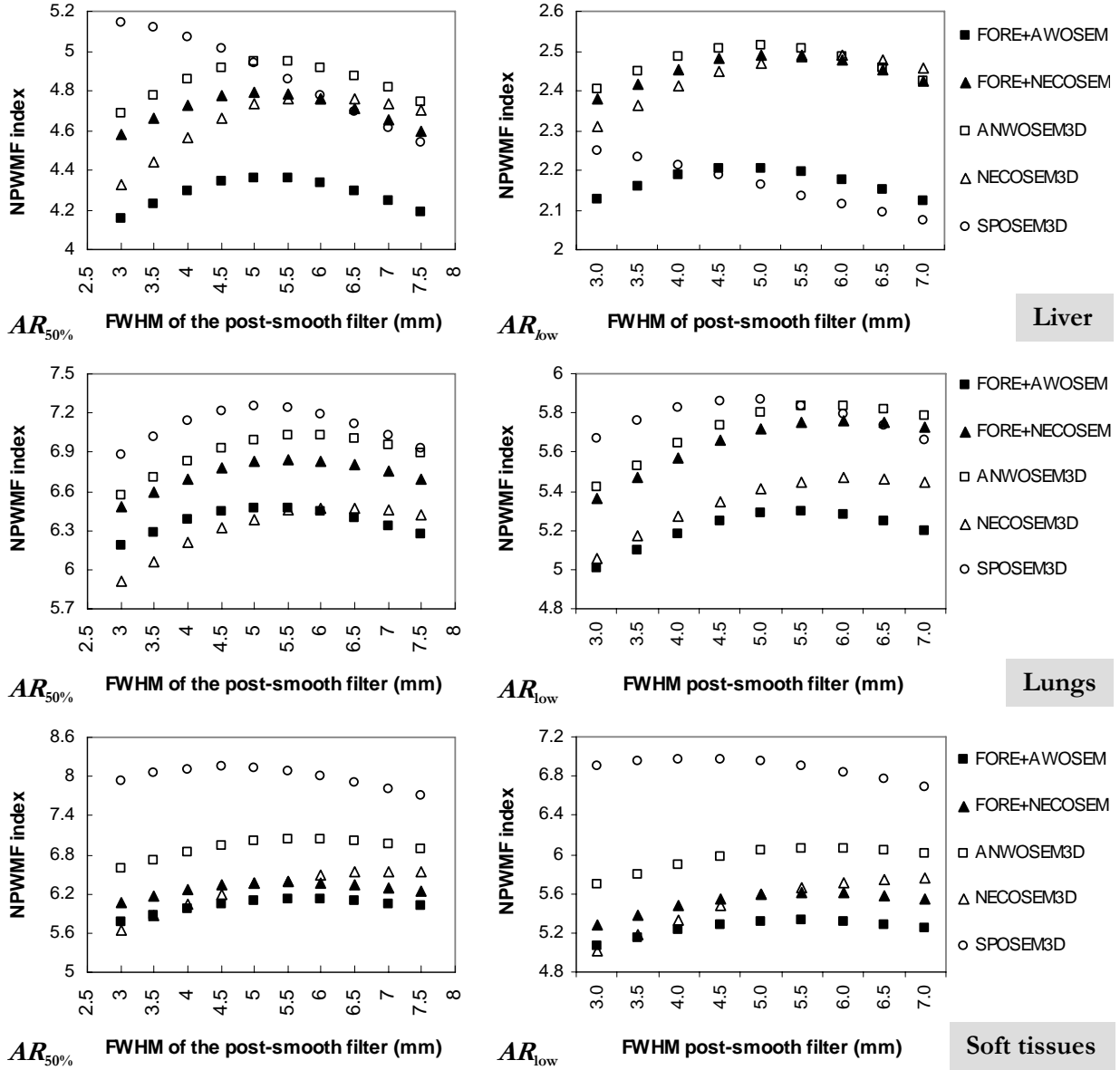


Figure 4.7 – Variation with the post-smoothing filter's FWHM of the NPWMF index averaged over all tumors in the liver (top), lungs, and soft tissues (bottom row), after 4 iterations (15 iterations of SPOSEM3D). Left column: $AR_{50\%}$; right column: AR_{low} .

¹² This refers to all the OSEM versions except SPOSEM3D.

In Figure 4.8 we show a coronal view of the reconstructed phantom with FORE+AWOSEM, FORE+NECOSEM, ANWOSEM3D, and NECOSEM3D after 4 iterations, as well as a reconstruction with the SPOSEM3D algorithm after 15 iterations (8 subsets in all cases). Images were post-smoothed with a gaussian filter (FWHM = 5.5 mm) and are shown for the activity ratio $A_{50\%}$.

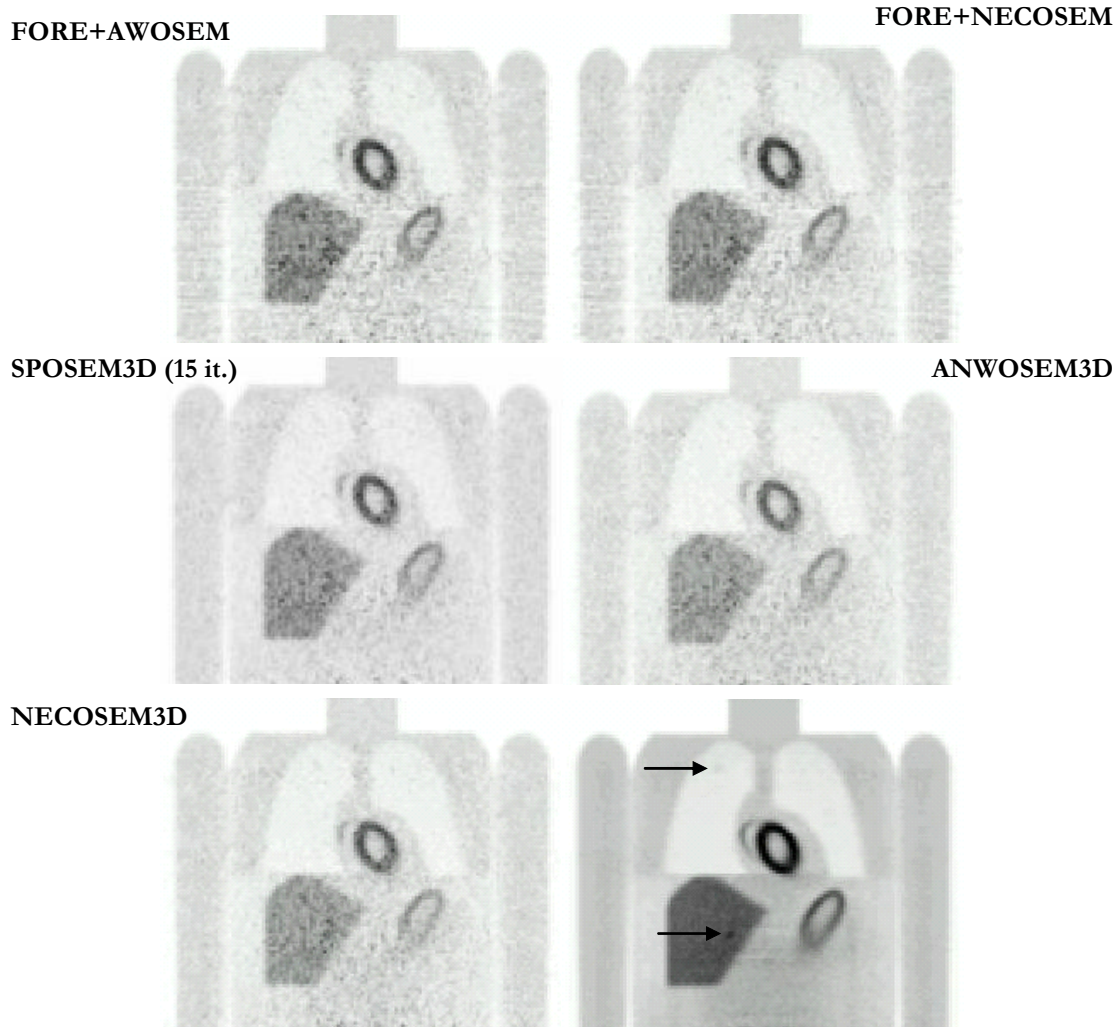


Figure 4.8 – Coronal view of the reconstructed phantom. All the images were obtained after 4 iterations, except for SPOSEM3D (15 iterations). Image in the bottom right is the average image taking 25 reconstructions with ANWOSEM3D after 4 iterations. Post-smoothing filter: FWHM = 5.5mm; activity ratio: $AR_{50\%}$. Arrows point to the tumors visible on the average image.

Deserving mention is the fact that, for a few realizations of the reconstructed images, typically one in each set of 25, as the number of NECOSEM3D iterations increases, the activity becomes concentrated in a few points on the border of the object – something that we can infer by comparison with a regular reconstructed image. Whenever this effect is verified, after 8 iterations, for example, the image contrast is virtually reduced to zero, because all the activity is distributed among a few points in the reconstructed volume. For the purpose of image quality assessment through the use of the NPWMF observer, these images were discarded.

4.4.2 – Study of contrast versus noise

The tumor-to-background contrast is used as a complementary figure-of-merit to compare the performance of the reconstruction algorithms as a function of the number of iterations. In Figure 4.9 are plotted the values for the contrast, averaged across all the tumors in the liver, lungs, and soft tissues, as a function of noise values (averaged in the same way), based on images for activity ratio $AR_{50\%}$ and using a post-smoothing filter with a FWHM of 5.5 mm. It should be stressed that since the noise increases with the number of iterations (a very well known result for this sort of reconstruction algorithms), plots in Fig. 4.9 can also be interpreted as showing the variation of the tumor-to-background contrast with the number of iterations. Therefore, in this figure, each noise value corresponds to a different iteration.

The higher number of iterations adopted for the SPOSEM3D (15 iterations) is based on the slower convergence of this algorithm, which is clearly shown in Fig. 4.9.

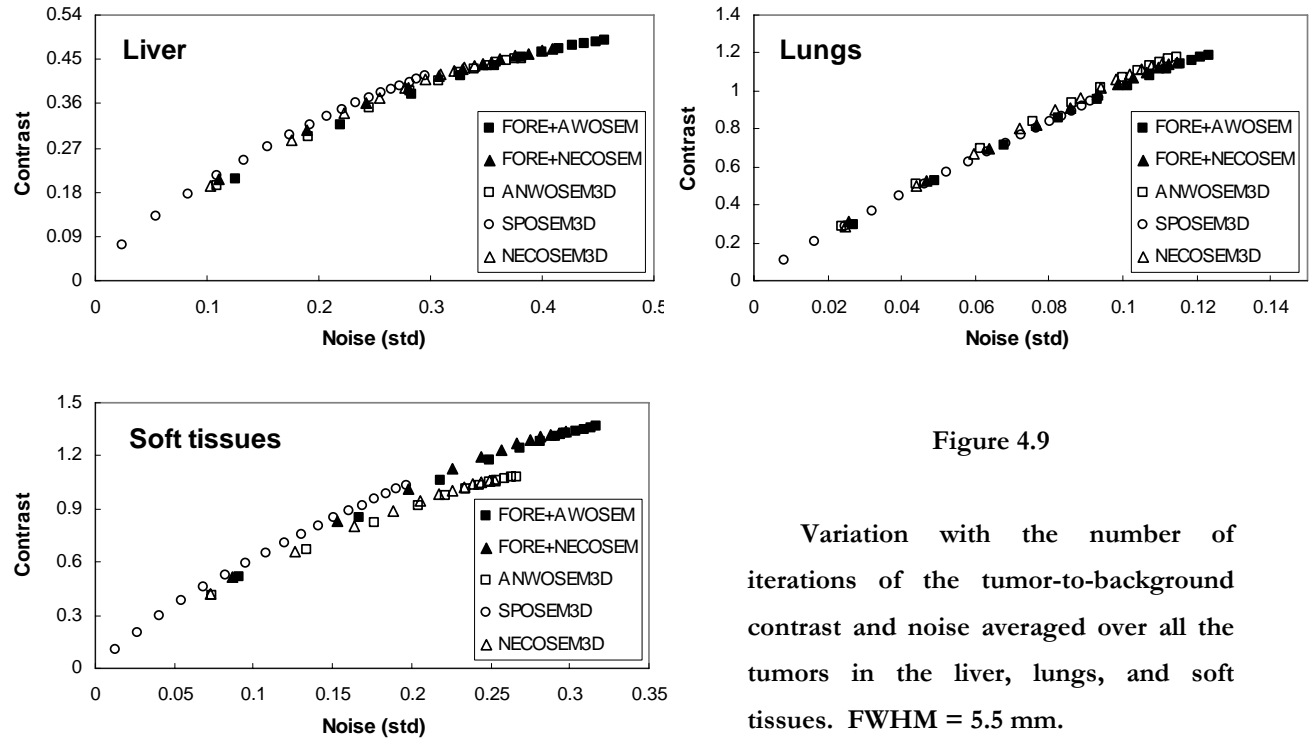


Figure 4.9

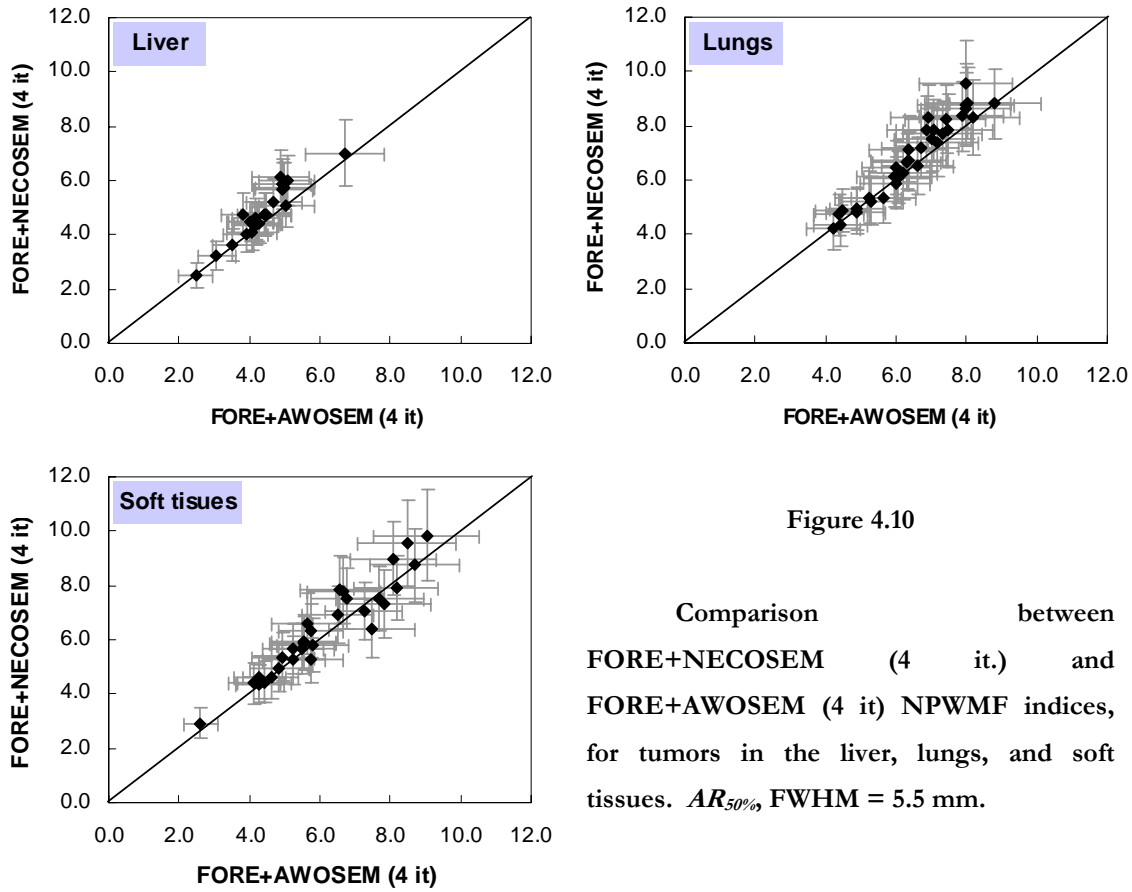
Variation with the number of iterations of the tumor-to-background contrast and noise averaged over all the tumors in the liver, lungs, and soft tissues. FWHM = 5.5 mm.

OSEM implementation		LIVER	LUNGS	SOFT TISSUES
FORE+AWOSEM (4 it)	Contrast	0.414	0.854	1.176
	d_{NPW}	4.357	6.471	6.113
	Noise	0.327	0.082	0.249
FORE+NECOSEM (4 it)	Contrast	0.393	0.821	1.125
	d_{NPW}	4.785	6.845	6.386
	Noise	0.280	0.076	0.226
ANWOSEM3D (4 it)	Contrast	0.385	0.836	0.919
	d_{NPW}	4.945	7.023	7.036
	Noise	0.282	0.076	0.204
SPOSEM3D (15 it)	Contrast	0.397	0.895	0.953
	d_{NPW}	4.855	7.237	8.077
	Noise	0.273	0.086	0.177
NECOSEM3D (4 it)	Contrast	0.372	0.801	0.890
	d_{NPW}	4.757	6.451	6.398
	Noise	0.255	0.072	0.188

Table 4.2 – Tumor-to-background detectability index, contrast and noise, averaged across all the tumors on each organ, for different reconstruction schemes and number of iterations. (FWHM = 5.5 mm).

4.4.3 – NPWMF observer study

On Table 4.2 we list the detectability indices, d_{NPWF} , for the five different weighted OSEM reconstructions. We also show the corresponding values for the tumor-to-background contrast (as defined in section 4.3.2) and noise. Values in the table are the average values across all the tumors in each organ, i.e., $\langle d_{NPWF} \rangle_{organ}$, $\langle C_j \rangle_{organ}$, and $\langle \sigma(a_j) \rangle_{organ}$ (the average noise). Images were obtained for the activity ratio $AR_{50\%}$ and smoothed using a gaussian filter with a FWHM equal to 5.5 mm.



We also investigated how the results obtained for two particular algorithms compare with each other, in one particular organ and after a convenient number of iterations (based on previous results from the study of contrast versus noise), taking all the tumors in that organ. All the scatter plots are for $AR_{50\%}$ and a post-smoothing filter's FWHM equal to 5.5 mm. Error bars

associated with the NPWMF index for each tumor were set considering the standard deviation derived from the error propagation formula, as described in section 4.3.1.

In Figure 4.10 we compare the indices resulting from FORE+NECOSEM with those obtained with FORE+AWOSEM, after 4 iterations, for all the tumors in the liver, lungs and soft tissues. Activity ratio in the phantom is $AR_{50\%}$ and images were subjected to a post-smoothing filter with FWHM equal to 5.5 mm.

Figure 4.11 is for the comparison between ANWOSEM3D (4 iterations) and FORE+AWOSEM (4 iterations). As in the former case, post-smoothing filter's FWHM equal to 5.5 mm, original activity ratio: $AR_{50\%}$.

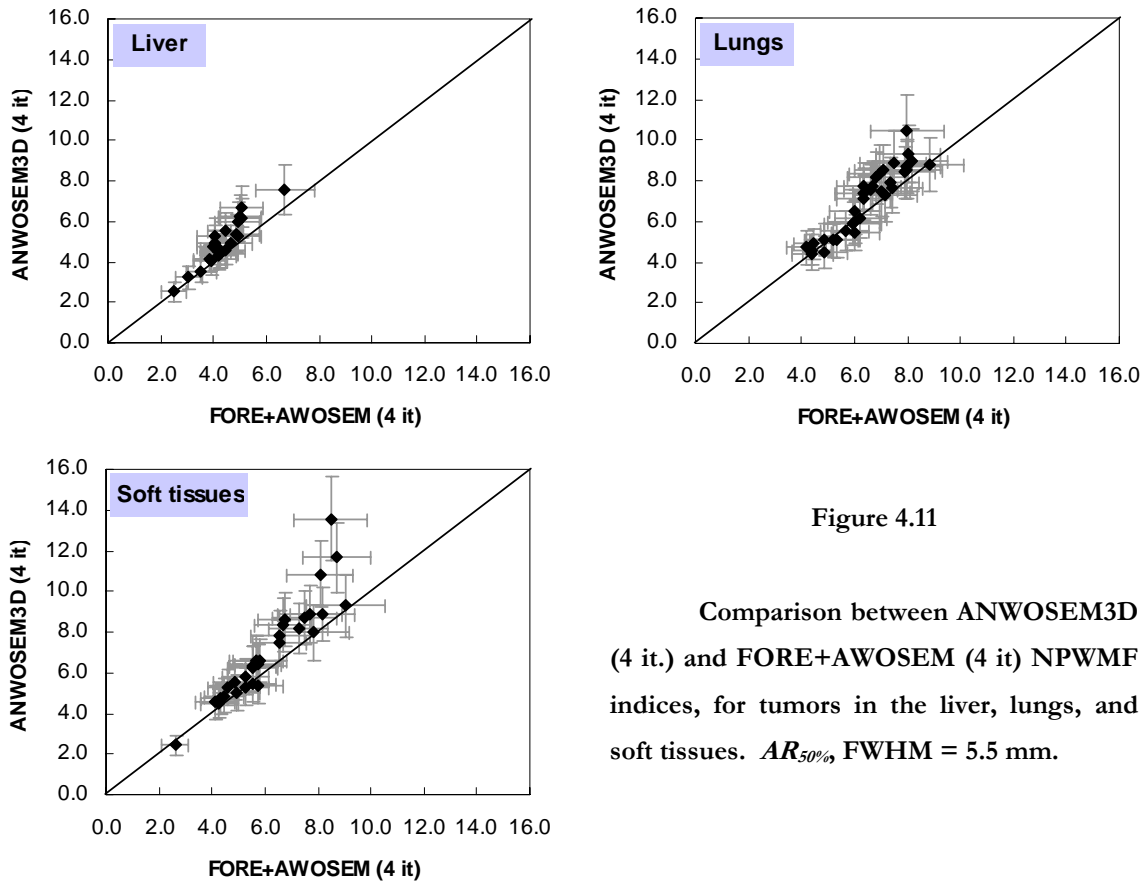


Figure 4.11

Comparison between ANWOSEM3D (4 it.) and FORE+AWOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm.

Figure 4.12 is for the comparison between ANWOSEM3D and FORE+NECOSEM after 4 iterations and Figure 4.13 for the comparison between NECOSEM3D and FORE+AWOSEM (both after 4 iterations). The activity ratio in the original phantom and the post-smoothing filter are the same as previously: $AR_{50\%}$, and FWHM = 5.5 mm.

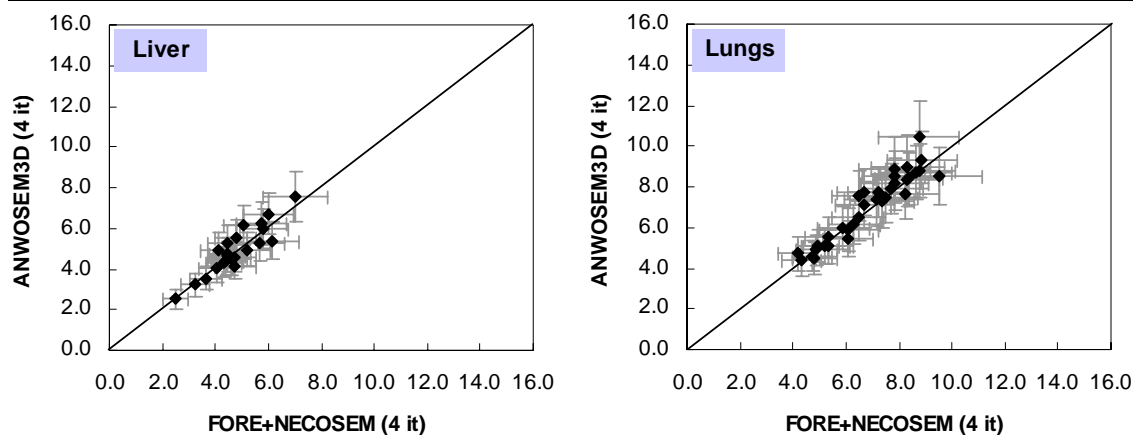


Figure 4.12

Comparison between ANWOSEM3D (4 it.) and FORE+NECOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$ FWHM = 5.5 mm.

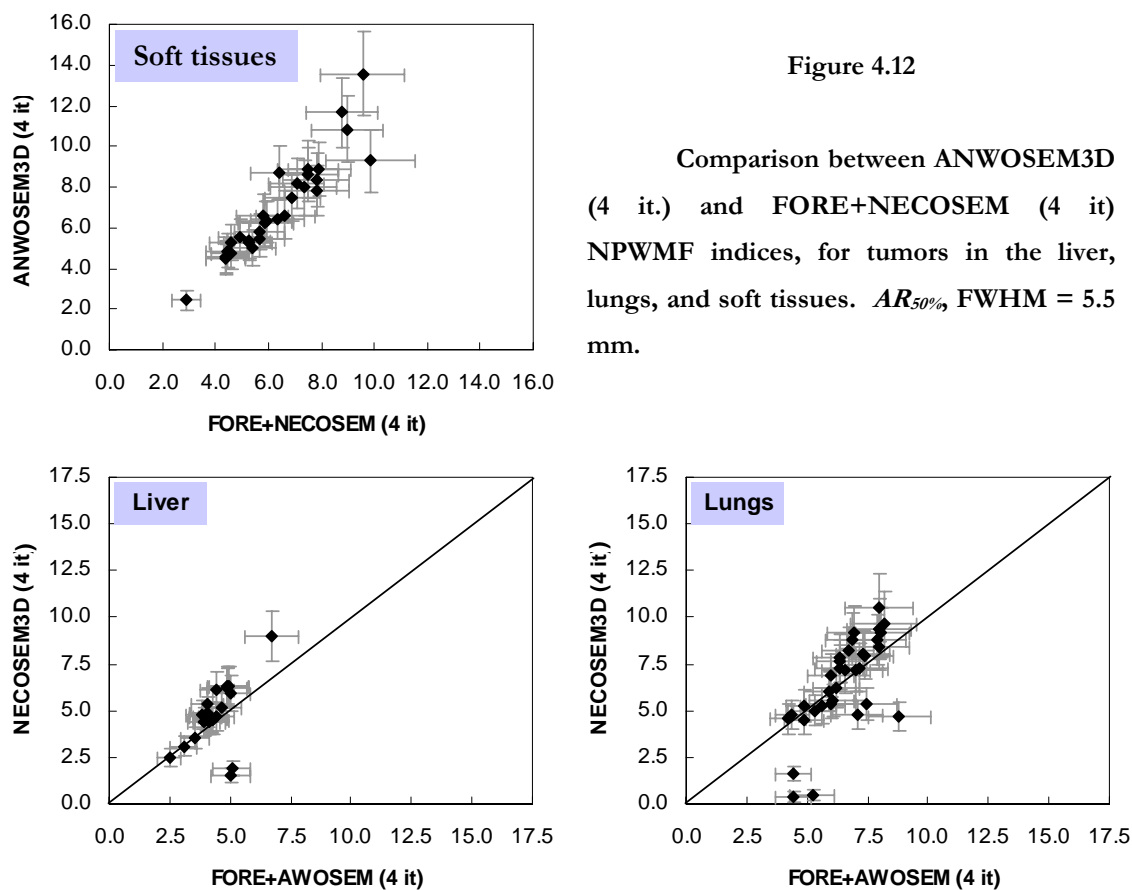


Figure 4.13

Comparison between NECOSEM3D (4 it.) and FORE+NECOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$ FWHM = 5.5 mm.

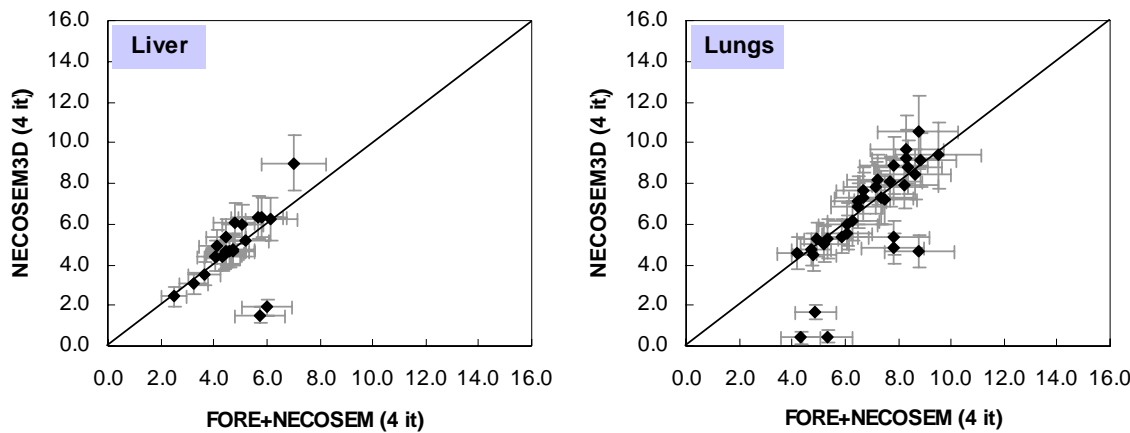


Figure 4.14

Comparison between NECOSEM3D (4 it.) and FORE+NECOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm.

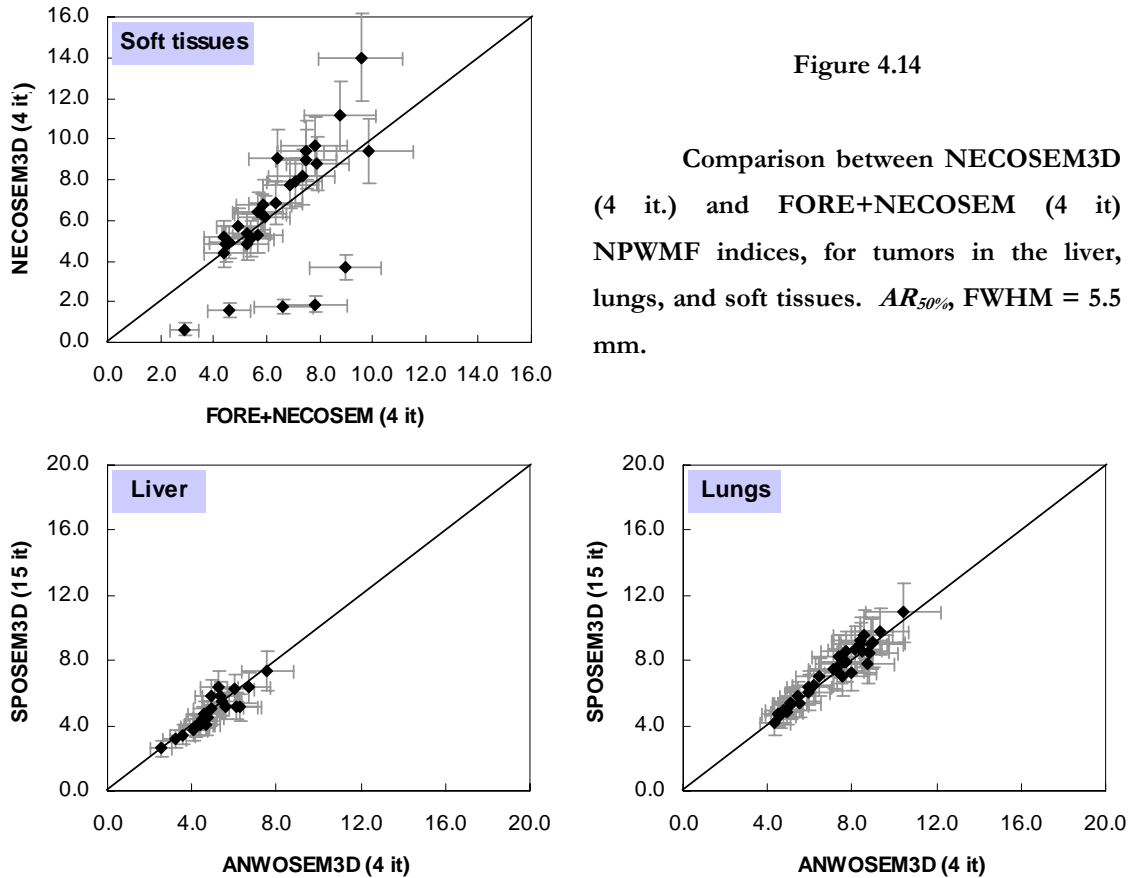


Figure 4.15

Comparison between SPOSEM3D (15 it.) and ANWOSEM3D (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm.

Finally, in Fig. 4.14 are compared NECOSEM3D and FORE+NECOSEM (once again after 4 iterations of these algorithms), and in Fig. 4.15 we compare two fully 3D OSEM algorithms: SPOSEM3D (15 iterations) and ANWOSEM3D (4 iterations). The activity ratio in the original phantom and the post-smoothing filter are still the same: $AR_{50\%}$ and FWHM = 5.5 mm.

4.5 – Discussion and conclusions

During this study it was possible to find a value for the FWHM of the post-smoothing filter, which, applied to the reconstructed images, results in a maximum value for the NPWMF detectability index. This is the case for all the studied OSEM versions. This value is not absolute, i.e., it is not exactly the same for all the activity ratios and organs, but a FWHM of 5.5 mm, for 1 cm diameter spherical tumors, is the optimum value in most of the cases or is close to this optimum. Also relevant is the fact that such an optimum value is not dependent from the number of iterations used.

On the other hand, the NPWMF detectability index can not be used to define the optimum number of iterations, since it was verified that this index is monotonically decreasing with the number of iterations. Due to this, we adopted the tumor-to-background contrast level as a figure-of-merit, to define the number of iterations.

Considering the variation of the contrast with the number of iterations (Fig. 4.9), we observe the slower convergence of SPOSEM3D compared to the other studied implementations. This result is not surprising since there is an additive term that is responsible for the effect. The fact that three times more iterations (15 instead of 4) were needed to reach similar tumor-to-background contrast, is dependent on the ratio between scattered plus random coincidences and unscattered true coincidences. Thus, this ratio between the number of iterations is valid just for this particular study.

For the other reconstruction algorithms studied, there are no significant variations in the way the contrast depends on the number of iterations, except for the case of FORE+AWOSEM where the noise seems to increase faster with the number of iterations.

From the direct comparison of the NPWMF indices for different tumors in each region, it is clear that FORE+NECOSEM has an overall better performance than FORE+AWOSEM. The same conclusion could be drawn observing the figures where ANWOSEM3D is plotted

against FORE+AWOSEM: the former results in higher detectability indices than FORE+AWOSEM. The difference between ANWOSEM3D and FORE + NECOSEM is less evident than in the previous case. Similar results are found when comparing SPOSEM3D with FORE+AWOSEM and FORE+NECOSEM.

The analysis of Figure 4.13 suggests less obvious differences between NECOSEM3D and FORE+NECOSEM. However, from the observation of a similar plot where NECOSEM3D is compared with FORE+AWOSEM (Fig. 4.13), it is possible to argue for a better performance of the former over the last reconstruction algorithm. In relation to the other 3D algorithms, no relevant differences were found, except is the soft tissues, where SPOSEM3D, after 15 iterations, performs better than ANWOSEM (4 iteration).

Thus, it seems that the additional complexity associated with NECOSEM3D has no direct positive impact on the detectability of tumors on images reconstructed using this version of the weighted OSEM algorithm.

The behavior of the fully 3D NECOSEM reconstruction depends very much on the profile of the NEC weights, due to rounding and other numerical artifacts that could result from the processing of these weights. We observe that sharp variations on the profile of the NEC weight lead to a reconstruction very sensitive to the noise on the sinograms. In some cases, the algorithm proceeds towards reconstructed images that look similar to images obtained with the other algorithms, in terms of visible structures and contrast. For other cases, however, some points drain all the activity during the reconstruction process: the final image after a few iterations is reduced to zero everywhere except in these points. The referred strategy of constraining the NEC weight to be equal to one for sinogram bins inside the object but close to its border (as well as outside the object), is an attempt to reduce sharp variations on the NEC weight profile. The use of the attenuation map to achieve this goal proved useful but didn't resolve the problem: we still had to eliminate, on average, one noisy realization per each set of 25. The threshold on the attenuation is a way to define how deep inside the object we are constraining the use of the expression established in section 3.1.1 to calculate the 3D NEC weight. Since the attenuation map is usually accessible, in theory it should be possible to tune the threshold so that reconstruction proceeds correctly. This is, nevertheless, a rather empirical choice.

We finally note that the trends observed on the results we have been describing are, in general, very similar for the two tumor-to-background activity ratios: $AR_{50\%}$ and AR_{low} .

In the next chapter we will describe the use of some of these reconstruction algorithms with real clinical data, in a context where, among other constraints, there is no knowledge about

the variance or the mean. Such fact has strong implications in the implementation of NECOSEM and a strategy must be devised to overcome this lack of information.

5 – Comparing NEC-based OSEM with other weighted OSEM approaches for clinical 3D whole-body PET imaging

In this chapter we compare the performance of different weighted OSEM approaches for the reconstruction of 3D clinical whole-body PET data. Among these approaches, we aim to verify if the use of an accurate statistical model for OSEM, i.e., if the use of a NEC-based OSEM, is affordable with clinical data and if it produces images with an improved quality (again, for the clinical task of tumor detection) when compared to the clinical widespread FORE+AWOSEM reconstruction algorithm.

The results described in the former chapter for simulated data show improvements in tumour detectability when a realistic statistical model for the data is used. This is the case of FORE+NECOSEM when compared with FORE+AWOSEM. However, these results are for simulated data, which means that relevant information concerning data is available and can be used directly in the reconstruction stage. In particular, as referred previously, this is critical for the use of NECOSEM, since the NEC weighting factor, w_{NEC} is defined as the ratio between the mean and the variance (section 3.1.1):

$$w_{NEC} = \frac{\langle y_i \rangle}{\sigma^2(y_i)} \quad \text{Eq. 5.1}$$

Simulated data provide the ideal conditions from this point of view since the NEC weighting factors can be determined exactly.

This is not the case for clinical data, for which there is no other value associated with a particular bin than its own value, i.e., there is no information about the mean and the variance associated with each data bin. Therefore, in order to use a NEC based reconstruction, it is necessary, first, to estimate these moments for each data bin. However, even if this is possible, these will be just estimated and not exact values, which means that they contain an error factor. How far this error translates into the NECOSEM performance is part of the research work this

chapter is devoted to: FORE+NECOSEM is compared not only to FORE+AWOSEM but also to fully 3D weighted OSEM approaches.

Another important point regarding clinical data is to verify if the use of an accurate statistical model is advantageous for over weighted patients, whose resulting whole-body PET images are very noisy.

In the following sections we will be describing how data was acquired and post-processed (corrected for some effects) before matching the requirements for being given as input data to the reconstruction algorithms. The estimation of the weighting factors, in particular for NEC, is also detailed. The chapter proceeds with a description of the reconstruction stage and, finally, the most important results are presented and discussed.

ID	Weight (kg)	Height (cm)	Dose (MBq)	Age (y)	Sex	BMI
1	52	179	384.8	56	M	16.2
2	45	164	370	76	F	16.7
3	48	167	370	42	F	17.2
4	65	178	188	71	F	20.5
5	65	174	351.5	52	M	21.5
6	63	170	370	80	M	21.8
7	59	158	384.8	64	F	23.6
8	75	178	338.5	44	M	23.7
9	75	178	381.1	44	M	23.7
10	90	186	547.6	44	M	26.0
11	86	180	499.5	44	M	26.5
12	83	174	484.7	75	M	27.4
13	90	181	444	45	M	27.5
14	67	156	414.4	74	F	27.5
15	90	178	370	70	M	28.4
16	94	180	407	71	M	29.0
17	79	160	407	71	M	30.9
18	96	171	?	41	F	32.8
19	92	164	429.2	54	F	34.2
20	87	158	396	73	F	34.9
21	85	152	370	73	F	36.8
22	95	160	370	76	F	37.1
23	115	175	555	51	M	37.6
24	110	168	403.3	72	M	39.0
25	120	164	414	64	F	44.6

Table 5.1 – A description of the 25 scans used in the present study, ordered according to their body mass index. Records for patient #26 were not available. (BMI = Weight(kg) / Height(m)²)

5.1 – Data acquisition and description

Clinical data were acquired in 3D mode using the Siemens/CTI HR+ scanner and corrected for attenuation, randoms and scattered photons, as well as for geometric effects (normalization). Data handling and processing will be detailed in the following section.

All the scans were obtained at the Service Hospitalier Frédéric Joliot, in accordance with the clinical protocols established at the institution. It should be noted, however, that the author was not present during the scans, which are performed, under the supervision of the medical staff, by a nurse or a technologist. Data were, therefore, chosen *a posteriori*.

This choice was made among the exams available, considering 25 patients with different body mass indices (BMI). Due to the interest in comparing reconstructed images for overweighted patients, most of the exams correspond to patients with large BMIs. In Table 5.1 we list some useful data regarding each exam.

5.2 – Data pre-processing

The designation of data pre-processing is meant to distinguish a set of operations performed over the data using proprietary routines available only in the scanner's dedicated computer, from the other conventional processing that can be done elsewhere (as long as output files from proprietary pre-processing become available and usable as input files at this stage) and is similar to the processing chain simulated data are also subjected to¹.

This pre-processing chain, which is graphically represented in Figure 5.1, is necessary due to some specific characteristics of the clinical data acquired with the HR+ scanner operating in regular conditions. In fact, in order to match the requirements of the subsequent stages, i.e., in order to use these data as input for the routines that were previously adopted for correcting simulated data for attenuation, normalization, and other operations (c.f. Fig. 4.2), clinical data must be first properly worked.

First of all, at the very beginning of the chain are the three files that can be downloaded from the server where all the clinical exams are stored. These files are:

¹ With the very important exception of the operations that are required for estimating the NEC weighting factors.

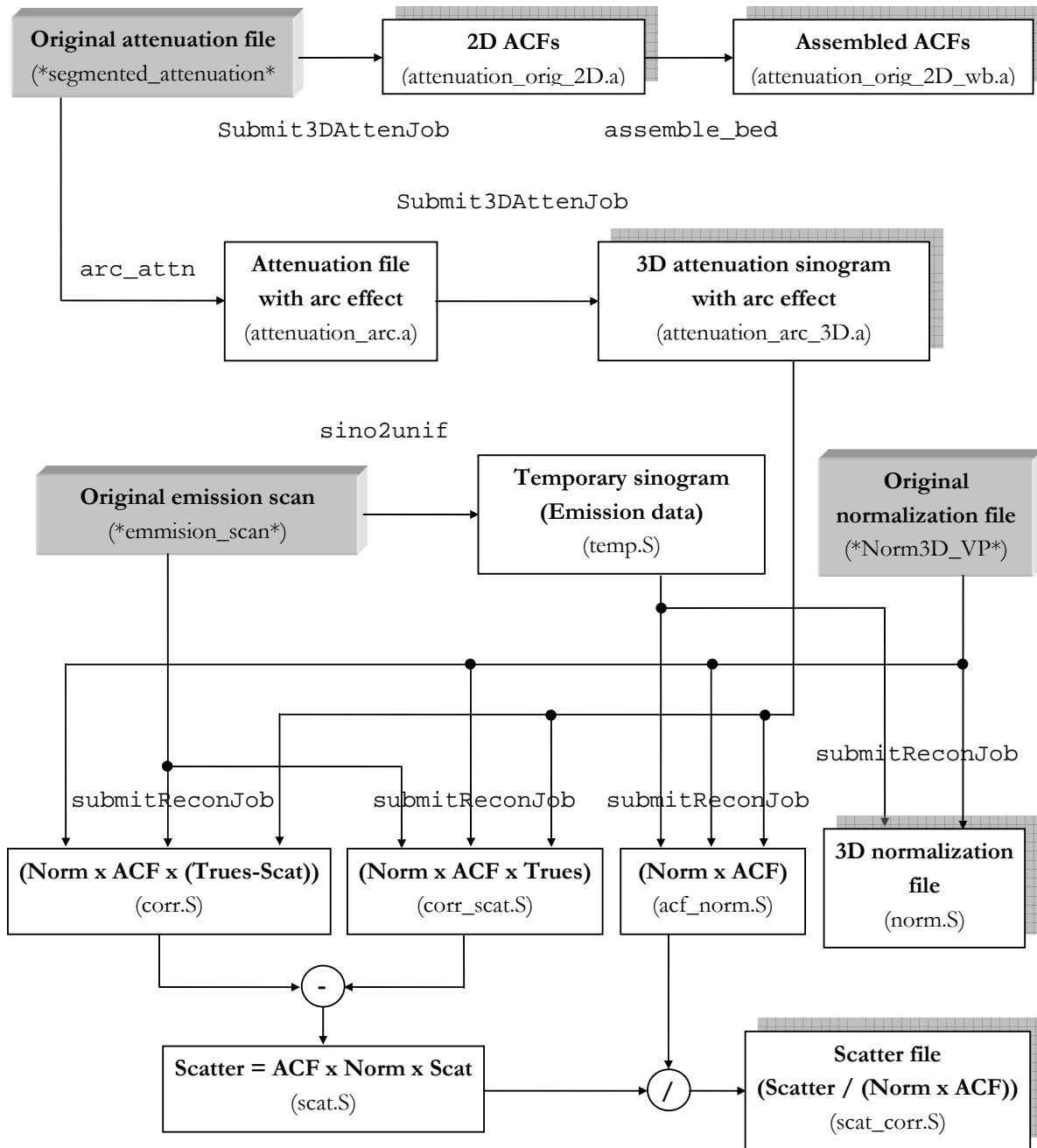


Figure 5.1 – A flowchart of the data pre-processing stage that uses dedicated routines available only in the scanner’s proprietary computer. Outside the boxes, but close to them, is the name of the routine involved in each process.

- the emission scan (*emmission_scan*): a file with the trues;
- the original attenuation file (*segmented_attenuation*), containing the ACFs in 2D corrected for the arc effect;

- the original normalization file (*Norm3D_VP*), not in sinogram format, accounting for geometric effects and a calibration factor.

As soon as these files become available, before using them outside the scanner’s proprietary computer, for ethical reasons the patient name is removed and the data made anonymous. This procedure was strictly followed.

Concerning the original attenuation file, the most important feature to be considered is the fact that it is already corrected for the arc effect (something that is automatically done by the proprietary routines used for its generation). However, as for the case of simulated data, raw data is normalized and corrected for the arc effect latter, by using the `normalize` routine. So, since the attenuation data is also input for that routine, it must not be corrected for the arc effect.

To match this requirement of the main normalization stage, the original attenuation file must be previously de-corrected for the arc effect, something that is done at the beginning of the pre-processing chain (as described in Fig. 5.1) by using the `arc_atten`² routine and results in the file designated as “attenuation_arc.a”. In addition to this important task, another routine – `submit3DAttenJob` – is used to compute 3D ACFs from 2D ACFs, and also for granting data the correct format (in accordance with the data organization requirements of the subsequent processing routines).

The scattered photons component – which is necessary for the calculation of the NEC weighting factors – is not directly available from the scanner, once more because of the characteristics of the proprietary software. However, it can be estimated indirectly from two different files that can be generated from the original emission sinogram, using the proper corrections.

As illustrated in Fig. 5.1, using as input the original emission sinogram, the 3D attenuation sinogram with arc effect, and the original normalization file, the `submitReconJob` routine can perform different corrections over the emission data and, therefore, generate different files. In particular, by properly setting the routine parameters’, one can order the following files:

- $corr_scat_i = N_i A_i T_i$, where N_i is for the normalization factors, A_i for the attenuation correction factors and T_i for the trues;
- $corr_i = N_i A_i (T_i - s_i)$, where s_i is for the scattered.

² This routine was written by Charles Pautrot [Pautrot, 03] as part of his research work developed in the S. H. F. J..

Once we have these files, their difference results in the following file:

$$corr_i - corr_scat_i = N_i \cdot A_i s_i$$

A final division by $N_i \cdot A_i$ results in the file `scat_corr.S`, that contains the data scattered component.

Strictly for 2D reconstruction purposes, and in spite of the routine name, a parallel pre-processing path is taken to convert the original 2D attenuation file's structure according to the latter processing requirements, keeping its 2D nature. In case of multi-bed acquisitions, an assemble bed operation is also performed.

5.3 – Data corrections and the estimation of data mean and variance

Given the goal of this study, the most critical part of the work is the estimation of the NEC weighting factors. As referred at the beginning of this chapter, the use of Eq. 5.1 requires knowledge about data mean and variance, and, since for clinical data this sort of information is not available (in strong contrast with simulated data), it must be estimated.

The trues, t_i^{3D} ($t_i^{3D} = (p_i^{3D} - r_i)$), are the raw clinical data. Correcting these data for attenuation and geometrical effects – normalization – can be mathematically expressed as follows:

$$y_i^{3D} = N_i \cdot A_i (t_i^{3D} - s_i) \tag{Eq. 5.2}$$

Therefore, the mean value and the variance of the 3D clinical data after corrections are given by Eq.5.3 and 5.4, respectively.

$$\langle y_i^{3D} \rangle = N_i \cdot A_i (\langle t_i^{3D} \rangle - s_i) \tag{Eq. 5.3}$$

$$\sigma^2(y_i^{3D}) = N_i^2 \cdot A_i^2 (\langle p_i^{3D} \rangle + \langle r_i \rangle) = N_i^2 \cdot A_i^2 (\langle t_i^{3D} \rangle + 2\langle r_i \rangle) \tag{Eq. 5.4}$$

5.3.1 – The estimation of the mean

From Eq. 5.3 it results the need to estimate the mean value for the trues, t_i^{3D} , in order to obtain the mean value of the corrected data. Since there is a single value per bin (the number of counts for the corresponding LOR) and there is no possibility to replicate the number of realizations of the exam, the original raw data were smoothed and the value of each bin after the smoothing operation assumed as an estimator of its own mean value.

The smoothing was done by using the Fast Fourier Transform (FFT) and a Gaussian filter in the Fourier space, using a routine designed by C. Pautrot, who has verified in his research work [Pautrot, 03] that such a method represents a good trade-off between the desired result and speed in the computational execution of the task. The Gaussian transaxial FWHM was set equal to 2.20 mm and the axial FWHM equal to 2.13 mm [Pautrot, 03].

Before smoothing, however, the trues must be corrected for geometrical effects and detector efficiency, i.e., must be normalized. Once they are normalized, they are smoothed as described above. Because the original file with the raw data will be latter submitted again to the normalization routine (a stage where not only the normalization but also the attenuation are considered and corrected for), the smoothed trues must, therefore, be de-corrected for the normalization.

Finally, for a question of convenience, the sinogram with the smoothed trues is concatenated to the original sinogram (with the trues), by maintaining the last in gate 1 and storing the latter in gate 3.

In order to validate de procedure for estimating a mean value for the trues, we used simulated data and compared, for a noisy sinogram obtained from the anthropomorphic phantom, the known mean values (stored in gate 3) with those values resulting from the smoothing of the simulated noisy values (gate 1)³.

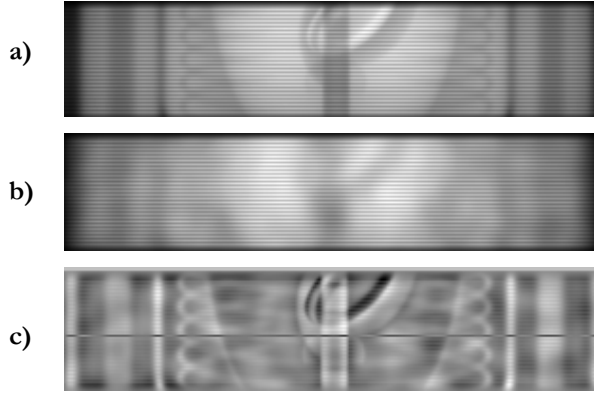
Therefore, we proceeded as follows: we took the noiseless emission scan (Fig. 4.2) and use the `noise` routine to generate the emission scan with noise (`*_nsy.S`). However, for the present study, we didn't correct for scattered photons, in order to replicate the conditions of clinical data. The noisy sinogram was stored in gate 1, the data variance in gate 2, and the mean value of the trues, in gate 3. This last sinogram (mean value in gate3), $\langle t_i^{3D} \rangle$, was used, as referred, as the reference for validating the method under study. The smoothed sinogram,

³ For noisy sinograms, data are stored in Gate 1, the associated variance in Gate 2, and the associated expectation in Gate 3; the latter two if option `-V` is used in the routine `noise`. In addition, the prompts or the trues are stored in `Data=0`, the randoms or delayed in `data=1`, and the scatters in `data=4` (cf. Appendix A.8).

$sm(t_i^{3D})$, was obtained according to the following equation, where $sm()$ is the Gaussian smoothing operation performed in the Fourier space and N_i the normalization factors:

$$sm(t_i^{3D}) = \frac{1}{N_i} sm(N_i \cdot t_i^{3D})$$

In Figure 5.2 a) we show one projection corresponding to $\langle t_i^{3D} \rangle$ (gate 3), for one bed position, segment 0 and azimuthal angle equal to 0° . In Figure 5.2 b), for the same bed position, segment and azimuthal angle, is shown the projection for the smoothed trues, $sm(t_i^{3D})$. Finally, in Fig. 5.2 c) is the image with the difference between the two projections: $\langle t_i^{3D} \rangle - sm(t_i^{3D})$.



All the images were obtained before normalization, which justifies the observed structure along the axial direction.

In Figure 5.3 is shown a radial profile drawn from the previous projections: $\langle t_i^{3D} \rangle$ and $sm(t_i^{3D})$. An addition profile for the simulated trues (gate 1), t_i^{3D} , is also presented. The line in Figure 5.2 c) indicates the axial position corresponding to the profiles.

As it can be observed from Figure 5.3, the differences found between the radial profiles for the smoothed trues and the trues' mean values are not significant. The trend observed for that particular profile is verified (not shown) for all the profiles obtained for different axial and bed positions. This leads us to conclude that smoothing the trues is a valid approach for obtaining an estimation of the trues' mean values, validating, therefore, the procedure to be used for clinical data.

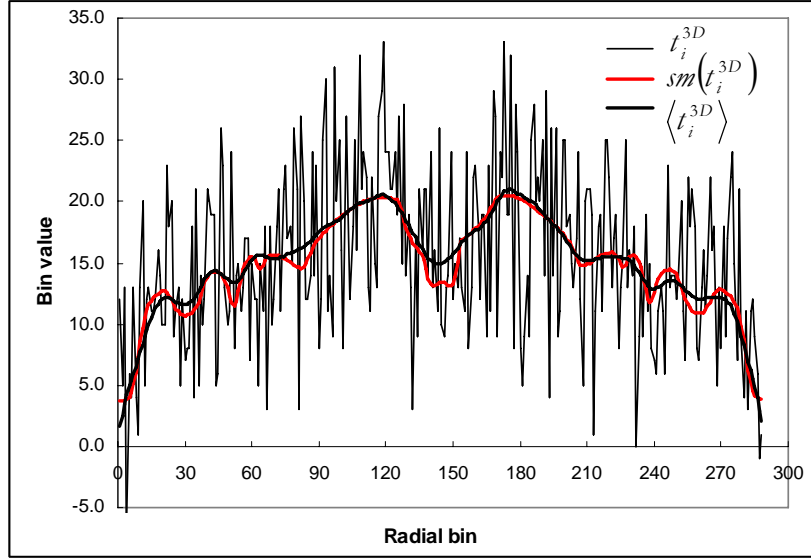


Figure 5.3 – A radial profile of the trues (gate 1), t_i^{3D} , the trues’ mean values (gate 3), $\langle t_i^{3D} \rangle$, and the smoothed trues, $sm(t_i^{3D})$. Profiles drawn from projections shown in Figure 5.2.

5.3.2 – The estimation of the variance

The estimation of the variance is done according to Eq. 5.4. As for the estimation of the mean value, we used simulated data to validate the approach described by Eq. 5.4 and adopted for clinical data.

We started by adding noise (noise routine) to a simulated emission sinogram previously obtained based on the anthropomorphic phantom. In this operation, we used the ASIM simulator’s options –V (to store the variance data on gate 2 and the mean value on gate 3) and –R (for simulating the trues and not the prompts, i.e., not correcting for scattered photons (Fig. 5.4). This is the sinogram designated by “emission scan with noise 1” in Fig. 5.4. Then, this sinogram was corrected for scattered photons, attenuation and normalization, in the so-called normalization step (using the normalize routine). The obtained sinogram contains in gate 2 the variance data based on which noise add been added to the sinogram, i.e., the exact variance data values.

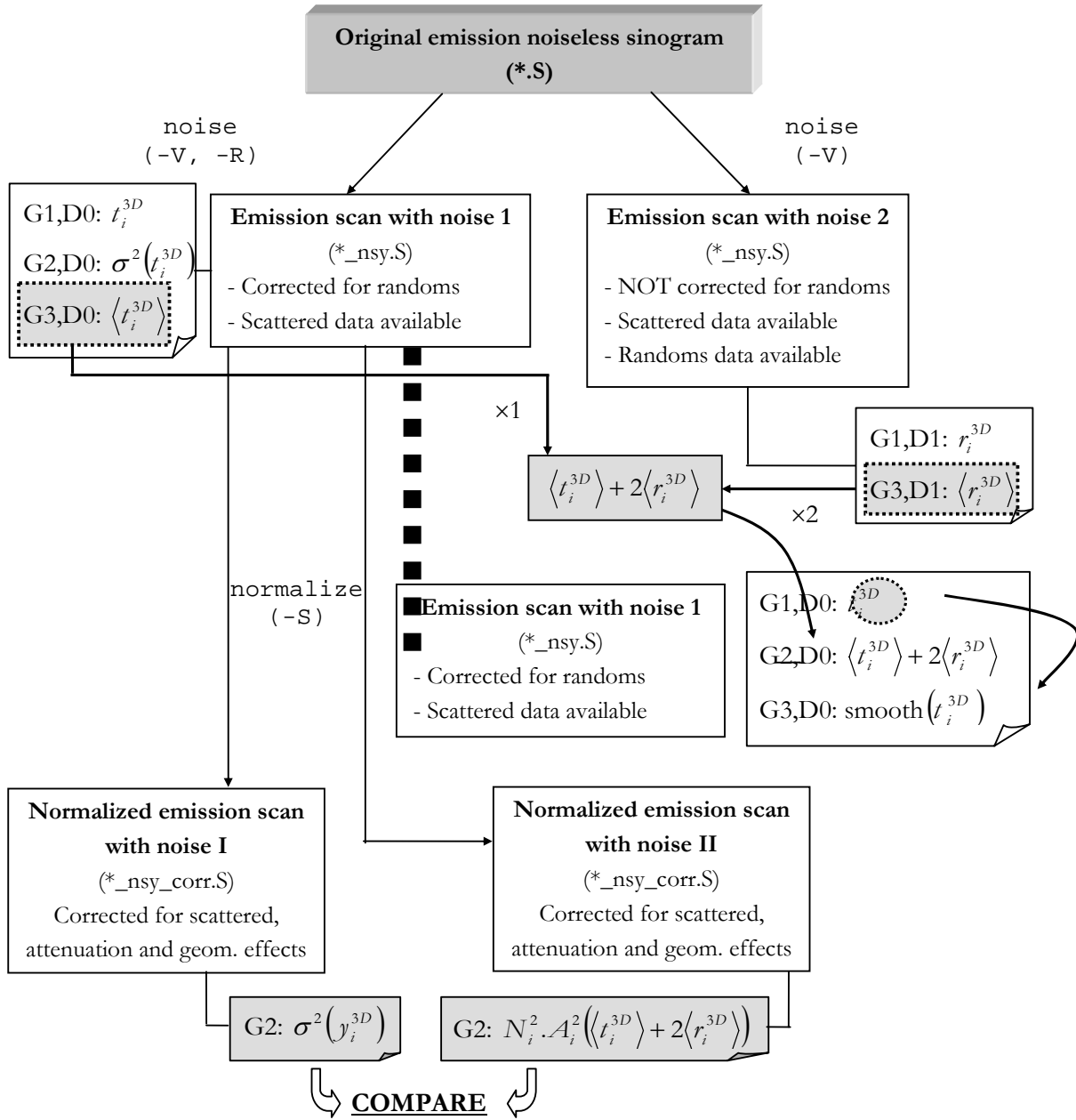


Figure 5.4 – A flowchart of the processing chain simulated data were subjected to in order to compare the exact data variance with the variance estimation process used for clinical data.

In parallel, for simulating the processing chain clinical data are subjected to, a second noisy emission sinogram was simulated, “emission scan with noise 2”. However, in contrast with the former noisy sinogram, in this stage this sinogram was not corrected for randoms. The idea was to use the simulator’s possibility of storing the randoms in a dedicated section of the

sinogram file⁴, corresponding to data category 1 (D1). Since the simulator's option $-V$ was set, in gate 3 are stored data mean values. So, in the sinogram section labeled D1G3 are stored the mean values for the randoms, $\langle r_i \rangle$. The mean values for the trues, $\langle t_i^{3D} \rangle$, were determined as described in section 5.3.1: smoothing the trues. Once these data were available, the $\langle t_i^{3D} \rangle + 2\langle r_i \rangle$ sinogram was generated.

Finally, in one copy on the first noisy sinogram - “emission scan with noise 1” - values in gate 2 were replaced by the $\langle t_i^{3D} \rangle + 2\langle r_i \rangle$ sinogram, and gate 3 was overwritten by the smoothed trues sinogram. After this replacement, the sinogram was corrected for scattered, attenuation and geometrical effects.

For validating the variance estimation procedure used with clinical data, two comparisons were made between the exact variance in gate 2 and the estimated variance using Eq. 5.4: before and after the normalization procedure.

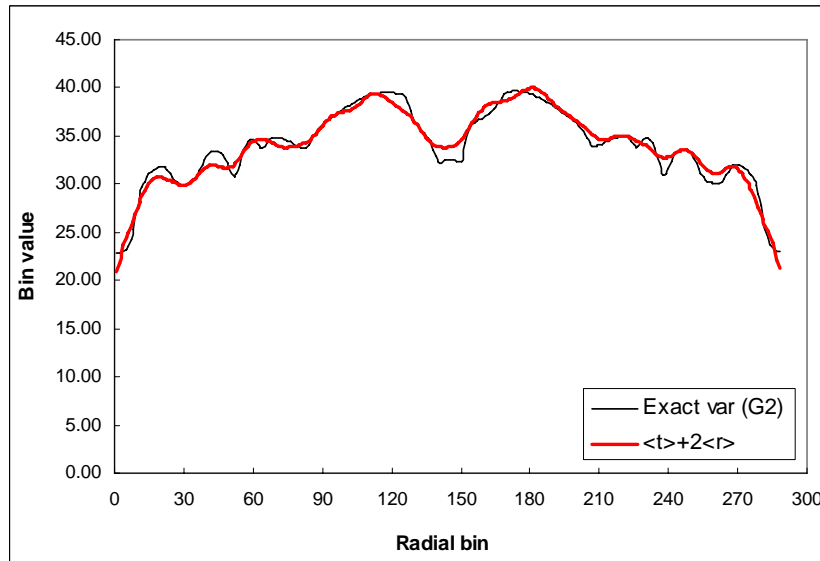


Figure 5.5 – A radial profile of the exact variance (gate 2) and $\langle t_i^{3D} \rangle + 2\langle r_i^{3D} \rangle$. Profiles drawn from projections at the chest level for segment 0 and $\phi = 0^\circ$.

⁴ As referred in Appendix A8, in the ECAT sinogram structure data are labelled according to 5 different parameters: *frame*, *plane*, *gate*, *data* and *bed*. The *data* label is used to distinguish between the trues (or prompts) (D0), the randoms (D1) and the scattered (D4).

In Fig. 5.5, the comparison is between the exact variance before normalization and $\langle t_i^{3D} \rangle + 2\langle r_i^{3D} \rangle$, i.e., the corresponding variance estimation. Among the four bed positions simulated, it was chosen one at the chest level. The projection corresponding to segment 0 and $\phi = 0^\circ$ was considered. The radial profile is for axial position 32 (among the 63 available).

The second comparison (Fig. 5.6) is between the exact variance (G2) after normalization and $N_i^2 \cdot \mathcal{A}_i^2 \left(\langle t_i^{3D} \rangle + 2\langle r_i^{3D} \rangle \right)$, as indicated at the bottom of Fig. 5.3. Once again, the projections were obtained at the chest level of the anthropomorphic phantom, for segment 0 and $\phi = 0^\circ$.

In both cases we can observe that the two profiles are very similar. This was also verified for other radial profiles at different transaxial levels, as well as for different axial profiles (plots not shown). In particular after normalization, as we observe in Fig. 5.6, profiles almost overlap, which is due to the larger values range of the y axis. Therefore, it is reasonable to assume that Eq. 5.4 can be used as a good way to estimate the variance for clinical data.

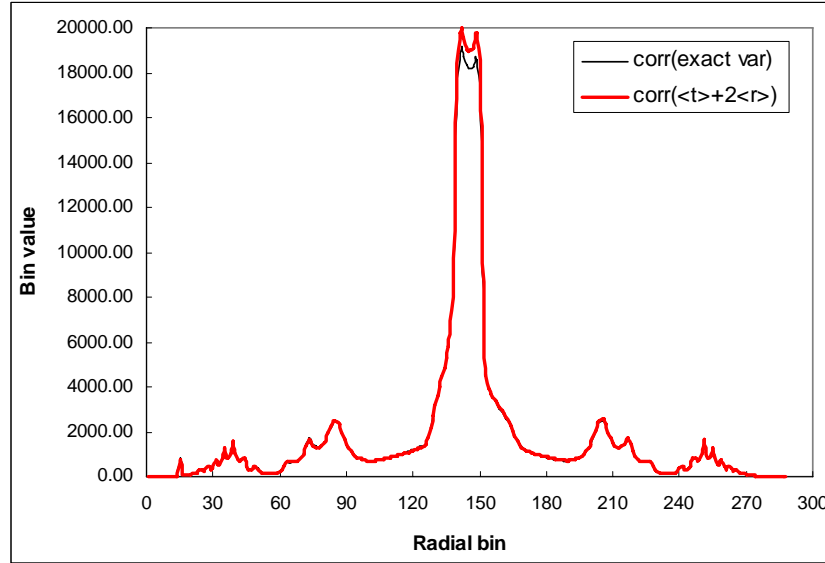


Figure 5.6 – A radial profile of the exact variance (gate 2) after normalization and $N_i^2 \cdot \mathcal{A}_i^2 \left(\langle t_i^{3D} \rangle + 2\langle r_i^{3D} \rangle \right)$. Profiles drawn from projections at the chest level for segment 0 and $\phi = 0^\circ$.

As referred, the use of Eq. 5.4 for estimating the variance requires a previous estimation of $\langle r_i \rangle$ - the mean of each value obtained from the random coincidences.

However, during the clinical exams on the HR+ scanner, random data are not collected. The delayed technique is used, but the randoms are subtracted online. Therefore, just the trues are stored and one has no access to delayed. In order to estimate the randoms, a posteriori, the strategy is to use the singles (cf. Fig. 1.12) recorded in each detector. This is done by considering that the random event count rate is a function of the square of the number of disintegrations per second and, for detectors a and b , r_{ab} is given by (cf. section 1.3.5 of the present thesis):

$$R_{ab} = 2\tau \cdot S_a S_b \quad \text{Eq. 1.1}$$

where S_a and S_b are the single event rates for detectors a and b , respectively, and 2τ is the width of the coincidence timing window.

Based on this approach, the mean value for the randoms is estimated as:

$$\langle r_i \rangle = 2\tau S_{ai} S_{bi} \quad \text{Eq. 5.5}$$

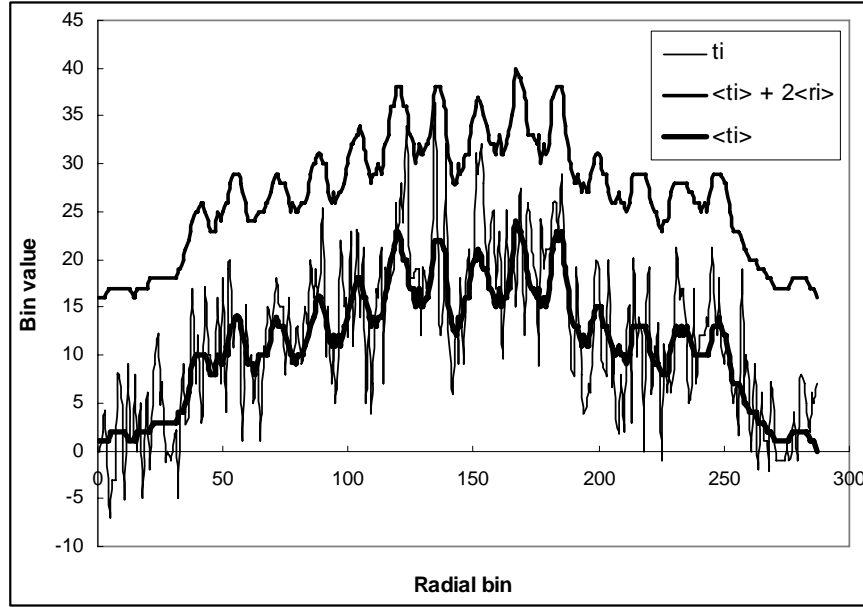


Figure 5.7 – A radial profile of the trues, ti , smoothed trues, $\langle ti \rangle$ and $\langle ti \rangle + 2\langle ri \rangle$. Data extracted from a transaxial plane crossing the chest of one of the patients, for segment 0 of a projection obtained for an azimuthal angle equal to zero.

It is important to underline that random data obtained as described above is not smoothed. This is because the single rate is high compared to the trues' rate and, therefore, the single are much less noisy than the trues [Pautrot, 03].

Once $\langle r_i \rangle$ and $\langle t_i^{3D} \rangle$ are estimated, the variance is calculated using Eq. 5.4. In Figure 5.7 we plot the a radial profile of the trues, t_i^{3D} , the smoothed trues, $\langle t_i^{3D} \rangle$, and $\langle t_i^{3D} \rangle + 2\langle r_i \rangle$, before normalization and attenuation correction. This profile was extracted from segment 0, azimuthal angle equal to 0° , and corresponds to one of the clinical exams used for the present study. The profile belongs to a transaxial plane crossing the chest.

In practice, i.e., considering the computation routines used for the different tasks, the original file with the trues (gate 1) are appended the values of $\langle t_i^{3D} \rangle + 2\langle r_i \rangle$ (gate 2), and $\langle t_i^{3D} \rangle$ (gate 3). The corrections for attenuation and normalization are done in the same stage, for the three types of data, in accordance with the corresponding equations. The result is, therefore, a single file, with the corrected data in gate 1 (trues), the variance in gate 2, and the mean value in gate 3 (cf. App. A.8).

5.4 – FORE and the weighting factors

Clinical data were supposed to be reconstructed in both 2D and fully 3D mode. Since data were acquired in 3D mode, it was necessary to rebin data into 2D. As for simulated data, the Fourier Rebinning (FORE) algorithm was used. So, once more the need to include the effect of FORE on the variance arose.

In order to model this effect it was used the approach described in chapter 3: a Single Slice Rebinning (SSRB) on the variance. This is mathematically formulated as in Eq. 3.5:

$$\sigma^2(y_i^{2D}) = K(\tilde{z}) SSRB \left\{ N_i^2 \cdot \mathcal{A}_i^2 \left(\langle p_i^{3D} \rangle + \langle r_i \rangle \right) \right\} \quad \text{Eq. 3.5}$$

where $K(\tilde{z})$ are the slice variance reduction factors.

In Figure 5. we plot the radial profiles of the data after FORE, for one bed position (at the chest level) extracted from one of the exams. The projection data were obtained for an

azimuthal angle equal to 0° and the 32nd transaxial plane. The gate number is the aforementioned: gate 1, trues; gate 2, variance; gate 3, mean value.

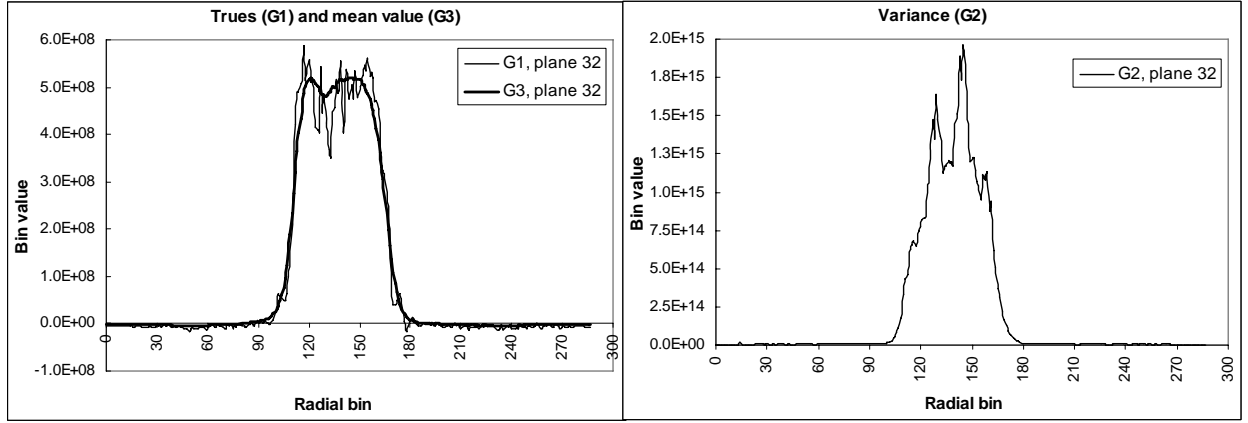


Figure 5.8 - Radial profiles of the data after FORE. The projection data were drawn was obtained for one bed position of one of the clinical exams (azimuthal angle equal to 0° ; transaxial plane 32). Gate 1 (G1), trues; Gate 2 (G2), variance; Gate 3 (G3), mean value.

An axial profile of the variance, along the scanner axis, i.e., at the centre of the FOV (radial bin #144), is plotted in Figure 5.9. The projection is once more shown for $\phi = 0^\circ$.

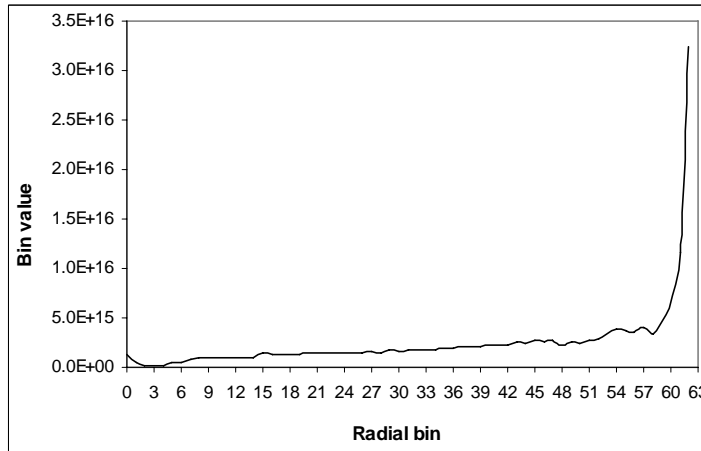


Figure 5.9

Radial profile of the variance after fore, at the centre of the FOV. Projection obtained for $\phi = 0^\circ$.

A flowchart of the whole data processing after appending the values of $\langle t_i^{3D} \rangle + 2\langle r_i \rangle$ (gate 2) and $\langle t_i^{3D} \rangle$ (gate 3) to the original file with raw data (trues), raw.S, is depicted in Figure 5.10

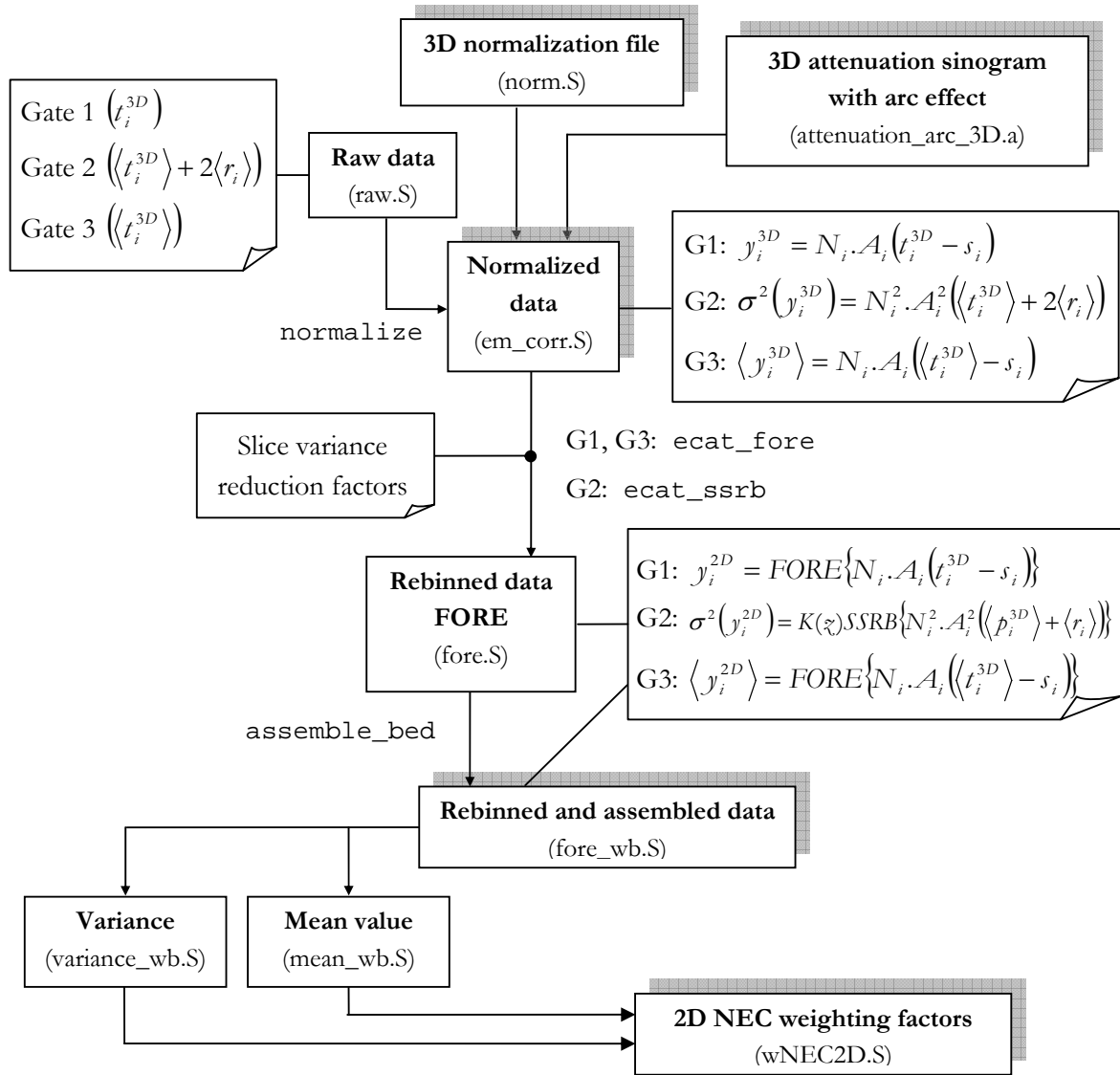


Figure 5.10 – A flowchart of the data processing stages and NEC weighting factors creation. Close to each box is indicated the routine used for generating the corresponding file.

After the rebinning operation (FORE applied to gates 1 and 3 and SSRB to gate 2), data are assembled so that instead of being split among different bed positions, they are concatenated into a virtual longer single bed. The result is a set of transaxial planes, whose number is equal to the sum of the number of planes for each bed position (63) minus a certain number of them corresponding to bed overlap. For example: for 7 bed positions (bp) with a bed overlap (bo) corresponding to 7 planes, the total number of transaxial planes after `assemble_bed` will be equal to $7 \text{ (bp)} \times 63 - 6 \times 7 \text{ (bo)} = 399$ planes.

The very important process of determining the 2D NEC weighting factors incorporates knowledge gained from the simulation studies. As explained in chapter 4, the use of 3D NEC weights raised the problem of how to deal with bins over the boundaries of the object⁵. The problem, we recall, is a consequence of the fact that mean values for bins over the object boundaries are very small. Since the computational implementation of the algorithm requires the reciprocals of the weights, i.e., $w_{NEC2D} = \sigma^2(y_i^{2D}) / \langle y_i^{2D} \rangle$, very low values of $\langle y \rangle$ correspond to very high values of w_{NEC} . Therefore, instead of using a ceiling in the OSEM algorithm to limit the pixel values that could result from this particular effect, the attenuation map is used as a mask to identify bins in the sinogram that are inside the object.

This is the role of the `poids_nec` routine: it takes the attenuation sinogram and identifies all the bins where the attenuation is less than 1.05. Since 1 is recorded for bins whose corresponding LOR does not cross the object – there is no attenuation along such an LOR – establishing a threshold in the attenuation as 1.05 is equivalent to assume that if the photon beam is attenuated less than 5% then the beam is not crossing the interior of the object but, eventually, its boundaries. Therefore, not only the estimated mean value will be very small, resulting in very high weighting factors, but also it won't be too much relevant for the final reconstructed image. The strategy implemented through the use of `poids_nec` is to set the NEC weighting factors for these bins equal to 1, replacing the otherwise extremely high values of these weighting factors.

Additionally, it is also possible to anticipate a situation where some bin inside the object (attenuation higher than 1.05) records a very small number of counts. Once again, the estimated NEC weight will be very high. To avoid the existence of a bin with a very high value, in relation its neighbors, of the corresponding NEC weighting factor, it was defined a ceiling which is established as a function of the other radial bins for the same axial and azimuthal coordinates [Pautrot, 03].

Figure 5.11 represents a profile of the 2D NEC weighting factors, for one of the clinical exams (drawn from a projection obtained for $\phi = 0^\circ$). The ceiling is clearly observed in the figure, and it is possible to infer that it is being used for bins over the object boundaries.

⁵ During the calculation of the 2D NEC weighting factors and subsequent 2D NECOSEM reconstruction, for simulated data, the problem was not found. However, the latter adopted solution remains valid for 2D and could have been used.

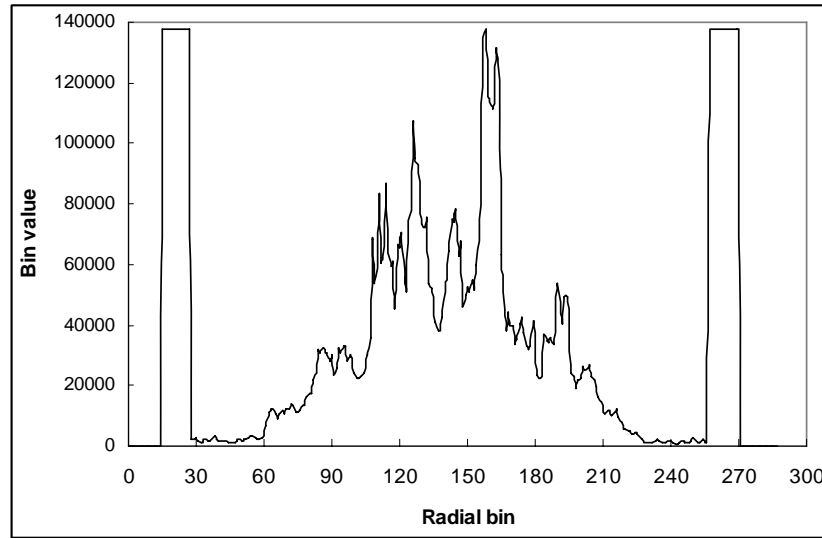


Figure 5.11 – A radial profile of the calculated 2D NEC weighting factors, after assemble bed, for one of the clinical exams (drawn form a projection obtained for $\phi = 0^\circ$).

The weighting factors to be used by the AWOSEM algorithm are the reciprocals of the assembled attenuation factors stored in the file “attenuation_orig_3D_wb.a”. For the fully 3D counterpart – the ANWOSEM3D algorithm – the file with the attenuation factors with arc effect, “attenuation_arc_3d.a”, is first normalized⁶ by using the normalization sinogram, “norm.S”, within the usual routine: `normalize`. Data in the resulting file, “wmult.S”, are, then, taken as the reciprocals of the weighting factors.

5.5 – Image reconstruction and final results

5.5.1 – Image reconstruction

After correcting data as previously described and establishing the weighting factors, images were reconstructed using two different strategies:

- 2D mode: using the AWOSEM2D and NECOSEM2D algorithms, after FORE;
- fully 3D mode: using the ANWOSEM algorithm.

A schematics of input and output data is shown in Figure 5.12.

⁶ Where, among others, data are corrected for the arc effect.

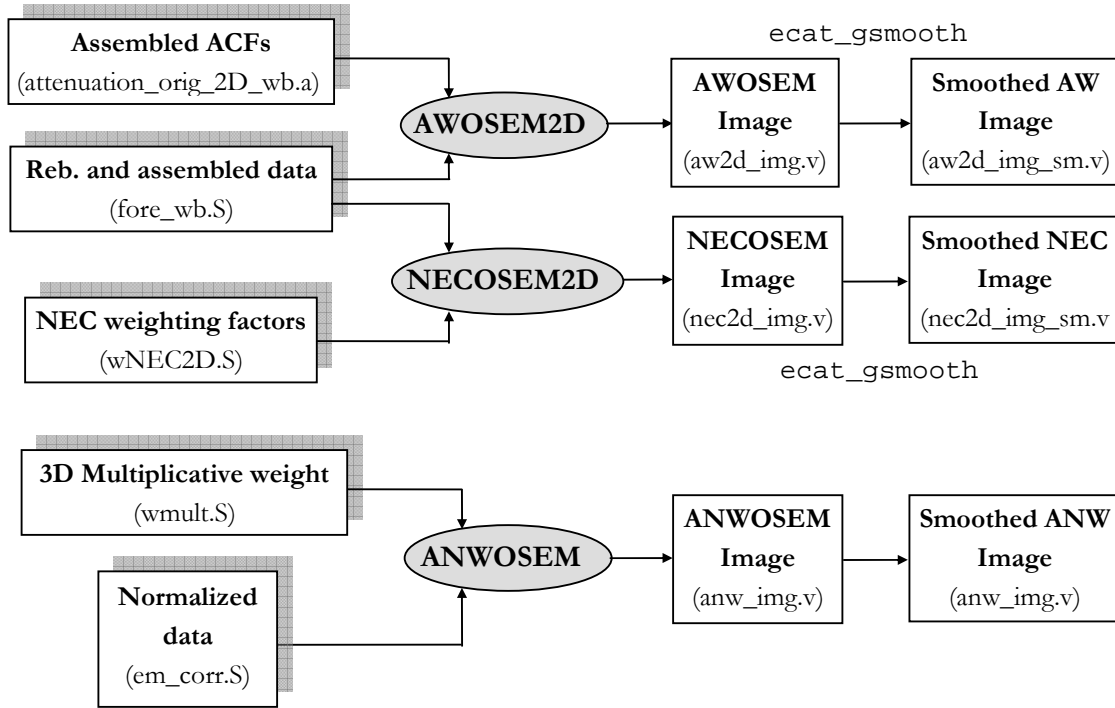


Figure 5.12 – Diagram of data input and output in the reconstruction stage, for the three algorithms: AWOSEM, NECOSEM, and ANWOSEM.

For all the reconstruction schemes, 8 subsets were used and a zoom factor of 1.5. Images were obtained after 4 iterations of the algorithm. This choice was made based on the usual clinical protocol at the S. H. F. J., where reconstructed images after 4 iterations of FORE+AWOSEM are used by the physician when performing the clinical task of tumor detection in whole-body PET images, and is coherent with the criteria previously adopted for the simulation studies.

Reconstructed images were finally post-smoothed using a Gaussian filter with a FWHM equal to 6 mm. This post-processing was adopted not only considering the simulation results (cf. section 4.5) but also the previous research work by C. Lartizien [Lartizien, 00; Lartizien, 03] and C. Pautrot [Pautrot, 03].

5.5.2 – Reconstructed images.

In this section we show the reconstructed images using the three reconstruction approaches. Among the set of patients whose exams were used for the present study, we chose

them in accordance to their BMI [Kuczmarski, 97]: one with a low (< 19.0) body mass index (BMI), ID #1 (cf. Table 5.1), another with a normal BMI ($19.0 - 24.9$), ID #8, and three obese patients (> 30.0): one of them with a B. M. I. equal to 30.9 (ID #17), whilst the others are markedly obese (B. M. I. values of 39.0 (ID #24) and 44.6 (ID #25)).

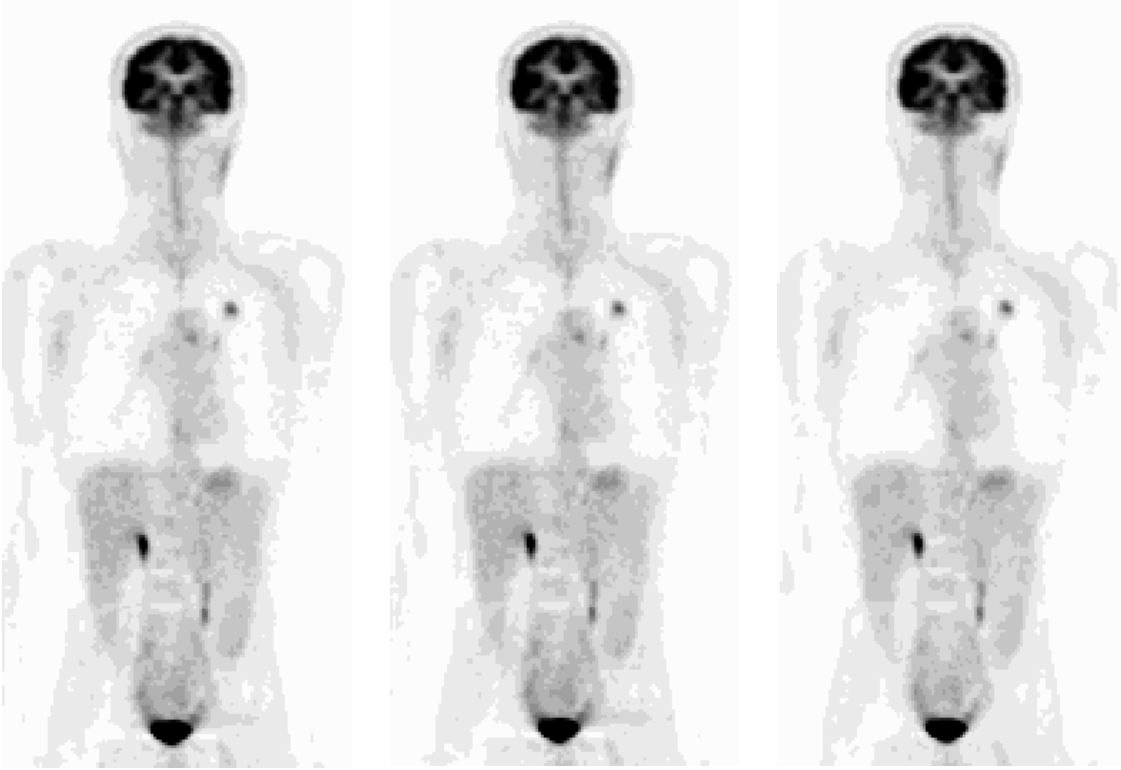


Figure 5.13 - Patient ID #1. B.M.I. = 16.2, injected dose = 384.8 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D.

In Figure 5.13 we show the reconstructed images after 4 iteration and a post-smoothing with a Gaussian filter (FWHM = 6 mm), for patient ID #1: B.M.I. = 16.2; injected dose = 384.8 MBq.

Figure 5.14 is for patient ID #8: B.M.I. = 23.7 and a dose of 338.5 MBq

Figure 5.15 is for patient ID #17: B.M.I. = 30.9 and a dose of 407 MBq. As is the former cases, images were obtained after 4 iterations and a Gaussian post-smoothing using a filter's FWHM of 6 mm. Figures 5.16 and 5.172 are for two very obese patients: patient #24, B.M.I. = 39.0, dose = 403.3 MBq; patient #25: B.M.I. = 44.6, dose = 414 MBq.

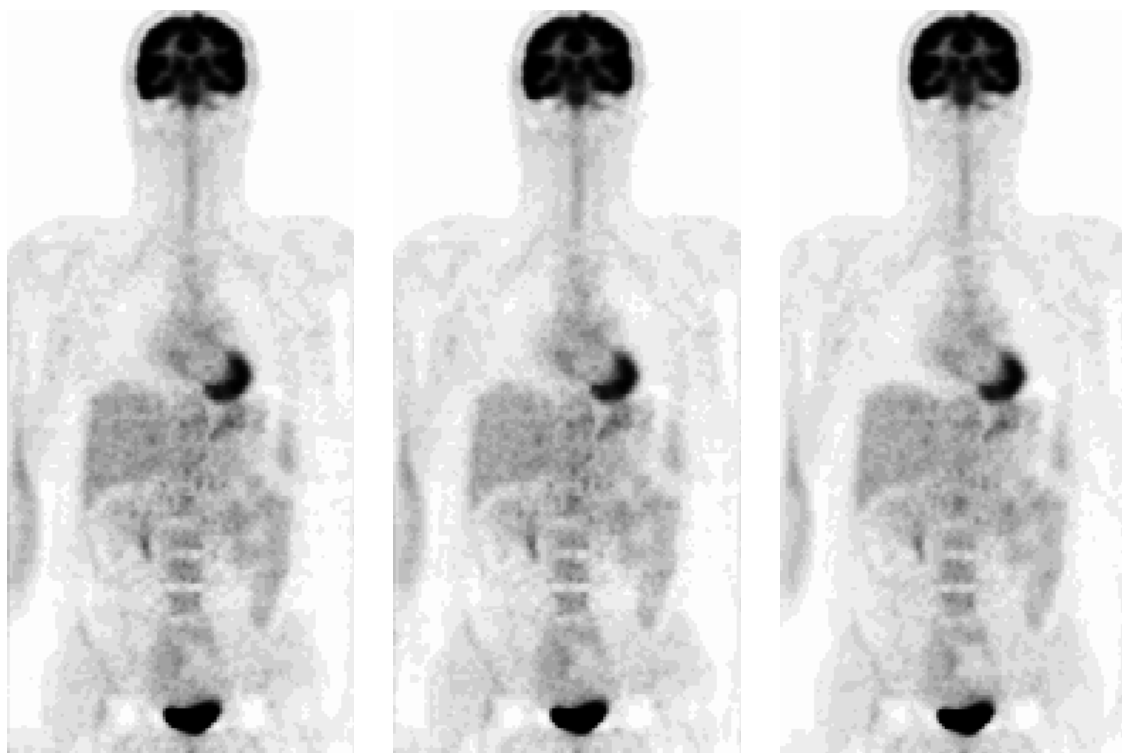


Figure 5.14 - Patient ID #8. B.M.I. = 23.7, injected dose = 338.5 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D.

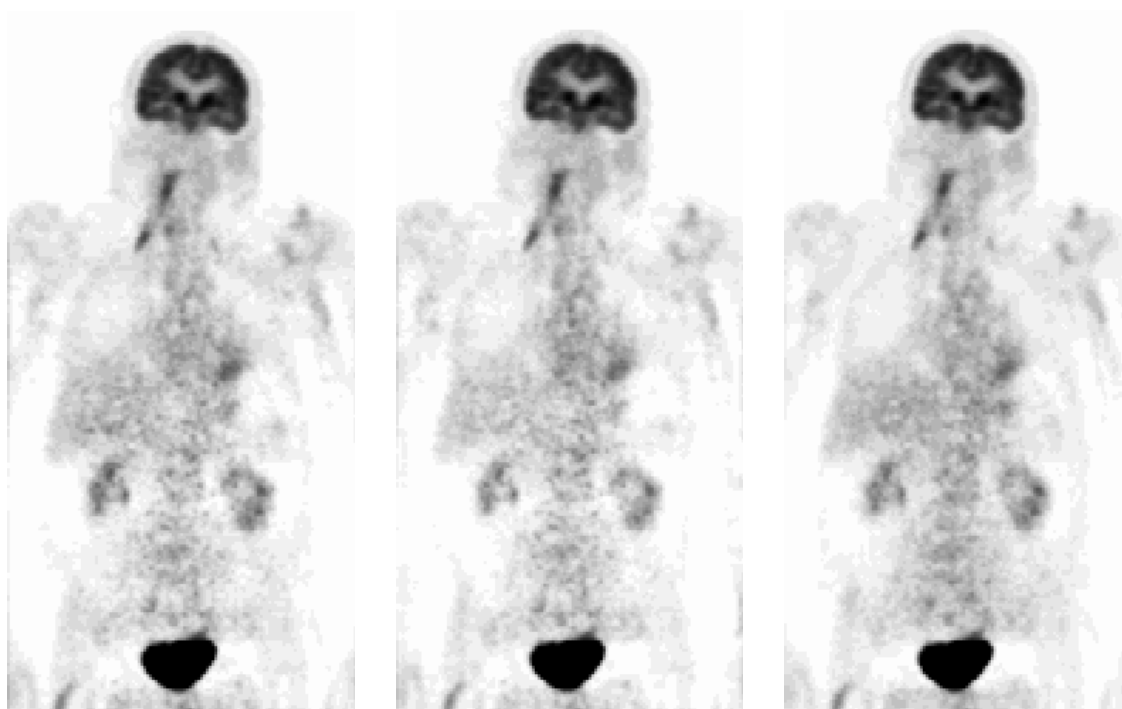


Figure 5.15 - Patient ID #17. B.M.I. = 30.9, injected dose = 407 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D.

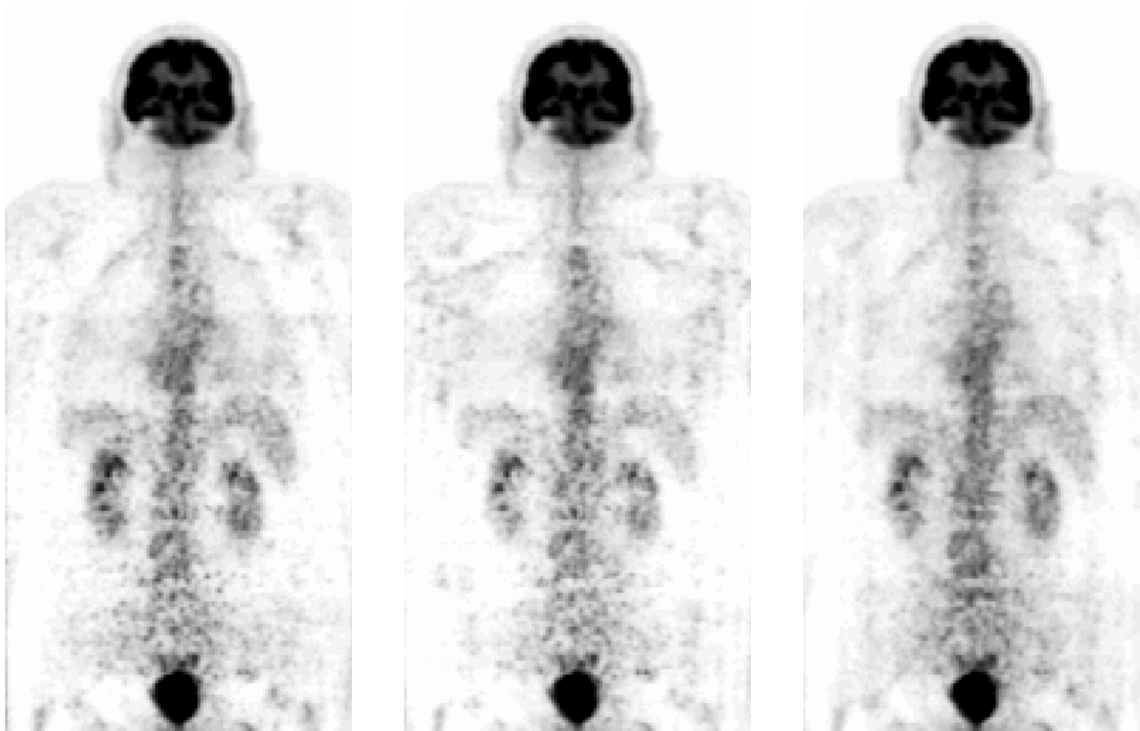


Figure 5.16: - Patient ID #24. I.B.M. = 39.0, injected dose = 403.3 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D.

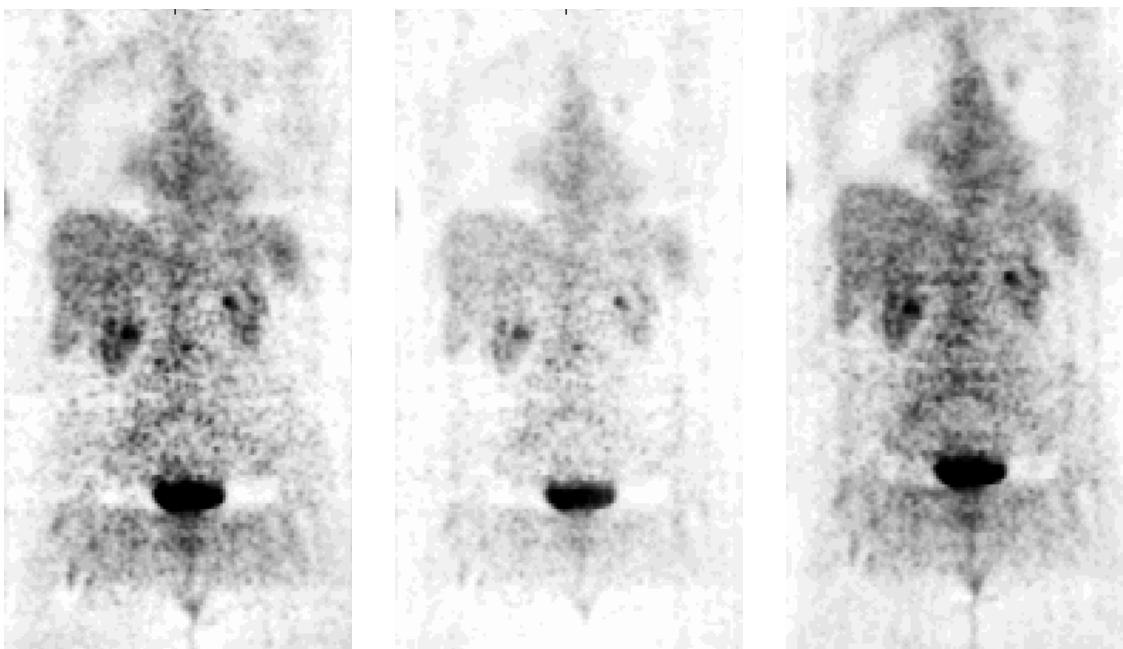


Figure 5.17 - Patient ID #25. B.M.I. = 44.6, injected dose = 414.3 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D.

5.6 – Discussion and conclusions

The first thing to retain is the complexity involved in the determination of the NEC weighting factors for clinical data. However, as shown in Figure 5.3 based on simulated data, smoothing the trues is a valid procedure for obtaining an estimation of the trues' mean values. In addition, the use of Eq. 5.4 for estimating the variance of clinical data was also validated by the study performed with simulated data (Fig. 5.6).

It is important to stress that even if there is some bias in the estimation of the mean and variance, hence a bias in the estimation of the NEC weight, this will only impact on the calculation of the weight. The expected image should not be bias because the weights are compensated for in the EM equation.

Handling the values for bins over the object boundaries requires, nevertheless, establishing ceilings and thresholds on the attenuation map (for defining which region corresponds, in fact, to the object), a procedure that is very dependent from the user judgment, i.e., there is not a definitive reason to use a specific value. It is not certain that the use of one particular threshold and ceiling will result for all the NECOSEM reconstructions of clinical data.

The reconstructed images were visually analyzed by a physician who is largely familiar with the images obtained for this sort of whole-body PET scans. She was asked to perform a preliminary visual inspection to check for some image feature suggesting an improvement of FORE+NECOSEM or AWOSEM3D over FORE+AWOSEM, in respect to the clinical task of tumor detection. The conclusion that could be drawn is the following: based on the set of images used for this study, no relevant differences were found between those reconstructed with FORE+AWOSEM and FORE+NECOSEM or ANWOSEM. This is the case even for obese patients, who are those for whom FORE+NECOSEM could eventually result in images with better quality.

So, from this observation, it is possible to infer that the additional complexity involved in FORE+NECOSEM does not necessarily translate into an increase on image quality. Moreover, there are cases where the contrary is true: FORE+NECOSEM reconstructed images are less useful than those obtained by using FORE+AWOSEM, for the present clinical task.

The comparison of FORE+AWOSEM and ANWOSEM3D does not show better performance of the last over the former. This is, however, a somewhat unexpected result. We stress, nevertheless, that this statement is valid just for that particular set of images, and no detailed observer study was performed in order to generalize it.

Final discussion and conclusion

As suggested by the title and reinforced in the Introduction, the goal of the research work described in the present document was to incorporate an accurate statistical modeling in the OSEM reconstruction algorithm, for whole-body PET imaging.

This was achieved through the use of weighted versions of OSEM, instead of the regular implementation of this algorithm. In this way, it was possible to restore the Poisson-like characteristics of the corrected data, as theoretically required by OSEM.

Another important assumption of this work was that all the collected data were 3D data. This had two different and relevant implications. First of all, two approaches were used for image reconstruction: fully 3D reconstructions (ANWOSEM3D, SPOSEM3D and NECOSEM3D) and 2D reconstructions after rebinning (FORE+AWOSEM and FORE+NECOSEM). The comparison between the performances of these two approaches was a very important issue of our investigation. In this context, differences in image reconstruction times became obvious: we have experienced that fully 3D reconstructions were 8 to 10 times more time consuming than FORE + 2D OSEM reconstructions, for the same number of iterations, underlining the importance of the rebinning operation.

Moreover, the use of the Fourier Rebinning (FORE) requires data corrected for geometrical and physical effects. Due to this, and under the FORE + 2D OSEM approach, input data given to the OSEM algorithm are not Poisson distributed. This is to say: what could otherwise be envisaged in a fully 3D reconstruction, i.e., accounting for the corrections in the system matrix (as in OP-OSEM¹), became forbidden by the use of FORE. Thus, the pertinence of studying the possibility of incorporating an accurate statistical model for the data was inevitable as soon as the FORE+2DOSEM approach was adopted. The same is true for most of the fully 3D reconstruction schemes often considered, since they take as input this sort of non-Poisson corrected data.

The strategy of using weighted versions of OSEM, as referred in Chapter 3, in some cases restores the Poisson-like characteristics of the data – NECOSEM and SPOSEM3D – but not always – AWOSEM and ANWOSEM, depending on the weight strategy used. Therefore, just in the former cases is the Poisson data requirement completely fulfilled. The comparison

¹ Not used in our work.

between the performances of these two groups of algorithms working over the same corrected non-Poisson data was done.

Schematically, there are two different perspectives behind the performance analysis we have done: the 2D versus 3D perspective (P1), and the complete and non-complete restoring of Poisson data (P2) (Fig. C1).

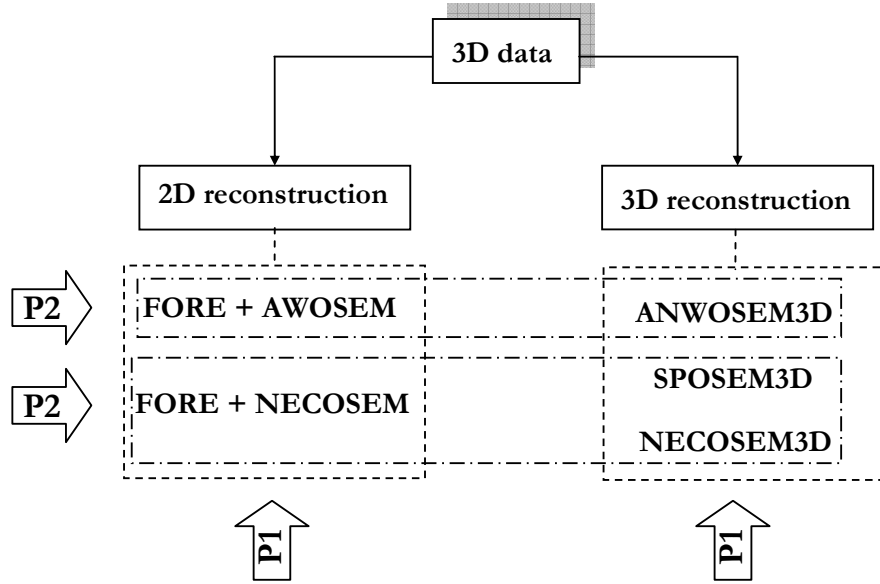


Figure C1 – The two different perspectives behind the developed performance analysis of reconstruction algorithms: P1, the 2D versus 3D perspective; P2, the complete versus non-complete restoring of Poisson data.

From the point of view of how data were obtained, two complementary studies were developed: one with simulated data produced by using the ASIM analytical simulator; another, with real clinical data resulting from exams undertaken on the S.H.F.J.'s HR+ scanner.

We started by an extensive simulation study, where the use of the ASIM simulator allowed us to have control over the data generated. In particular, it was possible to keep track of the mean value and of the variance used to add pseudo-random noise to the data, as well as to know, without any error, the attenuation correction factor (ACF) for each LOR. Considering the weighted OSEM reconstruction, this is the ideal situation: it was possible to calculate exactly, for each data bin, the corresponding weighting factor, in all the cases, i.e., for all the weighted reconstruction algorithms. Therefore, any additional effect introduced by the uncertainty associated with the weight estimation was discarded in the simulation study.

The simulation conditions were as realistic as possible, in terms of both phantom conception and simulator capability. The tumors' dimension, number and activity ratios were chosen to mimic situations that were validated in previous studies [*Lartizien, 03; Lartizien, 01*].

The very large amount of data² made human observers studies unaffordable. So, the alternative was to resort to numerical observers. The NPWMF numerical observer had been previously tested under similar imaging conditions, showing that its detectability index allows good correlation with human observer detectability [*Lartizien, 04*]. Therefore, we relied on this observer to compare the performance of the reconstruction algorithms. In addition, the noise and the tumor-to-background contrast (cf. section 4.3.2) were also used to access the image quality.

To make the task of comparing the performance of the algorithms feasible, we had to fix two parameters: the post-smoothing filter applied to reconstructed images, and the number of iterations to be used.

We started by studying the effect of the post-smoothing filter. The NPWMF detectability index was adopted to verify the existence of an optimum value for the FWHM of this Gaussian filter. Reconstructed images were smoothed using different values of the FWHM of the filter, something that was done for images obtained after different numbers of iterations. The results showed the existence of a maximum value for the NPWMF detectability index, when this index was plotted against the FWHM of the post smoothing filter. This was true for all the studied OSEM versions, but the value was found to be different for different organs. However, for 1 cm diameter spherical tumors, a FWHM of 5.5 mm was in most cases close to the optimum value. Also significant is the fact that this trend revealed to be independent from the number of iterations used to obtain the images over which the analysis was being done. So, these results allowed us to assume that the use of a Gaussian post-smoothing filter with a FWHM equal to 5.5 mm is the best compromise for the task of maximizing the tumor detectability, under the conditions established for our study.

On the contrary, the NPWMF index could not be used to define the optimum number of iterations: it was verified that this index does monotonically decrease with the number of iterations. Therefore, some other criterion had to be chosen, and we decided to use the tumor-to-background contrast level as a figure-of-merit to set the number of iterations. This allowed us to establish the equivalence between different weighted OSEM versions. First of all, we assumed as reference the tumor-to-background contrast level in FORE+AWOSEM reconstructed images, since it is widespread in clinical practice. After, we searched, for the other

² More than forty thousand files with images.

algorithms, the number of iterations resulting in images with an equivalent contrast for the corresponding organ. As it was possible to fit this criterion for a different number of FORE+AWOSEM iterations, we fixed in 4 the number of iterations for this reconstruction scheme, which is a number usually adopted in clinical routine, and used the aforementioned criterion to establish the number for the other algorithms.

As shown in Fig. 4.9, except for SPOSEM3D, the tumor-to-background contrast was found to be equivalent, after 4 iterations, for all the situations taken into account. In the SPOSEM3D case, we observed, without surprise, the slower convergence (when compared to the other algorithms), due to the presence of the additive term. So, in this case, in order to reach similar tumor-to-background contrasts it was necessary to iterate almost four times more than for the other algorithms: 15 iterations instead of 4. Nevertheless, since this difference is dependent on the ratio between scattered plus random coincidences and unscattered true coincidences, this ratio between the number of iterations is valid just for this particular study.

Based on the results described in the paragraphs above, we fixed both the FWHM of the post-smoothing filter and the number of iterations. Then, we were in position to start the direct comparison of the reconstruction algorithms, using the NPWMF detectability index as a mean to access image quality.

We did, at first, a comparison between the performance of FORE+AWOSEM and FORE+NECOSEM. Considering the NPWMF indices for different tumors in each region, it became clear that FORE+NECOSEM had an overall performance better than FORE+AWOSEM. The same conclusion could also be drawn when comparing ANWOSEM3D and FORE+AWOSEM (the former resulted in higher detectability indices than FORE+AWOSEM). Less evident differences were found when comparing ANWOSEM3D with FORE+NECOSEM. The comparison between SPOSEM3D and the two FORE + 2D OSEM algorithms resulted similar to those obtained for ANWOSEM3D.

The differences between NECOSEM3D and FORE + NECOSEM revealed to be much less obvious. On the contrary, the obtained results allowed us to argue for a better performance of NECOSEM3D over FORE+AWOSEM. In relation to the other 3D algorithms, no relevant differences were found, except for one situation concerning the soft tissues.

So, one important conclusion is that: even for the case of simulated data, where one has the possibility to calculate the exact NEC weights, the additional complexity associated with NECOSEM3D has no direct positive impact on the detectability of tumors.

Still referring to the simulation study, in spite of being possible to calculate the exact values of the NEC weights, we verified that the behavior of the fully 3D NECOSEM depends

very much on the profile of the noiseless NEC weights. This is due to rounding and other numerical artifacts that could result from the processing of these weights. We observed that sharp variations on the profile of the NEC weights lead to reconstructions very sensitive to the noise present on the sinograms. This is so more critical as in some cases the reconstruction proceeded as if some points drained all the activity during the reconstruction process: the final image after a few iterations was reduced to zero everywhere except in these points. We tried to avoid this problem by using the attenuation map to establish the borders of the object, and artificially setting the NEC weight equal to one in every sinogram bin corresponding to LORs not crossing the object or crossing its very boundaries. In such a way we were trying to limit sharp variations on the NEC weights profiles. This strategy proved to decrease the number of reconstructions where some points behaved as drains, but didn't completely avoid the existence of these cases.

This fact reinforces the complexity associated with the use of NECOSEM3D in practice, which, as referred, proved not to perform better than the other 3D weighted versions of OSEM.

With this set of results, obtained from simulated data, we stepped forward to the application of some of these algorithms to real clinical data. However, for the reasons adduced above, we discarded NECOSEM3D reconstructions. Moreover, due to the fact that, even after 15 iterations, SPOSEM3D was shown to produce images whose usefulness for the task of tumor detection is similar to those obtained with 4 iterations of ANWOSEM3D, we decided to use just the latter for the case of fully 3D reconstructions of clinical data.

These data were chosen among the exams that had been done at the S.H.F.J., on the HR+ scanners, and were stored on the proper backup computer. We retrieved 25 of them, corresponding to patients with different body mass indices, in the range of 16.2 kg/m^2 up to 44.6 kg/m^2 .

Data were reconstructed with FORE+AWOSEM, FORE+NECOSEM and ANWOSEM3D. Markedly different from the simulation study is the fact that one has no access to the mean value and the variance of each bin: these statistical moments must be estimated. Therefore, the NEC weights were dependent from this estimation, and, in consequence, were no more exact neither noiseless values (as for simulation studies).

Our study suggests smoothing the trues as a procedure for obtaining an estimation of the trues' mean values, as well as the use of Eq. 5.4 for estimating the variance of clinical data. Thus, the possibility to have a valid procedure for the calculation of the NEC weights exists. The drawback is that this procedure can introduce some bias in the estimations of the two referred moments and, hence, in the estimation of the NEC weight. However, it must be stressed that

the expected image should not be biased because the weights are compensated for in the EM equation.

As for the simulation study, estimating NEC weights for bins over the object boundaries required establishing ceilings and thresholds on the attenuation map (for defining which region corresponds, in fact, to the object). This procedure is very dependent from the user judgment, because there is not a definitive reason to use a specific value. In other words: we used a threshold of 5% on the attenuation to define the object boundaries, but we have no strong arguments for object against the use of 8%, for example. It is also not certain that the use of one particular threshold and ceiling will work in the same way for all the NECOSEM reconstructions of clinical data.

Reconstructed images were visually analyzed by a physician who is largely familiar with the images obtained for this sort of whole-body PET scans. She was asked if there was any image feature suggesting an improvement of FORE+NECOSEM or AWOSEM3D over FORE+AWOSEM, in respect to the clinical task of tumor detection. Based on the set of images used for this study, the answer was: there are no relevant differences between those reconstructed with FORE+AWOSEM and FORE+NECOSEM or ANWOSEM. This is the case even for obese patients, who are those for whom FORE+NECOSEM could eventually result in images with better quality.

The fact that a single physician was involved in this study doesn't give room for drawing definitive conclusions. However, the results we have obtained are in-line with the conservative perspective, i.e., the additional complexity involved in FORE+NECOSEM does not necessarily translate into an increase on image quality. Moreover, in some cases the contrary was found to be true: FORE+NECOSEM reconstructed images less useful, for the present clinical task, than those obtained by using FORE+AWOSEM.

Concluding in the opposite direction – i.e., arguing by the advantages of FORE+NECOSEM –, under the conditions of our study, would certainly be more daring. However, we assume that more work, involving more physicians, is certainly necessary to strengthen the scope of our conclusion.

In any case, we believe that we gave an additional contribution to show that is possible to use NECOSEM with clinical data. As said above, we also believe that the additional complexity introduced by the estimation of the NEC weights will probably, in practice, slant the complexity-usefulness trade-off to the side of costs, hence making this algorithm not advantageous for reconstructing clinical data, if the clinical task is to detect tumors in whole-body PET imaging.

The results we have obtained for the AwnOSEM3D reconstruction algorithm, on the other hand, are somehow surprising, not only considering the simulation study we have done before, but also other additional work developed within the research group at the S.H.F.J.. Therefore, one suggestion for the future is to proceed to the investigation of the fully 3D reconstruction algorithms. The inclusion of a PSF modeling within the OSEM algorithm, in order to increase the spatial resolution, is already being done with success [Sureau, 06]. Since there is not the constraint to correct data prior to the reconstruction, incorporating the aforementioned corrections inside the reconstruction stage is also a very interesting field with large room for future developments.

A.1 – Coordinate Systems in 2D and 3D

A.1.1 – 2D case

In 2D, the two coordinate systems to be considered are co-planar. One of them is fixed, (x, y) , and could be thought as having an origin coincident with the scanner axis (independent from the relative orientation of the plane). The other, (x', y') , is a rotating coordinate system and the relative position of their axis from the corresponding axis of the former system defines the azimuthal angle, ϕ . These systems are illustrated in Fig. A1.1.

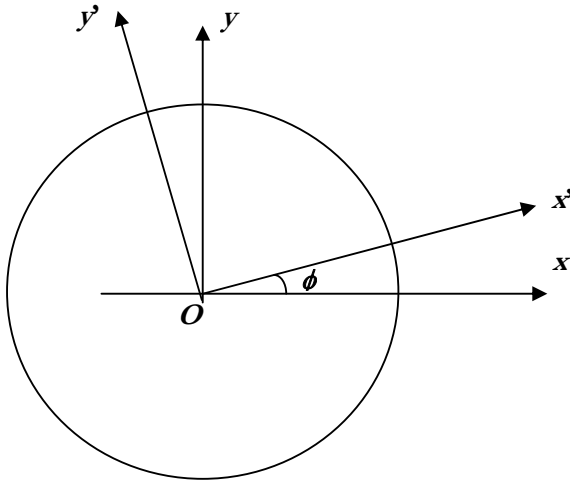


Figure A1.1

The two coordinate systems. (x, y) is fixed; (x', y') is the rotating system and ϕ is azimuthal angle.

The relation between the coordinates in the two systems is established in equations A1.1 – A1.4 [Townsend, 93; Bruyant, 02]:

$$\begin{cases} x' = x \cos \phi + y \sin \phi & \text{Eq. A1.1} \\ y' = -x \sin \phi + y \cos \phi & \text{Eq. A1.2} \end{cases}$$

$$\begin{cases} x = x' \cos \phi - y' \sin \phi & \text{Eq. A1.3} \\ y = x' \sin \phi + y' \cos \phi & \text{Eq. A1.4} \end{cases}$$

A.1.2 – 3D case

In 3D the situation is somehow more complex, but the equations relating the two coordinate systems, (x, y, z) fixed and (x', y', z') rotating, are easily deduced if the whole rotation is split and understood in terms of three sequential steps (Fig. 1): one axis renaming and two rotations around two different axis.

The previous axis renaming (step 1, shown in Fig. 1a) is relevant for the sake of coherence with the integration variables that are usually found both in literature and

reconstruction software. As consequence of this operation, in the new rotating system the scanner axis is coincident with axis y'' ; in-plane integration along a LOR translates into an integration on variable z'' ; and the radial position is described by variable x'' . The second step (Fig. 1b) is a rotation around the scanner axis (z or y'). It defines the azimuthal angle which is usually identified with ϕ . Finally, there is another rotation, this time around x' axis, that changes the orientation of the planes inside which integrations are done, i.e., where LORs are to be considered. This last step results in the introduction of the polar angle, θ , a particular feature of 3D PET insofar as for the 2D case all the considered planes are direct planes (planes normal to the scanner axis).

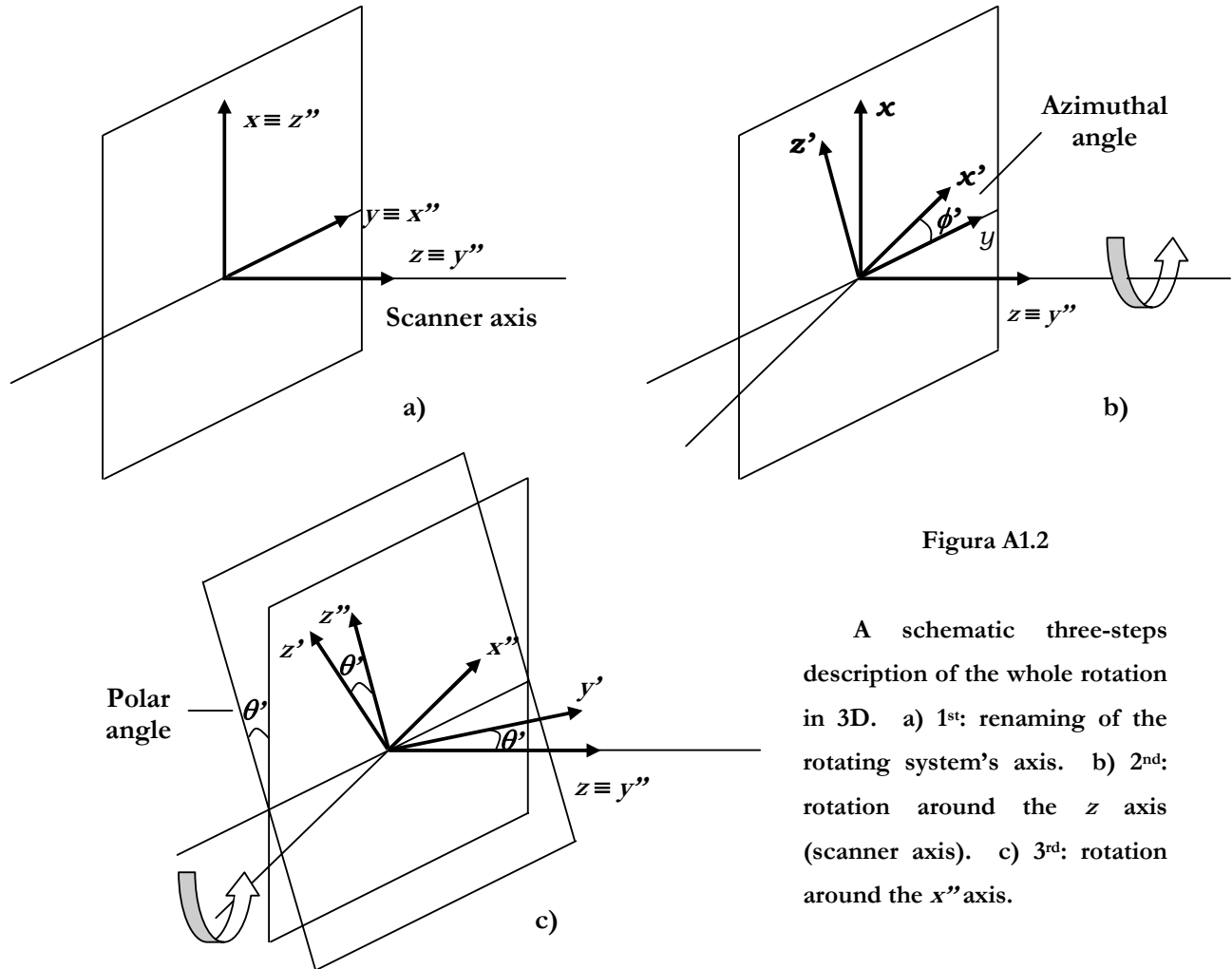


Figura A1.2

A schematic three-steps description of the whole rotation in 3D. a) 1st: renaming of the rotating system's axis. b) 2nd: rotation around the z axis (scanner axis). c) 3rd: rotation around the x'' axis.

The operations above result in relations between coordinates that could be mathematically described as following:

$$\begin{cases} x = x'' \\ y = y'' \\ z = z'' \end{cases} \implies \begin{cases} x = x'' \sin \phi' + z'' \cos \phi' \\ y = x'' \cos \phi' - z'' \sin \phi' \\ z = y'' \end{cases} \implies \begin{cases} x'' = x' \\ y'' = y' \cos \theta' - z' \sin \theta' \\ z'' = y' \sin \theta' + z' \cos \theta' \end{cases}$$

and so, using $\theta' = -\theta$ and $\phi' = -\phi$ [Kak, 88; Defrise, 98]:

$$\begin{cases} x = -x' \sin \phi - y' \cos \phi \sin \theta + z' \cos \phi \cos \theta \\ y = x' \cos \phi - y' \sin \phi \sin \theta + z' \sin \phi \cos \theta \\ z = y' \cos \theta + z' \sin \theta \end{cases} \quad \begin{array}{l} \text{Eq. A1.5} \\ \text{Eq. A1.6} \\ \text{Eq. A1.7} \end{array}$$

A.2 – The Radon Transform

If $f(x, y)$ is an arbitrary function defined on some domain D of R^2 and L is any line in the plane, then the mapping defined by the line integral (projection) of f along all possible lines L is the two-dimensional Radon transform of f , $\mathfrak{R}f$ or $\hat{f}(x', \phi)$, provided the integral exists [Deans, 83].

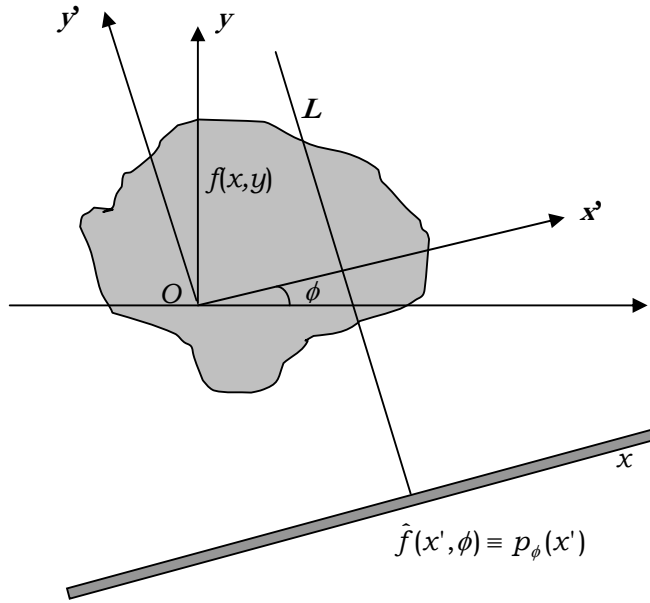


Figure A2.1

One example of a line in D , along which $f(x, y)$ is to be integrated.

Due to the different applications of the Radon transform, some variants of its mathematical expression could be found in literature. In the field of emission computed tomography, the widespread formulation is the following [Toft, 96; Deans, 83; Herman, 79]:

$$\hat{f}(x', \phi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(x' - x \cos \phi - y \sin \phi) dx dy \quad \text{Eq. A2.1}$$

which is equivalent to:

$$\hat{f}(x', \phi) = \int_{-\infty}^{+\infty} f(x' \cos \phi - y' \sin \phi, x' \sin \phi + y' \cos \phi) dy' \quad \text{Eq. A2.2}$$

In equation A2.2 we recognize the line integral of $f(x, y)$ along a line parallel to the axis y' . Radon showed that if f is continuous and has a compact support, then $\mathfrak{R}f$ is uniquely determined

by integrating along all lines L . When $\hat{f}(x', \phi)$ is known for only certain values of x' and ϕ , we have a sample of the Radon transform [Deans, 83].

Among the properties of the Radon transform [Deans, 83; Defrise, 02], there are:

- Linearity: $\mathfrak{R}[c_1 f + c_2 g] = c_1 \hat{f} + c_2 \hat{g}$
- Invariance for translation (for object space and sinogram space): $\mathfrak{R}[\mathfrak{T}_{a,b} f] = \mathfrak{T}_{a,b}[\mathfrak{R} f]$
- Invariance for rotation (for object space and sinogram space): $\mathfrak{R}[C_\theta f] = C_\theta[\mathfrak{R} f]$

It is interesting to note that Cormack, in 1964 [Cormack, 63], described a method of finding a real function in a finite region of a plane giving its line integrals along all straight lines intersecting the region, without an explicit reference to the Radon transform. The properties of the set of one-dimensional integral equations that are identified as solutions to this problem are also referred. One application that Cormack devised for this method was the determination of the distribution of positron annihilations when there is an inhomogeneous distribution of the positron emitter in matter. Thus, after the seminal paper by Radon, in 1919, (whose translation in English can be found in [Deans, 83]), this is probably the work where this formulation is applied to emission computed tomography.

The discrete Radon transform (DRT), its relation with the former classical (continuous) counterpart and inversion procedures are describe, for example, in [Beylkin, 87; Fishburn, 97; Svalbe, 04]. Among other things, it was shown [Beylkin, 87] that a discrete version of the Radon transform can be inverted only approximately if its inversion is directly based on a discretization of the Radon's formula (Eq. A2.2).

A.3 – The Backprojection Operator

Assuming an arbitrary function $\Psi(x', \phi)$, where $x' = x \cos \phi + y \sin \phi$ is the x coordinate of the conventional in-plane rotating system (Fig. A3.1), the backprojection operator, B , considered the adjoint to the forward projection [Defrise, 98], is defined as [Deans, 83; Herman, 80]:

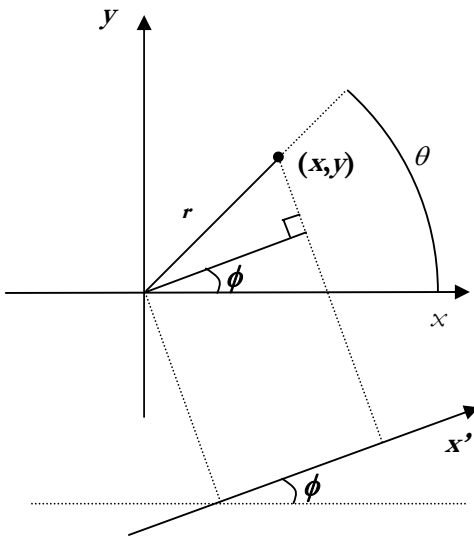
$$B\Psi(x', \phi) = \int_0^\pi \Psi(x \cos \phi + y \sin \phi, \phi) d\phi \quad \text{Eq. A3.1}$$

In terms of polar coordinates, $x = r \cos \theta$ and $y = r \sin \theta$.

$$B\Psi(r, \theta) = \int_0^\pi \Psi(r \cos(\theta - \phi), \phi) d\phi \quad \text{Eq. A3.2}$$

Thus, if $\Psi(x', \phi)$ is identified with the projection $\hat{f}(x', \phi)$, then, for a fixed angle ϕ , the incremental contribution to $B\Psi$ at the point (x, y) is the value of $\hat{f}(x', \phi)$ multiplied by $d\phi$. It is possible to show [Deans, 83] that the true image convolved with $1/r$ yields the backprojected image, $b(x, y)$: $b(x, y) = f(x, y) ** 1/r$, where $r = (x^2 + y^2)^{1/2}$ and $**$ is a 2D convolution.

In practice, $\hat{f}(x', \phi)$ is known for a discrete set of angles, ϕ , and radial points, x' . For this reason, a discrete implementation of backprojection is required. If the object is divided into voxels, denoted as $f(x_l, y_m)$ where x_l and y_m are the coordinates of the centre, this discrete implementation can be established as follows [Peters, 81]:



$$f(x_l, y_m) = \sum_{n=0}^{N-1} \hat{g}_n(x_l \cos(n\pi / N), y_m \sin(n\pi / N))$$

where $\hat{g}_n(x')$ is an interpolated version of $\hat{f}_n(x'_k)$:

$$\hat{g}_n(x') = \sum_{k=1}^N b(x'_k - x') \hat{f}_n(x'_k)$$

and $b(x')$ is an interpolation kernel.

Figure A3.1

A geometrical interpretation of the backprojection operator.

A.4 – Central Section Theorem

The Central Section Theorem [Defrise, 02; Kak, 88; Herman, 79] (also know as Central-Slice, Fourier-Slice or Projection-Slice Theorem), relates the two-dimensional Fourier transform of one function with the Radon transform of the same function.

Let us start by considering the 1D Fourier transform of function $p(x', \phi)$, in the present context identified with the function used to describe the projection data:

$$P(v_{x'}, \phi) = \int_{-\infty}^{+\infty} p(x', \phi) \exp(-2i\pi x' v_{x'}) dx'$$

Using the Radon transform it is possible to link this data function with the object function, i.e., with the 2D function describing the activity distribution, $f(x', y')$:

$$\begin{aligned} P(v_{x'}, \phi) &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(x', y') dy' \right] \exp(-2i\pi x' v_{x'}) dx' \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x', y') \exp[-2i\pi(x' v_{x'} + y' v_{y'})] dx' dy' \Big|_{v_{y'}=0} \end{aligned}$$

And finally:

$$P(v_{x'}, \phi) = F(v_{x'}, v_{y'}) \Big|_{v_{y'}=0} \quad \text{Eq. A4.1}$$

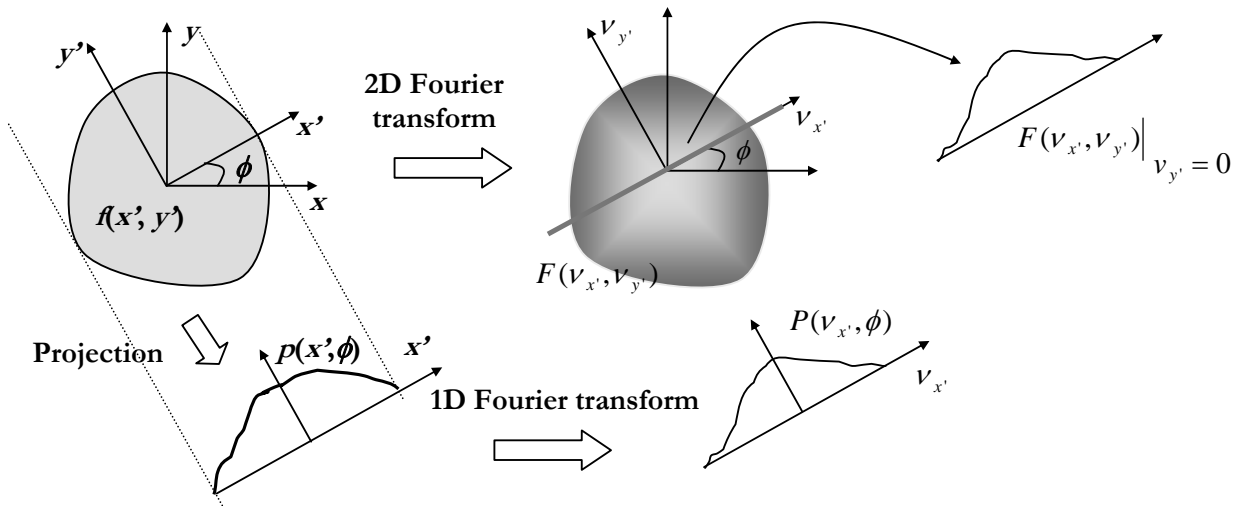


Figure A4.1 – An illustration of the 2D Central Section Theorem.

Equation A4.1 is the mathematical formulation of the Central Section Theorem, which could be stated as following: the 1D Fourier transform of a projection at an angle ϕ is equivalent to the value along a line through the origin (at the same angle) of the 2D Fourier transform of the activity distribution. This is illustrated in Figure A4.1.

The 3D extension of this theorem [Defrise, 98; Lewitt, 03] is derived from the 3D Fourier transform of $f(x, y, z)$, $F(v_x, v_y, v_z)$:

$$F(v_x, v_y, v_z) = \iiint_{\Omega} f(x, y, z) \exp[-i(v_x x + v_y y + v_z z)] dx dy dz$$

We now consider the values of the 3D Fourier transform of $f(x, y, z)$ on the 2D plane $v_y = 0$ through the origin of the 3D Fourier space:

$$F(v_x, 0, v_z) = \iint_{\Omega} \left[\int_{-\infty}^{\infty} f(x, y, z) dy \right] \exp[-i(v_x x + v_z z)] dx dz$$

The inner integral in the equation above can be identified with the 2D projection of $f(x, y, z)$ on the plane xOz for $\phi = \theta = 0$, where ϕ and θ are the azimuthal and polar angles, respectively. Thus¹:

$$F(v_x, 0, v_z) = \iint_{\Omega} p_{\phi=0, \theta=0}(x, z) \exp[-i(v_x x + v_z z)] dx dz \quad \text{Eq. A4.2}$$

Denoting by $P_{\phi, \theta}(v_{x'}, v_{y'})$ the Fourier transform of $p_{\phi, \theta}(x', y')$:

$$P_{\phi, \theta}(v_{x'}, v_{y'}) = \iint_{\Omega} p_{\phi, \theta}(x', y') \exp[-i(v_{x'} x' + v_{y'} y')] dx' dy'$$

we can use this result in Eq. A4.2, obtaining a final relation between $F(v_x, 0, v_z)$ and the 2D Fourier transform of the parallel ray projection at the orientation $\phi = \theta = 0$, for lines of integration perpendicular to the plane $y = 0$:

¹ For the sake of simplicity and since we are deriving the relation for the special case of $\phi = \theta = 0$, we do not consider the usual change in the name of the axis when establishing the relation between coordinates in the fixed and rotating systems.

$$F(v_x, 0, v_z) = P_{\phi=0, \theta=0}(v_x, v_z)$$

Using the rotation property of the Fourier transform, we can state the general relation, which is, in fact, the 3D Central Section Theorem: the 2D Fourier transform of a parallel-ray projection at a specific orientation (ϕ, θ) is equal to the values of the 3D Fourier transform on the 2D plane v_y through the origin having the corresponding orientation. This can be mathematically expressed as follows:

$$P_{\phi, \theta}(v_{x'}, v_{z'}) = F(v_{x'}, v_{y'}, v_{z'}) \Big|_{v_{y'}=0} \quad \text{Eq. A4.3}$$

A very important condition for the validity of this theorem is that the set of parallel projection should contain just non-truncated projections [DeFrise, 03].

A.5 – The Likelihood Function

The likelihood function provides information about a joint probability. It could be defined as following [Casella, 90]:

- Let $\tilde{X} = (X_1, X_2, \dots, X_n)$ be a sample, $X_k \stackrel{d}{=} X \cap F_X(x | \theta)$, depending from a parameter, θ .

$$\text{i) If } X \text{ is a random discrete vector: } X = \begin{cases} x_k & k \in K \\ p_k(\theta) = P[X=x_k] \end{cases}$$

the likelihood function, given $\tilde{X} = x = (x_1, x_2, \dots, x_n)$, is the joint probability mass function to be evaluated for various possible values of θ .

$$L(\theta | \tilde{x}) = L(\theta | x_1, x_2, \dots, x_n) = \prod_{k=1}^n p_{x_k}(\theta)$$

- ii) For the continuous case, denoting $f_x(x | \theta) = \frac{d}{dx} F_x(x | \theta)$ as the probability density function, the likelihood function, given $\tilde{X} = x = (x_1, x_2, \dots, x_n)$, is:

$$L(\theta | \tilde{x}) = L(\theta | x_1, x_2, \dots, x_n) = \prod_{k=1}^n f_x(x_k | \theta)$$

The value of the parameter maximizing the sample probability is the value for which the likelihood function attains its maximum, thus referred as the maximum likelihood estimator (MLE) of the parameter θ based on sample X , $\hat{\theta}(X)$.

According to this definition, the task of obtaining the maximum likelihood estimator involves the evaluation of the extrema of a product of n values, something that could be revealed very hard in practice. As a consequence, the logarithm plays an important role, since it converts products in sums and, very important, it is a strictly increasing function. From the latter property results that the extrema of $L(\theta | x)$ and $\log L(\theta | x)$ coincide.

Let us now consider a random Poisson variable, $X \cap P(\lambda)$, with λ unknown. Using a sample $\tilde{x} = (x_1, x_2, \dots, x_n)$:

$$\begin{aligned} x_k &= 0, 1, \dots \\ X &= \begin{cases} p_k = e^{-\lambda} \frac{\lambda^{x_k}}{x_k!} \end{cases} \end{aligned}$$

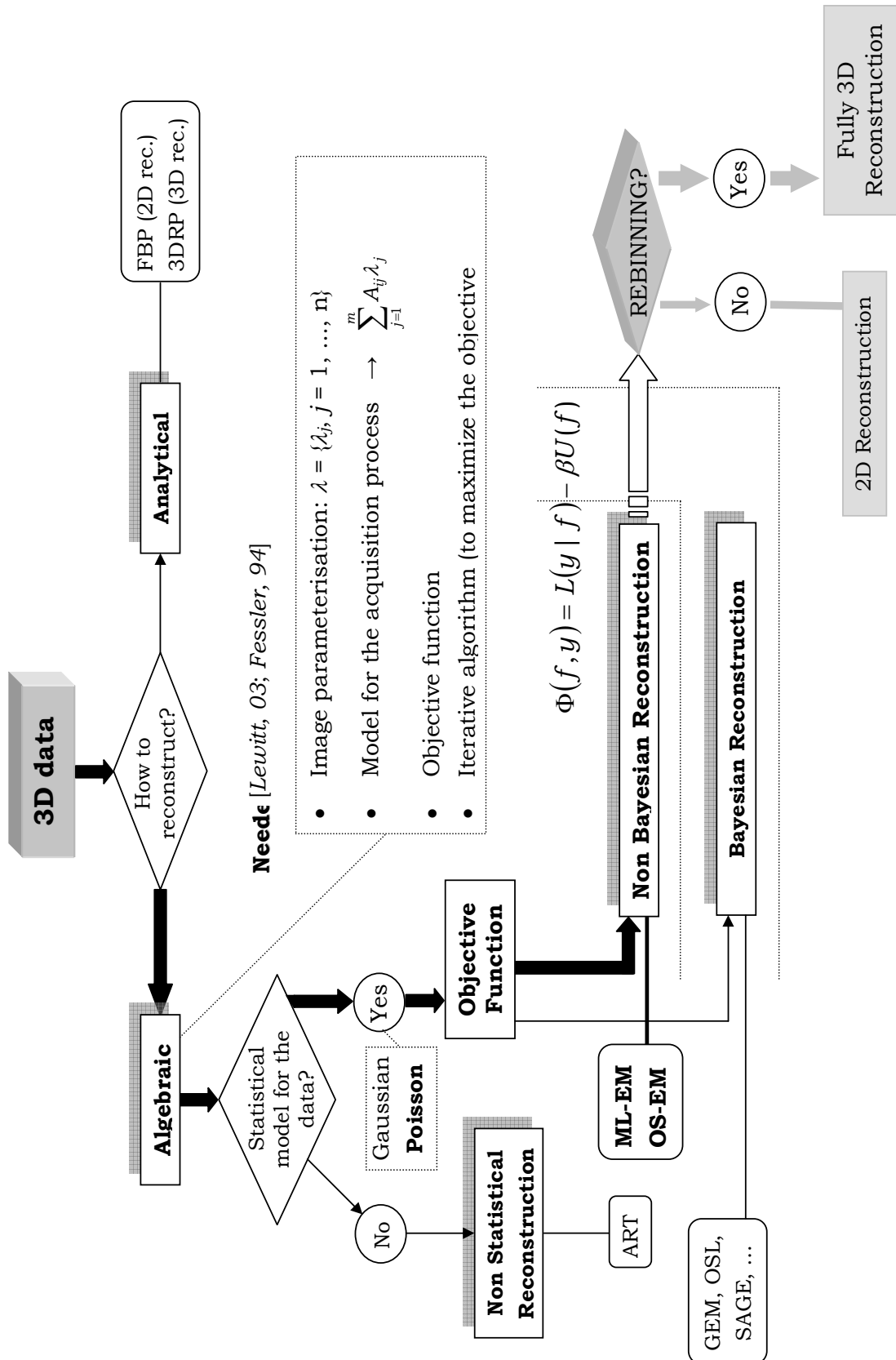
the likelihood function is:

$$L(\lambda \mid \underset{\sim}{x}) = \prod_{k=1}^n e^{-\lambda} \frac{\lambda^{x_k}}{x_k!}$$

and the log-likelihood:

$$\log L(\lambda \mid \underset{\sim}{x}) = \sum_{k=1}^n [-\lambda + x_k \log \lambda - \log(x_k!)]$$

A.6 – A Reconstruction Techniques Diagram



A.7 – The Non-Prewhitening Matched Filter (NPWMF) Numerical Observer

The ultimate goal of medical imaging is to produce good clinical images. However, a “good clinical image” is a very slippery concept since an image might be very useful for one specific task but not so relevant for another. Image quality can thus be defined in terms of how well desired information can be extracted from the image [Barrett, 93]. This means that the specific purpose images are to be used for by the observer is critical for the definition of what is a good clinical image.

Once the clinical task is perfectly established, a criterion is required to access image quality². This is where numerical observers may play an important role. In fact, if medical images are to be used for tumor detection, psychophysical studies with human observers are the accepted standard for assessing image quality [Gifford, 05]. Nevertheless, in practice, human observer studies can only be performed when involving just a relatively small number of images and/or test strategies. Otherwise, the task would be too much time consuming. The alternative is, therefore, the use of numerical observers – which are intended to replace the human observer entirely (or, in other cases, to increase the human performance) [Barrett, 93].

A numerical observer is a statistical decision model where a value for a statistical parameter, λ , is calculated from the available image and compared to a theoretical threshold. Considering the approach for calculating the value of such parameter, numerical observers can be grouped into two main groups: ideal (or Bayesian), and linear [Lartizien, 01] - the λ parameter associated with the statistical test is a linear function of the image. The former are used to set an upper bound to the performance of the observer [Barrett, 93; Lartizien, 01], while the latter are essentially adopted to mimic the performance of human observers [Lartizien, 01].

Linear observers are intrinsically different from Bayesian algorithms because they do not rely in the assumption of any probability distributions: they just use information about the first and second order moments drawn from the image [Lartizien, 01].

The Non-Prewhitening Matched Filter (NPWMF) observer is a linear observer used to predict the performance of human observers [Barrett, 93]. The test statistics (linear) on its basis is given by [Barrett, 93; Lartizien, 01; Kim, 04]:

$$\lambda_{NPW} = (\langle g_1 \rangle - \langle g_0 \rangle)^T g$$

where g is a sample image vector and $\langle g_k \rangle$ is an ensemble mean of classe k ($k = 1$ for signal-present and $k = 0$ for signal-absent). As referred, no detailed knowledge of the noise statistics is needed, and the observer forms the scalar product of a mask proportional to the signal profile to be detected $(\langle g_1 \rangle - \langle g_0 \rangle)$, and the image under test, g . It shows an optimal behavior for a signal whose position is known and constant over a background of Gaussian white noise [Lartizien, 01].

The corresponding signal-to-noise or detectability index for the NPWMF observer is given by [Lartizien, 01; Lartizien, 04]:

$$d_{NPW} = \frac{\langle \mathbf{s} \cdot \mathbf{g}_1 \rangle - \langle \mathbf{s} \cdot \mathbf{g}_0 \rangle}{\sqrt{\frac{1}{2} \cdot (\sigma^2(\mathbf{s} \cdot \mathbf{g}_1) + \sigma^2(\mathbf{s} \cdot \mathbf{g}_0))}} \quad \text{Eq. A7.1}$$

where \mathbf{g}_0 and \mathbf{g}_1 are the vectors of the image voxel values of the signal-absent (no lesion) and signal-present (with lesion) class, \mathbf{s} the vector of the expected signal profile, $\langle \rangle$ and $\sigma^2(\cdot)$ the ensemble average and variance for the ensemble.

² A definition of image quality must specify [Barrett, 02]: the task (what information is to be obtained from the image), the observer (how will information be extracted), and the object and image statistics (what sort of objects are to be observed and what is the measurement of noise).

A.8 – The EXACT HR+ PET scanner

The EXACT HR+ is a PET scanner developed by the CPS Innovations (Knoxville, USA). Its performance has been extensively studied and many detailed descriptions can be found in the literature, being either global descriptions or focusing some parts of the scanner. We can refer, for example: [Brix, 97], [Adam, 97] (which are among the first devoted to this scanner), [Ferreira, 01] or [Humm, 03] (where the values for some parameters are listed for different scanners). In addition, there are also the scanner manuals.

The purpose of this section is to underline some parameters which are fundamental for the reconstruction process and whose values are required by the OSEM algorithm.



Figure A8.1 - The ECAT EXACT HR+ scanner (from www.medical.siemens.com)

The HR+ scanner has a ring geometry: its structure consists in four detection rings (82.7 cm diameter), coupled along the axial direction. Each detector ring is made of 72 detector blocks, whose configuration is depicted in Fig. 1.6: 8×8 (radial \times axial) detection elements, referred as crystals³, resulting from the fine score of a BGO crystal. These individual matrix crystals, which are, in fact, pseudo-crystals, have the following dimensions: $4.39 \times 4.05 \times 30.0$ mm³, respectively in the transaxial, axial and radial direction. Each detector block is optically coupled to four circular PMTs (Fig. 1.6).

From the paragraph above, it becomes clear that each ring contains 576 crystals (72×8) and there are 32 (4×8) crystals along the scanner axis. Considering the cross planes, it is, therefore, possible to define 63 planes along the axial direction. These planes can be used in 2D or 3D mode, an option that is controlled by the use of retractable tungsten septa.

³ In practice, one detector block is assumed as a 8×8 matrix of crystals.

To perform transmission scans, the HR+ has three linear sources of ^{68}Ge , separated by 120° , that go round the patient at, about, 10 cm from the crystals and are removed to their shielding compartment whenever they are not being used [Ferreira, 01].

The timing window for the HR+ is 12 ns large and the default energy window is fixed between 350 keV and 650 keV. Slightly different values can be found in the literature in respect to the performance (resolution, sensitivity, scatter fraction, etc.) of this scanner. So, we refer here the work developed by Ferreira [Ferreira, 01], who has exhaustively investigated the performance of one of the S.H.F.J's. HR+ scanners according to the NEMA protocol. A more general comparison of the performance of many different scanners can be found in [Humm, 03].

The most important characteristics of the EXACT ECAT HR+ scanner, from the point of view of reconstruction, are resumed in Table A8.1. Other relevant values concerning the performance of this scanner can be found in [Humm, 03], for example.

Crystal	BGO
Patient port	56.2 cm
Number of structural rings (block rings)	4
Number of blocks per structural rings	72
Total number of blocks	288
Number of crystal per block	64 (8×8)
Number of detector rings ¹	32 (4×8)
Number of crystals per ring	576 (72×8)
Detector ring diameter	82.7 cm
Crystal dimensions (transaxial \times axial \times radial)	$4.39 \times 4.05 \times 30.0 \text{ mm}^3$
Transaxial FOV	58.3 cm
Axial FOV	15.5 cm
Number of slices	63
Default (maximum) axial sampling (plane separation)	2.425 (2.25) mm
Default sinogram dimensions (transaxial bins \times views)	288×144

Table A8.1 – Some relevant characteristics of the ECAT EXACT HR scanner (adapted from [Ferreira, 01] and [Humm, 03]).

At the level of the computer implementation, OSEM gets information about the scanner from two distinct files: an header (scanner_model.h), and a source file (scanner_model.c). Inside the header is defined a structure whose inner variables correspond to the scanner parameters important for the reconstruction stage:

```

typedef struct _ScannerModel {
    /* (...) */
    int dirPlanes;           /* number of direct planes */
    int defElements;         /* default number of elements */
    int defAngles;           /* default number of angles */
    int crystals_per_ring;   /* number of crystal */
    float crystalRad;        /* detector radius in mm */
    float planesep;          /* plane separation in mm */
    float binsize;           /* bin size (mm): spac. transaxial elem. */
    /* (...) */
    int defRingdiffmax;      /* default maximum ring difference */
    int defspan3d;           /* default span factor for 3D scan */
    /* (...) */
    int defbedverlap;        /* default amount of rings overlap for multibed
                             scans */
} ScannerModel;

```

As the name suggests, binsize is the size of each detection bin, along the transaxial direction, i.e., the data sampling for this direction. It reflects not only the double sampling (cf. section 1.3.2) but also the spatial resolution for the transaxial direction. The values for all of these variables are set from the definitions inside `scanner_model.c`:

```

ScannerModel _scanner_model = { /* (...) */ 32, 288, 144, 576, 412.5, 2.425,
2.247, /* (...) */ 22, 9, /* (...) */ 5 };

```

With this information, it is possible to design the Michelogram corresponding to a default acquisition with the HR+ scanner (Fig. A8.2).

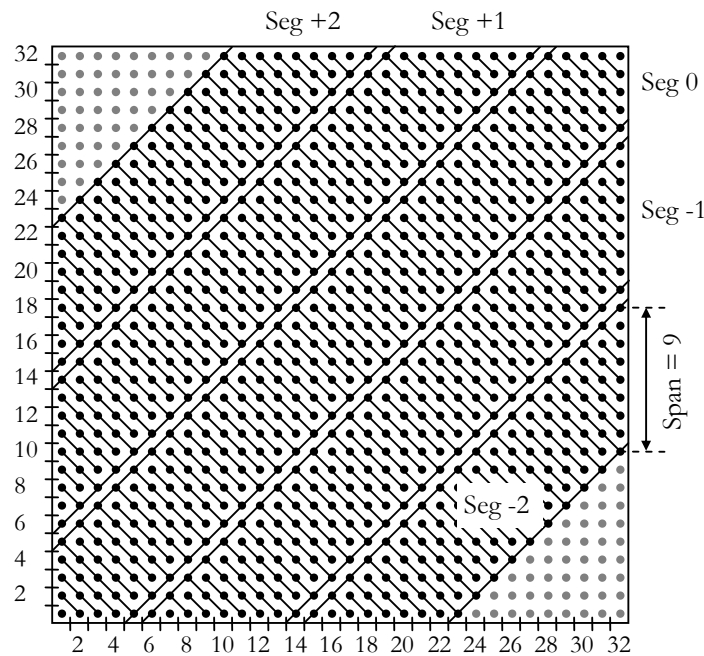


Figure A8.2 – The Michelogram corresponding to a default acquisition with the HR+ scanner.

For 32 rings, the use of cross planes along the axial direction results in 63 transaxial planes ($\theta = 0^\circ$). In addition, considering a maximum ring difference of 22, the planes represented by the grey dots in the Michelogram (upper left and right bottom corners) are discarded, i.e, data are not collected for LORs inside these planes. For the others (black dots in the Michelogram), data are stored using a span of 9 and, therefore, resulting in 5 segments (Fig A8.2): +2, -2 (containing data acquired along LORs corresponding to the larger values of θ); +1, -1, and 0 (made by the 63 transaxial planes).

For segments +1 and -1, the number of samples along the axial coordinate is 53, whereas for segments +2 and -2 there are data just for 35 axial samples.

The conditions referred in the paragraphs above correspond to a default acquisition. They were used for both the simulation and clinical studies carried on under the research work that is being described in the present thesis.

It is finally important to note that the HR+ scanner uses the ECAT7 file structure for the output and input of data. Within an ECAT7 sinogram, for example, data are labeled according to 5 parameters: `frame`, `plane`, `gate`, `data`, and `bed`. Therefore, sinograms acquired for different bed positions can be concatenated within a single file by setting properly the `bed` parameter. The same is true for different types of data - trues, randoms, or scattered - through the use of the `data` parameter (each type as a unique numerical code).

In the simulation work, the `gate` parameter played a very important role, since it was through its use that one could have stored in the same sinogram the trues (gate 1), the variance (gate 2), and the mean value (gate 3).

All the structure is described properly in a look-up table stored within the ECAT7 file.

LIST OF FIGURES

Figure 1.1	Typical whole-body FDG PET image, obtained from the HR+ scanner, for a patient with a body mass index equal to 23.7 and a dose of 338.5 MBq: a) coronal cross section; b) sagittal cross section	8
Figure 1.2	Schematic representation of a LOR within a transaxial plane	10
Figure 1.3	Physical aspects of positron-electron annihilation: positron range and non-colinearity ($\Delta\theta$)	11
Figure 1.4	A schematic representation of a block detector. a) Standard PMT arrangement in a standard block detector; b) Quadrant-sharing detector	19
Figure 1.5	2D and 3D data acquisition schemes for PET	20
Figure 1.6	The azimuthal angle and the radial sampling within a plane. (d – width of the FOV) ...	22
Figure 1.7	Three examples of Michelograms corresponding to three different acquisitions: a) 2D acquisition; b) 3D acquisition without mashing; c) 3D acquisition with “mash”. Axial location is along the bottom left to top right diagonal direction	23
Figure 1.8	a) Data acquisition within a transaxial plane ($\theta = 0^\circ$; $\tilde{r} = \tilde{r}_0$); b) Sinogram structure for this plane; c) The variation of the radial coordinates of three point sources as a function of the azimuthal angle	24
Figure 1.9	Coordinate system for sinogram parameterization. R is the ring radius, and L the scanner length	25
Figure 1.10	Link between projection coordinates (y) and sinogram coordinates (\tilde{r})	25
Figure 1.11	Different types of detected events in PET (prompts): a) true; b) multiple; c) single; d) random; e) scattered	26
Figure 1.12	The attenuation path length of two annihilation photons emitted from a point source inside the object. L is the total length of the LOR	30
Figure 1.13	Illustration of the arc effect in a transaxial plane of a ring tomograph. LOR separation, d , is smaller at the edge of the FOV ($d_1 < d_2$)	35
Figure 1.14	Parallax effect: the apparent width of the detector element (LOR) increases with the increasing of the radial offset	42
Figure 2.1	A single point source in the field-of-view of a PET scanner. $\overline{S_1}$ is the attenuation path of one annihilation photon; $\overline{S_2}$ is the attenuation path of the other photon	48
Figure 2.2	A general representation of the reconstruction strategies for 2D and 3D data	51
Figure 2.3	Sampling in the frequency domain	54
Figure 2.4	Orlov Spheres for different acquisition modes (2D and 3D) in a conventional cylindrical whole-body PET scanner	58
Figure 2.5	The 3DRP algorithm. The starting point is a 2D set of projections extracted from the acquired data	60
Figure 2.6	A schematic representation of the algebraic D-D model	61
Figure 2.7	Flow chart of an iterative image reconstruction technique	63

Figure 2.8	An example of a subset arrangement used by OSEM. Among all collected projections around the scanner (not represented), inside the same iteration for each new update is used just one subset (A, B or C). The order by which subsets are used is relevant and not necessarily A-B-C	82
Figure 2.9	The three different regions of the (k, n) space, in terms of rebinning approach: SSRB region, FORE region, and NC (Non-Consistency region)	95
Figure 2.10	SSRB: an illustration of the set of oblique LORs transformed into a single transaxial LOR	96
Figure 3.1	Schematics of the NEC definition	102
Figure 3.2	A schematic description of the OSEM core's interface	106
Figure 4.1	Axial (a), coronal (b) and sagittal views of the anthropomorphic phantom used for the simulations. Three tumors are visible and indicated by arrows: one the left lung, one in the right lung, and another in the liver	112
Figure 4.2	A flowchart of the file generation chain for each replica of the anthropomorphic phantom (reconstruction stage not included)	116
Figure 4.3	Radial profiles of the multiplicative (ANW) and NEC 3D weighing factors. Profiles are for the second bed position, axial plane 42 of segment 0 and azimuthal angle equal to 0, for one replica of the anthropomorphic phantom	119
Figure 4.4	The reciprocal values of the slice variance reduction factors, $K(z)$	120
Figure 4.5	A schematic representation of the five different weighted versions of OSEM used for reconstructing simulated data	121
Figure 4.6	Variation with number of iterations of the NPWMF index averaged over all tumors in the liver (top), lungs, and soft tissues (bottom row). Left column: $AR_{50\%}$; right column: AR_{low} . Post-reconstruction smoothing filter FWHM = 5.5 mm	127
Figure 4.7	Variation with the post-smoothing filter's FWHM of the NPWMF index averaged over all tumors in the liver (top), lungs, and soft tissues (bottom row), after 4 iterations (15 iterations of SPOSEM3D). Left column: $AR_{50\%}$; right column: AR_{low} ...	128
Figure 4.8	Coronal view of the reconstructed phantom. All the images were obtained after 4 iterations, except for SPOSEM3D (15 iterations). Image in the bottom right is the average image taking 25 reconstructions with ANWOSEM3D after 4 iterations. Post-smoothing filter: FWHM = 5.5mm; activity ratio: $AR_{50\%}$. Arrows point to the tumors visible on the average image	129
Figure 4.9	Variation with the number of iterations of the tumor-to-background contrast and noise averaged over all the tumors in the liver, lungs, and soft tissues. FWHM = 5.5 mm	131
Figure 4.10	Comparison between FORE+NECOSEM (4 it.) and FORE+AWOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm	132
Figure 4.11	Comparison between ANWOSEM3D (4 it.) and FORE+AWOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm	133
Figure 4.12	Comparison between ANWOSEM3D (4 it.) and FORE+NECOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm	134
Figure 4.13	Comparison between NECOSEM3D (4 it.) and FORE+NECOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm	134

Figure 4.14	Comparison between NECOSEM3D (4 it.) and FORE+NECOSEM (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm	135
Figure 4.15	Comparison between SPOSEM3D (15 it.) and ANWOSEM3D (4 it) NPWMF indices, for tumors in the liver, lungs, and soft tissues. $AR_{50\%}$, FWHM = 5.5 mm	135
Figure 5.1	A flowchart of the data pre-processing stage that uses dedicated routines available only in the scanner's proprietary computer. Outside the boxes, but close to them, is the name of the routine involved in each process	142
Figure 5.2	Projections obtained from the anthropomorphic phantom, for one be position, $\phi = 0^\circ$, and segment 0. a) $\langle t_i^{3D} \rangle$, b) $sm(t_i^{3D})$, c) $\langle t_i^{3D} \rangle - sm(t_i^{3D})$	146
Figure 5.3	A radial profile of the trues (gate 1), t_i^{3D} , the trues' mean values (gate 3), $\langle t_i^{3D} \rangle$, and the smoothed trues, $sm(t_i^{3D})$. Profiles drawn from projections shown in Figure 5.2 ...	147
Figure 5.4	A flowchart of the processing chain simulated data were subjected to in order to compare the exact data variance with the variance estimation process used for clinical data	148
Figure 5.5	A radial profile of the exact variance (gate 2) and $\langle t_i^{3D} \rangle + 2\langle r_i^{3D} \rangle$. Profiles drawn from projections at the chest level for segment 0 and $\phi = 0^\circ$	149
Figure 5.6	A radial profile of the exact variance (gate 2) after normalization and $N_i^2 \cdot A_i^2 (\langle t_i^{3D} \rangle + 2\langle r_i^{3D} \rangle)$. Profiles drawn from projections at the chest level for segment 0 and $\phi = 0^\circ$	150
Figure 5.7	A radial profile of the trues, t_i , smoothed trues, $\langle t_i \rangle$ and $\langle t_i \rangle + 2\langle r_i \rangle$. Data extracted from a transaxial plane crossing the chest of one of the patients, for segment 0 of a projection obtained for an azimuthal angle equal to zero	151
Figure 5.8	Radial profiles of the data after FORE. The projection data were drawn was obtained for one bed position of one of the clinical exams (azimuthal angle equal to 0° ; transaxial plane 32). Gate 1 (G1), trues; Gate 2 (G2), variance; Gate 3 (G3), mean value	153
Figure 5.9	Radial profile of the variance after fore, at the centre of the FOV. Projection obtained for $\phi = 0^\circ$	153
Figure 5.10	A flowchart of the data processing stages and NEC weighting factors creation. Close to each box is indicated the routine used for generating the corresponding file	154
Figure 5.11	A radial profile of the calculated 2D NEC weighting factors, after assemble bed, for one of the clinical exams (drawn form a projection obtained for $\phi = 0^\circ$)	156
Figure 5.12	Diagram of data input and output in the reconstruction stage, for the three algorithms: AWOSEM, NECOSEM, and ANWOSEM	157
Figure 5.13	Patient ID #1. B.M.I. = 16.2, injected dose = 384.8 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D	158
Figure 5.14	Patient ID #8. B.M.I. = 23.7, injected dose = 338.5 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D	159
Figure 5.15	Patient ID #17. B.M.I. = 30.9, injected dose = 407 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D	159

Figure 5.16	Patient ID #24. I.B.M. = 39.0, injected dose = 403.3 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D	160
Figure 5.17	Patient ID #25. B.M.I. = 44.6, injected dose = 414.3 MBq. Right : FORE+AWOSEM ; centre : FORE+NECOSEM ; left : ANWOSEM3D	161
Figure C1	The two different perspectives behind the developed performance analysis of reconstruction algorithms: P1, the 2D versus 3D perspective; P2, the complete versus non-complete restoring of Poisson data	164
Figure A1.1	The two coordinate systems. (x, y) is fixed; (x', y') is the rotating system and ϕ is azimuthal angle	171
Figure A1.2	A schematic three-steps description of the whole rotation in 3D. a) 1 st : renaming of the rotating system's axis. b) 2 nd : rotation around the z axis (scanner axis). c) 3 rd : rotation around the x'' axis	172
Figure A2.1	One example of a line in D , along which $f(x, y)$ is to be integrated	174
Figure A3.1	A geometrical interpretation of the backprojection operator	176
Figure A4.1	An illustration of the 2D Central Section Theorem	177
Figure A8.1	The ECAT EXACT HR+ scanner (from www.medical.siemens.com)	185
Figure A8.2	The Michelogram corresponding to a default acquisition with the HR+ scanner	187

LIST OF TABLES

Table 1.1	Some physical properties of positron-emitting nuclides used in PET (adapted from [Zanzonico, 04])	12
Table 1.2	Some properties of scintillator materials used for PET (adapted from [Humm, 03])	15
Table 3.1	The order by which the pairs of subsets and segments are handled, in order to grant OSEM balanced projections. This is for 8 subsets and 5 segments (-2, -1, 0, 1, 2 ordered as 3, 1, 0, 2, 4), inside a single iteration	108
Table 4.1	Lesion to tissue activity ratios	115
Table 4.2	Tumor-to-background detectability index, contrast and noise, averaged across all the tumors on each organ, for different reconstruction schemes and number of iterations. (FWHM = 5.5 mm)	131
Table 5.1	A description of the 25 scans used in the present study, ordered according to their body mass index. Records for patient #26 were not available. (BMI = Weight(kg) / Height(m) ²)	140
Table A8.1	Some relevant characteristics of the ECAT EXACT HR scanner (adapted from [Ferreira, 01] and [Humm, 03])	186

BIBLIOGRAPHY

- [**Adachi, 94**] Adachi S, Chiba M, Hirose T, et al., "Precise measurements of e^+e^- annihilation at rest into 4 photons and the search for exotic particles", *Physical Review A* 49 (5): 3201-3208 Part A May 1994.
- [**Adam, 97**] Adam LE, Zaers J, Ostertag H, et al., "Performance evaluation of the whole-body PET scanner ECAT EXACT HR+ following the IEC standard", *IEEE Trans. Nuc. Science* 44 (3): 1172-1179 Part 2 Jun 1997.
- [**Ahn, 03**] Ahn S, Fessler JA, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms", *IEEE Trans. Med. Imaging* 22 (5): 613-626 May 2003.
- [**Alenius, 97**] Alenius S, Ruotsalainen U, "Bayesian image reconstruction for emission tomography based on median root prior", *Eur. J. Nuc. Medicine* 24 (3): 258-265 Mar 1997.
- [**Alessio, 06**] Alessio A, Sauer K, Kinahan P, "Analytical reconstruction of deconvolved Fourier rebinned PET sinogram", *Phys. Med. Biology* 51 (1) 77-93 Jan 2006.
- [**Andersen, 89**] Andersen AH, "Algebraic reconstruction in CT from limited views", *IEEE Trans. Med. Imaging* 8 (1): 50-55 Mar 1989.
- [**Badawi, 96**] Badawi RD, Marsden PK, Cronin BF, et al., "Optimization of noise-equivalent count rates in 3D PET", *Phys. Med. Biology* 41 (9) 1755-1776 Sep 1996.
- [**Badawi, 99a**] Badawi RW, Miller MP, Bailey DL, et al., "Randoms variance reduction", *Phys. Med. Biology* 44 (4) 941-954 Apr 1999.
- [**Badawi, 99b**] Badawi RD, Marsden PK, "Developments in component-based normalization for 3D PET", *Phys. Med. Biology* 44 (2) 571-594 Feb 1999.
- [**Badawi, 00a**] Badawi RD, Kolhmyer SG, Harrison RL, et al., "The effect of camera geometry on singles flux, scatter fraction and trues and randoms sensitivity for cylindrical 3D PET – a simulation study", *IEEE Trans. Nuc. Science* 47 (3): 1228-1232 Part 3 Jun 2000.
- [**Badawi, 00b**] Badawi RD, Ferreira NC, Kohlmyer SG, et al., "A comparison of normalization effects on three whole-body cylindrical 3D PET systems", *Phys. Med. Biology* 45 (11) 3253-3266 Nov 2000.
- [**Baete, 04a**] Baete K, Nuyts J, Van Paesschen W, et al., "Anatomical-based FDG-PET reconstruction for the detection of hypo-metabolic regions in epilepsy", *IEEE Trans. Med. Imaging* 24(4): 510-519 Apr 2004.
- [**Baete, 04b**] Baete K, Nuyts J, Van Laere K, et al., "Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET", *Neuroimage* 23 (1): 305-317 Sep 2004.
- [**Bai, 03**] Bai C, Kinahan PE, Brasse D, et al., "An analytic study of the effects of attenuation on tumor detection in whole-body PET oncology imaging", *J. Nuc. Medicine* 44 (11) 1855-1861 Nov 2003.
- [**Bailey, 94**] Bailey DL, Meikle SR, "A convolution-subtraction scatter correction method for 3D PET", *Phys. Med. Biology* 39 (3) 411-424 Mar 1994.

-
- [**Bailey, 97**] Bailey DL, Meikle SR, Jones T, "Effective sensitivity in 3D PET: the impact of detector dead time on 3D system performance", *IEEE Trans. Nuc. Science* 44(3): 1180-1185 Part 2 Jun 1997.
- [**Bailey, 98a**] Bailey DL, "Transmission scanning in emission tomography", *Eur. J. Nuc. Medicine* 25 (7) 774-787 Jul 1998.
- [**Bailey, 98b**] Bailey DL, "Quantitative procedures in 3D PET", in *"The Theory and Practice of 3D PET"*, Eds. B. Bendriem and D. W. Townsend, Kluwer Academic Publishers, 1998.
- [**Bailey, 03a**] Bailey DL, "Data acquisition and performance characterization in PET", in *"Positron Emission Tomography: Basic Science and Clinical Practice"*, Eds. P. E. Valk, D. L. Bailey, D. W. Townsend and M. N. Maisey, Springer, 2003.
- [**Bailey, 03b**] Bailey DL, Karp JS, Surti S, "Physics and instrumentation in PET", in *"Positron Emission Tomography: Basic Science and Clinical Practice"*, Eds. P. E. Valk, D. L. Bailey, D. W. Townsend and M. N. Maisey, Springer, 2003.
- [**Balcerzyk, 00**] Balcerzyk M, Moszynski M, Kapusta M, et al., "YSO, LSO, GSO and LGSO. A study of energy resolution and nonproportionality", *IEEE Trans. Nuc. Science* 47(4): 1319-1323 Part 1 Aug 2000.
- [**Barrett, 93**] Barrett HH, Yao J, Rolland JP, et al., "Model observers for assessment of image quality", *Proc. National Academy of Sciences* 90 (21) 9758-9765 Nov 1 1993.
- [**Barrett, 94**] Barrett HH, Wilson DW, Tsui BMW, "Noise properties of the EM algorithm. 1. Theory", *Phys. Med. Biology* 39 (5) 833-846 May 1994.
- [**Barrett, 02**] Barrett HH, "Task-based assessment of image quality", *NSS-MIC Short Course, Norfolk, November 2002*.
- [**Barrett, 04**] Barrett H, Meyers K, "Foundations of Image Science", *John Wiley & Sons, 2004*.
- [**Bergmann, 05**] Bergmann H, Dobrozemsky G, Minear G, et al., "An inter-laboratory comparison study of image quality of PET scanners using the NEMA NU 2-2001 procedure for assessment of image quality", *Phys. Med. Biology* 50 (10) 2193-2207 May 21 2005.
- [**Berko, 56**] Berko S, Hereford FL, "Experimental studies of positron interactions in solids and liquids", *Reviews of Modern Physics*. 28 (3): 299-307 1956.
- [**Bevington, 94**] Bevington PR, Robinson DK, "Data Reduction and Error Analysis for the Physical Sciences", Second Edition, *McGraw-Hill International Editions, 1994*.
- [**Beylkin, 87**] Beylkin G, "Discrete Radon transform", *IEEE Trans. Acoustic, Speech and Signal Process.* 35 (2): 162-172 Feb 1987.
- [**Bowsher, 96**] Bowsher J, Johnson V, Turkington T, et al., "Bayesian reconstruction and use of anatomical a priori information for emission tomography", *IEEE Trans. Med. Imaging* 15 (5): 673-686 Oct 1996.
- [**Brasse, 05**] Brasse D, Kinahan PE, Lartizien C, et al., "Correction methods for random coincidences in fully 3D whole-body PET: impact on data and image quality", *J. Nuc. Medicine* 46(5) 859-867 May 2005.
- [**Brix, 97**] Brix G, Zaers J, Adam LE, et al., "Performance evaluation of a whole-body PET scanner using the NEMA protocol", *J. Nuc. Medicine* 38(10) 1614-1623 Oct 1997.

-
- [**Brooks, 76**] Brooks RA, Di Chiro G, "Principles of computer-assisted tomography (CAT) in radiographic and radioisotopic imaging", *Phys. Med. Biology* 21 (5) 689-732 1976.
- [**Browne, 96**] Browne J, De Piero AR, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomograph", *IEEE Trans. Med. Imaging* 15(5): 687-699 Oct 1996.
- [**Bruyant, 02**] Bruyant PP, "Analytic and iterative reconstruction algorithms in SPECT", *J. Nuc. Medicine* 43 (10) 1342-1358 Oct 2002.
- [**Buonocore, 81**] Buonocore MH, Brody WR, Macovski A, "A natural pixel decomposition for two-dimensional image reconstruction", *IEEE Trans. Biomed. Engineering* 28 (2): 69-78 1981.
- [**Byrne, 96**] Byrne CL, "Block-iterative methods for image reconstruction from projections", *IEEE Trans. Image Process.* 5 (5): 792-794 May 1996.
- [**Byrne, 98**] Byrne CL, "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods", *IEEE Trans. Image Process.* 7 (1): 100-109 Jan 1998.
- [**Casella, 90**] Casella G, Berger RL, "Statistical Inference", *Duxbury Press*, 1990.
- [**Casey, 86**] Casey ME, Hoffman EJ, "Quantitation in Positron Emission Computed Tomography: 7. A technique to reduce noise in accidental coincidence measurements and coincidence efficiency calibration", *J. Computer Assisted Tomography* 10 (5): 845-850 1986.
- [**Casey, 95**] Casey ME, Gadagkar H, Newport D, "A component based method for normalization in volume PET", *Proc. 3rd Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Aix-les-Bains)*, 67-71 1995.
- [**Chen, 91**] Chen CM, Lee SY, Cho ZH, "Parallelization of the EM algorithm for 3-D PET image reconstruction", *IEEE Trans. Med. Imag.* 10 (4): 513-522 Dec 1991.
- [**Cherry, 03**] Cherry S, Sorenson J, Phelps M, "Physics in Nuclear Medicine", 3rd Edition, *Saunders*, 2003.
- [**Chiba, 98**] Chiba M, Hamatsu R, Hirose T, et al., "Four and five photon decay from positronium and the lifetime", *Nucl. Inst. Methods in Physics Research Section B* 143 (1-2): 121-126 Aug 1998.
- [**Choi, 98**] Choi HK, Munson DC, "Direct-Fourier reconstruction in tomography and synthetic aperture radar", *Intern. Journal of Imaging Systems and Technology* 9 (1): 1-13 1998.
- [**Christensen, 95**] Christensen NL, Hammer BE, Heil B, et al., "Positron emission tomography within a magnetic field using photomultiplier tubes and lightguides", *Phys. Med. Biology* 40 (4) 691-697 Apr 1995.
- [**Colsher, 80**] Colsher JG, "Fully tree-dimensional positron emission tomography", *Phys. Med. Biology* 25 (1): 103-115 1980.
- [**Comtat, 98**] Comtat C, Kinahan PE, Defrise C, et al., "Fast reconstruction of 3-D PET data with accurate statistical modeling", *IEEE Trans. Nuc. Science* 45(3): 1083-1089 Part 2 Jun 1998.
- [**Comtat, 99**] Comtat C, Kinahan P, Defrise M, et al., "Simulating whole-body PET scanning with rapid analytical methods", *Nucl. Sci. Symp. Conf. Rec., Seattle, USA, Vol. 3*, 1260-1264, 24-30 Oct 1999.
- [**Comtat, 02**] Comtat C, Kinahan PE, Fessler J, et al., "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels", *Phys. Med. Biology* 47 (1) 1-20 Jan 2002.

-
- [**Comtat, 04**] Comtat C, Bataille F, Michel C, et al., "OSEM-3D reconstruction strategies for the ECAT HRRIT", *Nucl. Sci. Symp. Conf. Rec., Rome, Italy, Vol. 6, 3492-3496, 16-22 Oct 2004*.
- [**Cormack, 63**] Cormack AM, "Representation of a function by its line integrals, with some radiological applications", *Journal of Applied Physics* 34 (9): 2722-2727 1963.
- [**CruzRivera, 95**] CruzRivera JL, DiBella EVR, Wills DS, et al., "Parallelized formulation of the maximum likelihood-expectation maximization algorithm for fine-grain message-passing architecture", *IEEE Trans. Med. Imaging* 14 (4): 758-762 Dec 1995.
- [**Daube-Witherspoon, 87**] Daube-Witherspoon ME, Muehllehner G, "Treatment of axial data in three-dimensional PET", *J. Nuc. Medicine* 28 (11) 1717-1724 Nov 1987.
- [**Daube-Witherspoon, 01**] Daube-Witherspoon ME, Matej S, Karp JS, et al., "Application of the rao action maximum likelihood algorithm with spherical basis functions to clinical PET imaging", *IEEE Trans. Nuc. Science* 48 (1): 24-30 Feb 2001.
- [**Daube-Witherspoon, 02**] Daube-Witherspoon ME, Karp JS, Casey ME, et al., "PET performance measurements using the NEMA NU 2-2001 standard", *J. Nuc. Medicine* 43 (10) 1398-1409 Oct 2002.
- [**Dawson, 06**] Dawson P, "Functional imaging in CT", *Eur J. Radiology* 60 (3) 331-340 Dec 2006.
- [**Deans, 83**] Deans SR, "The Radon Transform and Some of its Applications", *John Wiley & Sons, 1983*.
- [**DeBenedetti, 54**] DeBenedetti S, Siegel RT, "The three-photon annihilation of positrons and electrons", *Physical Review* 94 (4): 955-959 May 1954.
- [**Defrise, 95a**] Defrise M, Clack R, Townsend DW, "Image reconstruction from truncated, two-dimensional, parallel projections", *Inverse Problems* 11 (2): 287-313 Apr 1995.
- [**Defrise, 95b**] Defrise M, "A factorization method for the 3D x-ray transform", *Inverse Problems* 11 (5): 983-994 Oct 1995.
- [**Defrise, 97**] Defrise M, Kinahan PE, Townsend DW, et al., "Exact and approximate rebinning algorithms for 3-D PET data", *IEEE Trans. Med. Imaging* 16 (2): 145-158 Apr 1997.
- [**Defrise, 98**] Defrise M, Kinahan P, "Data acquisition and image reconstruction for 3D PET", in *"The Theory and Practice of 3D PET"*, Eds. B. Bendriem and D. W. Townsend, *Kluwer Academic Publishers, 1998*.
- [**Defrise, 99**] Defrise M, Liu XA, "A fast rebinning algorithm for 3D positron emission tomography using John's equation", *Inverse Problems* 15 (4): 1047-1065 Aug 1999.
- [**Defrise, 01**] Defrise M, "A short reader's guide to 3D tomographic reconstruction", *Computerized Medical Imaging and Graphics* 25 (2): 113-116 Mar-Apr 2001.
- [**Defrise, 02**] Defrise M, "Image reconstruction from parallel projections", *NSS-MIC Short Course, Norfolk, November 2002*.
- [**Defrise, 03**] Defrise M, Kinahan PE, Michel C, "Image reconstruction algorithms in PET", in *"Positron Emission Tomography: Basic Science and Clinical Practice"*, Eds. P. E. Valk, D. L. Bailey, D. W. Townsend and M. N. Maisey, *Springer, 2003*.
- [**DeMan, 04**] De Man B, Basu S, "Distance-driven projection and backprojection in three dimensions", *Phys. Med. Biology* 49 (11) 2463-2475 Jun 2004.

-
- [**Demoment, 89**] Demoment G, “Image reconstruction and restoration: overview of common estimation structures and problems”, *IEEE Trans. Acoustic, Speech and Signal Process.* 37 (12): 2024-2036 Dec 1989.
- [**Dempster, 77**] Dempster AP, Laird NM, Rubin DB, “Maximum likelihood from incomplete data via de EM algorithm”, *J. Royal Statistical Society Series B-Methodological* 39 (1): 1-38 1977.
- [**DePierro, 93**] DePierro AR, “On the relation between the ISRA and the EM algorithm for positron emission tomography”, *IEEE Trans. Med. Imaging* 12 (2): 328-333 Jun 1993.
- [**DePiero, 01**] De Piero AR, Yamagishi MEB, “Fast EM-like methods for maximum a posteriori estimates in emission tomography”, *IEEE Trans. Med. Imaging* 20 (4): 280-288 Apr 2001.
- [**Derenzo, 86**] Derenzo SE, “Mathematical removal of positron range blurring in high resolution tomography”, *IEEE Trans. Nuc. Science* 33 (1): 565-569 Feb 1986.
- [**Farquhar, 98**] Farquhar T, Chatziioannou A, Chinn G, et al., “An investigation of filter choice for filtered back-projection reconstruction in PET”, *IEEE Trans. Nuc. Science* 45 (3): 1133-1137 Part 2 Jun 1998.
- [**Fahey, 02**] Fahey FH, “Data acquisition in PET imaging”, *J. Nuc. Med. Technology* 30 (2) 39-49 Jun 2002.
- [**Ferreira, 98**] Ferreira NC, Trébossen R, Bendriem B, “Assessment of 3-D PET quantitation: influence of out of the field of view radioactive sources and of attenuating media”, *IEEE Trans. Nuc. Science* 45 (3): 1670-1675 Part 4 Jun 1998.
- [**Ferreira, 01**] Ferreira NC, “Contribuição para a quantificação em Tomografia por Emissão de Positrões no modo 3D”, *Dissertação de doutoramento apresentada à Fac. de Medicina da Universidade de Coimbra*, 2001.
- [**Ferreira, 02**] Ferreira NC, Trébossen R, Lartizien C, et al., “A hybrid scatter correction for 3D PET based on an estimation of the distribution of unscattered coincidences: implementation on the EACT EXACT HR+”, *Phys. Med. Biology* 47 (9): 1555-1571 May 7 2002.
- [**Ferrell, 56**] Ferrell RA, “Theory of positron annihilation in solids”, *Reviews of Modern Physics* 28 (3): 308-337 Jul 1956.
- [**Fessler, 94**] Fessler JA, “Penalized weighted least-squares image reconstruction for positron emission tomography”, *IEEE Trans. Med. Imaging* 13 (2): 290-300 Jun 1994.
- [**Fessler, 95**] Fessler JA, Hero AO, “Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms”, *IEEE Trans. Image Process.* 4 (10): 1417-1429 Oct 1995.
- [**Fessler, 96**] Fessler JA, Rogers WL, “Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs”, *IEEE Trans. Image Process.* 5 (9): 1346-1358 Sep 1996.
- [**Fessler, 02**] Fessler JA, “Statistical methods for image reconstruction”, *NSS-MIC Short Course, Norfolk, November 2002*.
- [**Fishburn, 97**] Fishburn P, Schwander P, Shepp L, et al., “The discrete Radon transform and its approximate inversion via linear programming”, *Discrete Applied Math.* 75 (1): 39-61 May 1997.
- [**Flyckt, 02**] Flyckt S-O, Marmonier C, “Photomultiplier Tubes – Principles and Applications”, *Photonis Imaging Sensors* 2002.

-
- [**Gifford, 05**] Gifford HC, King MA, Pretorius PE, et al., "A comparison of human and model observers in multislice LROC studies", *IEEE Trans. Med. Imaging* 24 (2): 160-169 Feb 2005.
- [**Gindi, 93**] Gindi G, Lee M, Rangarajan A, et al., "Bayesian reconstruction of functional images using anatomical information as priors", *IEEE Trans. Med. Imaging* 12 (4): 670-680 Dec 1993.
- [**Green, 90**] Green PJ, "Bayesian reconstructions from emission tomography data using a modified EM algorithm", *IEEE Trans. Med. Imaging* 9 (1): 84-93 Mar 1990.
- [**Gottlieb, 00**] Gottlieb D, Gustafsson B, Forssén P, "On the direct Fourier method for computer tomography", *IEEE Trans. Med. Imaging* 19 (3): 223-232 Mar 2000.
- [**Gordon, 74**] Gordon R, "A tutorial on ART", *IEEE Trans. Nuc. Science NS21* (3): 78-93 1974.
- [**Gordon, 94**] Gordon R, Farrell R, Daley K, et al., "Solid state tritium detector for biomedical applications", *IEEE Trans. Nuc. Science* 41 (4): 1494-1499 Part 1 Aug 1994.
- [**Grootoink, 96**] Grootoink S, Spinks TJ, Sashin D, et al., "Correction for scatter in 3D brain PET using a dual energy window method", *Phys. Med. Biology* 41 (12): 2752-2774 Dec 1996.
- [**Hasegawa, 00**] Hasegawa T, Michel C, Kawashima K, et al., "A study of external end-shields for PET", *IEEE Trans. Nuc. Science* 47 (3): 1099-1103 Part 3 Jun 2000.
- [**Hebert, 89**] Hebert T, Leahy R, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors", *IEEE Trans. Med. Imaging* 8 (2): 194-202 Jun 1989.
- [**Hebert, 90a**] Hebert TJ, "Statistical stopping criteria for iterative maximum likelihood reconstruction of emission images", *Phys. Med. Biology* 35 (9): 1221-1232 Sep 1990.
- [**Hebert, 90b**] Hebert TJ, Leahy R, "Fast methods for including attenuation in the EM algorithm", *IEEE Trans. Nuc. Science* 37 (2): 754-768 Part 1 Apr 1990.
- [**Herman, 73**] Herman GT, Lent A, Rowland SW, "ART: mathematics and applications", *Journal of Theoretical Biology* 42(1): 1-32 1973.
- [**Herman, 80**] Herman GT, "Image Reconstruction from Projections – The Fundamentals of Computerized Tomography", *Academic Press*, 1980.
- [**Herman, 91**] Herman GT, Odhner D, "Performance evaluation of an iterative image reconstruction algorithm for positron emission tomography", *IEEE Trans. Med. Imaging* 12 (3): 336-346 Sep 1991.
- [**Herman, 93**] Herman GT, Meyer LB, "Algebraic reconstruction techniques can be made computationally efficient", *IEEE Trans. Med. Imaging* 12 (3): 600-609 Sep 1993.
- [**Hoffman, 81**] Hoffman EJ, Huang S-C, Phelps ME, et al., "Quantitation in Positron Emission Computed Tomography: 4. Effects of accidental coincidences", *J. Computer Assisted Tomography* 5 (3): 391-400 Jun 1981.
- [**Holte, 90**] Holte S, Schmidlin P, Lindén A, et al., "Iterative image reconstruction for positron emission tomography: a study of convergence and quantitation problems", *IEEE Trans. Nuc. Science* 37 (2): 629-635 Part 1 Apr 1990.
- [**Hsiao, 03**] Herman IT, Rangarajan A, Gindi G, "A new convex edge-preserving median prior with applications to tomography", *IEEE Trans. Med. Imaging* 22 (5): 580-585 May 2003.

- [**Hsiao, 04**] Hsiao IT, Rangarajan A, Khurd P, et al., "An accelerated convergent ordered subsets algorithm for emission tomography", *Phys. Med. Biology* 49 (11): 2145-2156 Nov 2004.
- [**Hsieh, 96**] Hsieh YL, Gullberg GT, Zeng GL, et al., "Image reconstruction using a generalized natural pixel basis", *IEEE Trans. Nuc. Science* 43 (4): 2306-2319 Part 1 Aug 1996.
- [**Huang, 99**] Huang S-C, "Image oscillation reduction and convergence acceleration for OS-EM reconstruction", *IEEE Trans. Nuc. Science* 44 (3): 603-607 Part 2 Jun 1999.
- [**Hudson, 94**] Hudson HM, Larkin RS, "Accelerated image reconstruction using ordered subsets of projection data", *IEEE Trans. Med. Imaging* 13 (4): 601-609 Dec 1994.
- [**Humm, 03**] Humm JL, Rosenfeld A, Del Guerra A, "From PET detectors to PET scanners", *Eur. J. Nuc. Medicine* 30 (11): 1574-1597 Nov 2003.
- [**Hutton, 97**] Hutton BF, Hudson HM, Beekman FJ, "A clinical perspective of accelerated statistical reconstruction", *Eur. J. Nuc. Medicine* 24 (7): 797-808 Jul 1997.
- [**Hwang, 06**] Hwang DS, Zeng GL, "Convergence study of an accelerated ML-EM algorithm using bigger step size", *Phys. Med. Biology* 51 (1): 237-252 Jan 21 2006.
- [**Jackson, 91**] Jackson JI, Meyer CH, Nishimura DG, et al., "Selection of a convolution function for Fourier inversion using gridding", *IEEE Trans. Med. Imaging* 10 (3): 473-478 Sep 1991.
- [**Johnson, 95**] Johnson C, Yan Y, Carson R, et al., "A system for the 3D reconstruction of retracted-septa PET data using the EM algorithm", *IEEE Trans. Nuc. Science* 42 (4): 1223-1227 Part 1 Aug 1995.
- [**Jones, 99**] Jones H, Mitra G, Parkinson D, et al., "A parallel implementation of the maximum likelihood method in positron emission tomography image reconstruction", *Computational Statistics & Data Analysis* 31 (4): 417-439 Oct 28 1999.
- [**Jones, 03a**] Jones T, "Historical development of functional in vivo studies using positron-emitting tracers", in *"Positron Emission Tomography: Basic Science and Clinical Practice"*, Eds. P. E. Valk, D. L. Bailey, D. W. Townsend and M. N. Maisey, Springer, 2003.
- [**Jones, 03b**] Jones JP, Jones WF, Kehren F, et al., "SPMD cluster-based parallel 3-D OSEM", *IEEE Trans. Nuc. Science* 50 (5): 1498-1502 Part 2 Oct 2003.
- [**Kacperski, 04**] Kacperski K, Spyrou NM, Smith FA, "Three-gamma annihilation imaging in positron emission tomography", *IEEE Trans. Med. Imaging* 23 (4): 525-529 Apr 2004.
- [**Kak, 88**] Kak AC, Slaney M, "Principles of Computerized Tomographic Imaging", *IEEE Press*, 1988.
- [**Kaufman, 87**] Kaufman L, "Implementing and accelerating the EM algorithm for positron emission tomography", *IEEE Trans. Med. Imaging* 6 (1): 37-51 Mar 1987.
- [**Kaufman, 93**] Kaufman L, "Maximum likelihood, least squares, and penalized least squares for PET", *IEEE Trans. Med. Imaging* 12 (2): 200-214 Jun 1993.
- [**Kazantsev, 00**] Kazantsev IG, Van de Walle R, Lemahieu I, "Ridge functions, natural pixels and minimal norm reconstruction", *IEEE Trans. Nuc. Science* 47 (3): 1118-1122 Part 3 Jun 2000.

-
- [**Kim, 04**] Kim Js, Kinahan PE, Lartizien C, et al., “A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging”, *IEEE Trans. Nuc. Science* 51 (1): 34-40 Part 1 Feb 2004.
- [**Kinahan, 89**] Kinahan PE, Rogers JG, “Analytic 3D image reconstruction using all detected events”, *IEEE Trans. Nuc. Science* 36 (1): 964-968 Part 1 Feb 1989.
- [**Kinahan, 03**] Kinahan PE, Hasegawa BH, Beyer T, “X-ray-based attenuation correction for Positron Emission Tomography/Computed Tomography Scanners”, *Sem. Nuc. Medicine* 33(3): 166-179 Jul 2003.
- [**Knoll, 89**] Knoll GF, “Radiation Detection and Measurement”, 2nd Edition, John Wiley & Sons, 1989.
- [**Kuczmarski, 97**] Kuczmarski RJ, Carroll MD, Flegal KM, et al., “Varying body mass index cutoff points to describe overweight prevalence among US adults: NHANES III (1988 to 1994)”, *Obesity Research* 6(6): 542-548 Nov 1997.
- [**Lange, 84**] Lange K, Carson R, “EM reconstruction algorithms for emission and transmission tomography”, *J. Computer Assisted Tomography* 8(2): 306–316, 1984.
- [**Lange, 90**] Lange K, “Convergence of EM image reconstruction algorithms with Gibbs smoothing”, *IEEE Trans. Med. Imaging* 9 (4): 439-446 Dec 1990.
- [**Leahy, 00a**] Leahy RM, Qi J, “Statistical approaches in quantitative positron emission tomography”, *Statistics and Computing* 10 (2): 147-165 Apr 2000.
- [**Lartizien, 00**] Lartizien C, Kinahan P, Comtat C, et al., “A tumor detection observer study comparing 2D versus 3D whole-body PET imaging protocols”, *Nucl. Sci. Symp. Conf. Rec., Lyon, France, Vol. 3*, 81-86, 15-20 Oct 2000.
- [**Lartizien, 01**] Lartizien C, “Optimisation des protocoles d’imagerie corps entier en tomographie par émission de positons pour la détection des hyperfixations du ¹⁸F-FDG en oncologie”, *Thèse pour obtenir le grade de Docteur de l’Université Paris XI*, 2001.
- [**Lartizien, 03**] Lartizien C, Kinahan PE, Swensson R, et al., “Evaluating image reconstruction methods for tumor detection in 3-Dimensional whole-body PET oncology imaging”, *J. Nuc. Medicine* 44 (2) 276-290 Feb 2003.
- [**Lartizien, 04**] Lartizien C, Kinahan PE, Comtat C, “Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography”, *Academic Radiology* 11 (6): 637-648 Jun 2004.
- [**Leahy, 00b**] Leahy R, Byrne C, “Recent developments in iterative image reconstruction for PET and SPECT”, *IEEE Trans. Med. Imaging* 19 (4): 257-260 Apr 2000.
- [**Leo, 94**] Leo WR, “Techniques for Nuclear and Particle Physics: A How-To Approach”, 2nd Edition, Springer, 1994.
- [**Levitan, 87**] Levitan E, Herman GT, “A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography”, *IEEE Trans. Med. Imaging* 6 (3): 185-192 Sep 1987.

-
- [**Levin, 99**] Levin CS, Hoffman EJ, "Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution", *Phys. Med. Biology* 44 (3): 781-799 Mar 1999.
- [**Lewitt, 86**] Lewitt RM, Muehllehner G, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum-likelihood estimation", *IEEE Trans. Med. Imaging* 5(1): 16-22 Mar 1986.
- [**Lewitt, 92**] Lewitt RM, "Alternatives to voxels for image representation in iterative reconstruction algorithms", *Phys. Med. Biology* 37 (3): 705-716 Mar 1992.
- [**Lewitt, 03**] Lewitt R, Matej S, "Overview of methods for image reconstruction from projections in emission computed tomography", *Proceedings of the IEEE* 91 (10): 1588-1611 Oct 2003.
- [**Liow, 91**] Liow J, Strother S, "Practical tradeoffs between noise, quantitation, and number of iterations for maximum likelihood-based reconstructions", *IEEE Trans. Med. Imaging* 10 (4): 563-571 Dec 1991.
- [**Liow, 93**] Liow JS, Strother SC, "The convergence of object dependent resolution in maximum likelihood based tomographic image reconstruction", *Phys Med. Biology* 37 (3): 55-70 Jan 1993.
- [**Lipinski, 97**] Lipinski B, Herzog H, Kops ER, et al., "Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information", *IEEE Trans. Med. Imaging* 16 (2): 129-136 Apr 1997.
- [**Liu, 99**] Liu X, Defrise M, Michel C, et al., "Exact rebinning methods for three-dimensional PET", *IEEE Trans. Med. Imaging* 18 (8): 657-664 Aug 1999.
- [**Liu, 01**] Liu X, Comtat C, Kinahan P, et al., "Comparison of 3-D reconstruction with 3D-OSEM and with FORE+OSEM for PET", *IEEE Trans. Med. Imaging* 20 (8): 804-814 Aug 2001.
- [**Llacer, 89**] Llacer J, Veklerov E, "Feasible images and practical stopping rules for iterative algorithms in emission tomography", *IEEE Trans. Med. Imaging* 8 (2): 186-193 Jun 1989.
- [**Llacer, 93**] Llacer J, Veklerov E, Coakley K, et al., "Statistical analysis of maximum likelihood estimator images of human brain FDG PET studies", *IEEE Trans. Med. Imaging* 12 (2): 215-231 Jun 1993.
- [**Lucignani, 06**] Lucignani G, "Time-of-flight and PET/MRI: recurrent dreams or actual realities?", *Eur. J. Nuc. Medicine* 33 (8): 969-971 Aug 2006.
- [**Lundqvist, 98**] Lundqvist H, Lubberink M, Tolmachev V, "Positron emission tomography", *Eur. Journ. Physics* 19 (2): 537-552 1998.
- [**Matej, 90**] Matej S, Bajla I, "A high-speed reconstruction from projections using direct Fourier method with optimized parameters – an experimental analysis", *IEEE Trans. Med. Imaging* 9 (4): 421-429 Dec 1990.
- [**Matej, 96**] Matej S, Lewitt RM, "Practical considerations for 3-D image reconstruction using spherically symmetric volume elements", *IEEE Trans. Med. Imaging* 15 (1): 68-78 Feb 1996.
- [**Matej, 98**] Matej S, Karp JS, Lewitt RM, et al., "Performance of the Fourier rebinning algorithm for PET with large acceptance angles", *Phys Med. Biology* 43 (4): 787-795 Apr 1998.

-
- [**McCready, 00**] McCready VR, "Milestones in nuclear medicine", *Eur. J. Nuc. Medicine* 27 (1): S49-S79 Suppl. S Jan 2000.
- [**Meikle, 94**] Meikle SR, Hutton BF, Bailey D, et al., "Accelerated EM reconstruction in total-body PET: potential for improving tumour detectability", *Phys. Med. Biology* 39 (10): 1689-1704 Oct 1994.
- [**Meikle, 03**] Meikle SR, Badawi RD, "Quantitative techniques in PET", in *"Positron Emission Tomography: Basic Science and Clinical Practice"*, Eds. P. E. Valk, D. Bailey, D. W. Townsend and M. Maisey, Springer, 2003.
- [**Melcher, 00**] Melcher CL, "Scintillation crystals for PET", *J. Nuc. Medicine* 41 (6): 1051-1055 Jun 2000.
- [**Michel, 99**] Michel C, Liu X, Sanabria S, et al., "Weighted schemes applied to 3D-OSEM reconstruction in PET", *Nucl. Sci. Symp. Conf. Rec., Seattle, USA, Vol. 3, 1152-1157, 24-30 Oct 1999*.
- [**Moisan, 97**] Moisan C, Vozza D, Loope M, "Simulating the performances of an LSO based position encoding detector for PET", *IEEE Trans. Nuc. Science* 44 (6): 2450-2458 Part 2 Dec 1997.
- [**Mueller, 97**] Mueller K, Yagel R, Cornhill J, "The weighted distance scheme: a globally optimizing projection ordering method for ART", *IEEE Trans. Med. Imaging* 16 (2): 223-230 Apr 1997.
- [**Mumcuoğlu, 94**] Mumcuoğlu EU, Leahy R, Cherry SR, et al., "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images", *IEEE Trans. Med. Imaging* 13 (4): 687-701 Dec 1994.
- [**Mumcuoğlu, 96**] Mumcuoğlu EU, Leahy R, Cherry SR, "Bayesian reconstruction of PET images: methodology and performance analysis", *Phys. Med. Biology* 41 (9): 1777-1807 Sep 1996.
- [**Natterer, 01**] Natterer F, Wübbeling F, "Mathematical Methods in Image Reconstruction", *SIAM – Monographs on Mathematical Modeling and Computation*, 2001.
- [**Nutt, 02**] Nutt R, "The history of Positron Emission Tomography", *Molecular Imaging and Biology* 4 (1): 11-26 Jan - Feb 2002.
- [**Nuyts, 01**] Nuyts J, Michel C, Dupont P, "Maximum-likelihood expectation-maximization reconstruction of sinograms with arbitrary noise distribution using NEC-transformations", *IEEE Trans. Med. Imaging* 20 (3): 365-375 May 2001.
- [**Nuyts, 02**] Nuyts J, "On estimating the variance of smoothed MLEM images", *IEEE Trans. Nuc. Science* 49 (3): 714-721 Part 1 Jun 2002.
- [**Nuyts, 03**] Nuyts J, Fessler JA, "A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution", *IEEE Trans. Med. Imaging* 22 (9): 1042-1052 Sep 2003.
- [**O'Sullivan, 85**] O'Sullivan JD, "A fast sinc function gridding algorithm for Fourier inversion in computer tomography", *IEEE Trans. Med. Imaging* 4 (4): 200-207 1985.
- [**Ollinger, 95**] Ollinger JM, "Detector efficiency and Compton scatter in fully 3D PET", *IEEE Signal Processing Magazine* 14 (1): 43-55 Jan 1997.
- [**Ollinger, 97**] Ollinger JM, Fessler JA, "Positron-emission tomography", *IEEE Signal Processing Magazine* 14 (1): 43-55 Jan 1997.

- [**Ore, 49**] Ore A, Powell JL, "Three-photon annihilation of an electron-positron pair", *Physical Review* 75 (11): 1696-1699 Jun 1949.
- [**Palmer, 92**] Palmer MR, Brownell GL, "Annihilation density distribution calculations for medically important positron emitters", *IEEE Trans. Med. Imaging* 11 (3): 373-378 Sep 1992.
- [**Palmer, 05**] Palmer MR, Zhu X, Parker JA, "Modeling and simulation of positron range effects for high resolution PET imaging", *IEEE Trans. Nuc. Science* 52 (5): 1391-1395 Part 1 Oct 2005.
- [**Pan, 91**] Pan TS, Yagle AE, "Numerical study of multigrid implementations of some iterative image reconstruction algorithms", *IEEE Trans. Med. Imaging* 10 (4): 572-588 1991.
- [**Pautrot, 03**] Pautrot C, "Reconstruction d'images par des méthodes statistiques en Tomographie par Émission de Positons pour l'imagerie en oncologie", *Rapport de Stage pour l'année de D.E.S.S.*, 2003.
- [**Peters, 81**] Peters TM, "Algorithms for fast back- and re-projection in computed tomography", *IEEE Trans. Nuc. Science* 28 (4): 3641-3647 Aug 1981.
- [**Qi, 98a**] Qi J, Leahy RM, Cherry SR, et al., "High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner", *Phys. Med. Biology* 43 (4): 1001-1013 Apr 1998.
- [**Qi, 98b**] Qi J, Leahy RM, Hsu CH, et al, "Fully 3D bayesian image reconstruction for the ECAT EXACT HR+", *IEEE Trans. Nuc. Science* 45 (3): 1096-1103 Part 2 Jun 1998.
- [**Qi, 99**] Qi J, Leahy RM, "A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data", *IEEE Trans. Med. Imaging* 18 (4): 293-305 Apr 1999.
- [**Qi, 06**] Qi J, Leahy RM, "Iterative reconstruction techniques in emission computed tomography", *Phys. Med. Biology* 51 (15): R541-R578 Aug 7 2006.
- [**Rafecas, 04**] Rafecas M, Böning G, Pichler BJ, et al, "Effect of noise in the probability matrix used for statistical reconstruction of PET data", *IEEE Trans. Nuc. Science* 51 (1): 149-156 Part 1 Feb 2004.
- [**Rajeevan, 92**] Rajeevan N, Rajgopal K, Krishna G, "Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography", *IEEE Trans. Med. Imaging* 7 (4): 273-278 Dec 1988.
- [**Ranganath, 88**] Ranganath MV, Dhawan AP, Mullani N, "A multigrid expectation maximization reconstruction algorithm for positron emission tomography", *IEEE Trans. Med. Imaging* 7 (4): 273-278 Dec 1988.
- [**Raylman, 96**] Raylman RR, Hammer BE, Christensen NL, "Combined MRI-PET scanner: a Monte Carlo evaluation of the improvements in PET resolution due to the effects of a static homogeneous magnetic field", *IEEE Trans. Nuc. Science* 43 (4): 2406-2412 Part 2 Aug 1996.
- [**Riederer, 78**] Riederer SJ, Pelc NJ, Chesler DA, "The noise power spectrum in computed X-ray tomography", *Phys. Med. Biology* 23 (3): 446-454 1978.
- [**Rockmore, 76**] Rockmore AJ, Macovski A, "A maximum likelihood approach to emission image reconstruction from projections", *IEEE Trans. Nuc. Science* 23 (4): 1428-1432 1976.
- [**Sánchez-Crespo, 04**] Sánchez-Crespo A, Andreo P, Larsson SA, "Positron flight in human tissues and its influence on PET image spatial resolution ", *Eur. J. Nuc. Medicine* 31 (1): 44-51 Jan 2004.

-
- [**Sánchez-Crespo, 06**] Sánchez-Crespo A, Larsson SA, “The influence of photon depth of interaction and non-collinear spread of annihilation photons on PET image spatial resolution”, *Eur. J. Nuc. Medicine* 33 (8): 940-947 Aug 2006.
- [**Sauer, 93**] Sauer K, Bouman C, “A local update strategy for iterative reconstruction from projections”, *IEEE Trans. Signal Processing* 41 (2): 534-548 Feb 1993.
- [**Shao, 94**] Shao LX, Freifelder R, Karp JS, “Triple energy window scatter correction technique in PET”, *IEEE Trans. Med. Imaging* 13 (4): 641-648 Dec 1994.
- [**Shepp, 74**] Shepp LA, Logan BF, “The Fourier reconstruction of a head section”, *IEEE Trans. Nuc. Science* NS21 (3): 21-43 1974.
- [**Shepp, 82**] Shepp LA, Vardi Y, “Maximum likelihood reconstruction for emission tomography”, *IEEE Trans. Med. Imaging* 1 (2): 113-122 Oct 1982.
- [**Siddon, 85**] Siddon RL, “Fast calculation of the exact radiological path for a three-dimensional CT array”, *Medical Physics* 12 (2): 252-255 1985.
- [**Snyder, 85**] Snyder DL, Miller MI, “The use of sieves to stabilize images produced with the EM algorithm for emission tomography”, *IEEE Trans. Nuc. Science* 32 (5): 3864-3872 1985.
- [**Snyder, 87**] Snyder DL, Miller MI, Thomas LJ, et al., “Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography”, *IEEE Trans. Med. Imaging* 6 (3): 228-238 Sep 1987.
- [**Soares, 00**] Soares EJ, Byrne CL, Glick SJ, “Noise characterization of block-iterative reconstruction algorithms: I. Theory”, *IEEE Trans. Med. Imaging* 19 (4): 261-270 Apr 2000.
- [**Soares, 05**] Soares EJ, Glick SJ, Hoppin W, “Noise characterization of block-iterative reconstruction algorithms: II. Monte Carlo Simulations”, *IEEE Trans. Med. Imaging* 24 (1): 112-121 Jan 2005.
- [**Sossi, 95**] Sossi V, Barney JS, Harrison R, “Effect of scatter from radioactivity outside of the field of view in 3D PET”, *IEEE Trans. Nuc. Science* 42 (4): 1157-1161 Aug 1995.
- [**Spinks, 98**] Spinks TJ, Miller MP, Bailey DL, et al., “The effect of activity outside the direct field of view in a 3D-only whole-body positron tomograph”, *Phys. Med. Biology* 43 (4): 895-904 1998.
- [**Stark, 81**] Stark H, Woods JW, Paul I, et al., “An investigation of computerized tomography by direct Fourier inversion and optimum interpolation”, *IEEE Trans. Biomed. Engineer.* 28 (7): 496-505 1981.
- [**Stayman, 00**] Stayman J, Fessler JA, “Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction”, *IEEE Trans. Med. Imaging* 19 (6): 601-615 Jun 2000.
- [**Stayman, 04**] Stayman J, Fessler J, “Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems”, *IEEE Trans. Med. Imaging* 23 (3): 269-284 Mar 2004.
- [**Stearns, 90**] Stearns C, Chesler D, Brownell G, “Accelerated image reconstruction for a cylindrical positron tomography using Fourier domain methods”, *IEEE Trans. Nuc. Science* 37 (2): 773-777 Part 1 Apr 1990.
- [**Stickel, 05**] Stickel JR, Cherry SR, “High-resolution PET detector design: modelling components of intrinsic spatial resolution”, *Phys. Med. Biology* 50 (2): 179-195 Jan 21 2005.

-
- [**Strother, 90**] Strother SC, Casey ME, Hoffman EJ, "Measuring PET scanner sensitivity: relating countrates to image signal-to-noise rations using noise equivalent counts", *IEEE Trans. Nuc. Science* 37 (2): 783-788 Part 1 Apr 1990.
- [**Strul, 03**] Strul D, Slates RB, Dahlbom M, et al., "An improved analytical detector response function mdel for multilayer small-diameter PET scanners", *Phys. Med. Biology* 48 (8): 979-994 Apr 21 2003.
- [**Sureau, 06**] Sureau FC, Comtat C, Reader AJ, et al., "Improved Clinical Parametric Imaging Using List-Mode Reconstruction via Resolution Modeling", *Nucl. Sci. Symp. Conf. Rec., Vol. 6, 3507-3510, Oct 2006.*
- [**Svalbe, 04**] Svalbe I, "Sampling properties of the discrete Radon tansform", *Discrete Applied Math.* 139 (1-3): 265-281 Apr 30 2004.
- [**Tai, 98**] Tai Y-C, Chatziioannou A, Dalbom M, et al., "Investigation of deadtime characteristics for simultaneous emission-transmission data acquisition in PET", *IEEE Trans. Nuc. Science* 45 (4): 2200-2204 Part 2 Aug 1998.
- [**Tanaka, 92**] Tanaka E, "Improved iterative image reconstruction with automatic noise artefact suppression", *IEEE Trans. Med. Imaging* 11 (1): 21-27 Mar 1992.
- [**Tanaka, 03**] Tanaka E, Kudo H, "Subset-dependent relaxation in block-iterative algorithms for image reconstruction in emission tomography", *Phys. Med. Biology* 48 (10) 1405-1422 May 21 2003.
- [**Thompson, 88**] Thompson CJ, "The effect of collimation on scatter fraction in multi-slice PET", *IEEE Trans. Nuc. Science* 35 (1): 598-602 Part 1 Feb 1988.
- [**Thompson, 93**] Thompson CJ, "The problem of scatter correction in positron volume imaging", *IEEE Trans. Med. Imaging* 12 (1): 124-132 Mar 1993.
- [**Toft, 96**] Toft P, "The Radon Transform – Theory and Implementation", *Ph. D. Thesis, Depart. of Mathematical Modelling, Section for Digital Signal Processing, Technical University of Denmark, 1996.*
- [**Townsend, 93**] Townsend DW, Defrise M, "Image reconstruction methods in positron emission tomography", *Lectures given in the Academic Training Programme of CERN, Geneva, 1993.*
- [**Townsend, 98**] Townsend DW, Isoardi RA, Bendriem B, "Volume imaging tomographs", in "The Theory and Practice of 3D PET", Eds. B. Bendriem and D. W. Townsend, Kluwer Academic Publishers, 1998.
- [**Tsui, 93**] Tsui BMW, Terry JA, Gullberg GT, "Evaluation of cardiac cone-beam SPECT using observer performance experiments and ROC analysis", *Investigative Radiology* (28): 1101-1112 1993.
- [**Vandenberghe, 01**] Vandenberghe S, D'Asseler Y, Van de Walle R, et al., "Iterative reconstruction algorithms in nuclear medicine", *Comput. Med. Imag. Graphics* 25 (2): 105-111 Mar-Apr 2001.
- [**Vardi, 85**] Vardi Y, Shepp LA, Kaufman L, "A statistical model for positron emission tomography", *J. of the American Statistical Association* 80 (389): 8-20 1985.
- [**Vardi, 98**] Vardi Y, Lee D, "Discrete Radon transform and its approximate inversion via the EM algorithm", *Intern. Journal of Imaging Systems and Technology* 9 (2-3): 155-173 1998.
- [**Veklerov, 87**] Veklerov E, Llacer J, "Stopping rule for the MLE algorithm based on statistical hypothesis-testing", *IEEE Trans. Med. Imaging* 6 (4): 313-319 Dec 1987.

-
- [**Veklerov, 88**] Veklerov E, Llacer J, Hoffman EJ, “MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule”, *IEEE Trans. Nuc. Science* 35 (1): 603-607 Part 1 Feb 1988.
- [**Vollmar, 02**] Vollmar S, Michel C, Treffert JT, et al., “*HeinzlCluster*: accelerated reconstruction for FORE and OSEM3D”, *Phys. Med. Biology* 47 (15) 2651-2658 Aug 2002.
- [**Vozza, 97**] Vozza D, Moisan C, Paquet S, “An improved model for the energy resolution of multicrystal encoding detectors for PET”, *IEEE Trans. Nuc. Science* 44 (2): 179-183 Apr 1997.
- [**Wilson, 93**] Wilson D, Tsui B, “Noise properties of filtered-backprojection and ML-EM reconstructed emission tomographic images”, *IEEE Trans. Nuc. Science* 40 (4): 1198-1203 Part 1 Aug 1993.
- [**Wilson, 94**] Wilson DW, Tsui BMW, Barrett HH, “Noise properties of the EM algorithm. 2. Monte-Carlo simulations”, *Phys. Med. Biology* 39 (5) 847-871 May 1994.
- [**Xia, 95**] Xia W, Lewitt RM, Edholm PR, “Fourier correction for spatially variant collimator blurring in SPECT”, *IEEE Trans. Med. Imaging* 14 (1): 100-115 Mar 1995.
- [**Yavus, 97**] Yavus M, Fessler J., “New statistical models for random-precorrected PET scans”, in “*Information Processing in Medical Imaging*”, Eds. J. Duncan, G. Gindi, *Lectures Notes in Computer Science*, Berlin, Germany, Springer-Verlag, vol. 1230, 190-203, 1997.
- [**Zanzonico, 04**] Zanzonico P, “Positron Emission Tomography: a review of basic principles, scanner design and performance, and current systems”, *Sem. Nuc. Medicine* 34 (2): 87-111 Apr 2004.
- [**Zeng, 01**] Zeng GL, “Image reconstruction – a tutorial”, *Comput. Med. Imag. Graphics* 25 (2): 97-103 Mar-Apr 2001.
- [**Zhuang, 94**] Zhuang W, Gopal SS, Hebert TJ, “Numerical evaluation of methods for computing tomographic projections”, *IEEE Trans. Nuc. Science* 41 (4): 1660-1665 Part 1 Aug 1994.