University of the Studies of Milano–Bicocca
Faculty of Mathematical, Physical and Natural Sciences
PhD Program in Physics and Astronomy
Coordinator: Prof. Giberto Chirico

Doctor of Philosophy Dissertation

Improvements in quality and quantification of 3D PET images

by

Eugenio Rapisarda

Matriculation number 059055

Supervisors:
Prof. Marco Paganoni Dott. Valentino Bettinardi

Milano, Italy, January 2012

To my beloved family

We cannot change the cards we are dealt, just
how we play the hand.

Randy Pausch

Ad hoc, ad loc and quid pro quo!
So little time, so much to know!

Jeremy Hillary Boob, Ph.D., the Nowhere Man

I have not failed 10,000 times. I have successfully
found 10,000 ways that will not work.

Thomas Alva Edison

People can come up with statistics to prove
anything. 14% of people know that.

Homer J. Simpson

Acknowledgements

This thesis presents the work started three years ago and which, today, allows me to reach the craved destination of the Doctor of Philosophy degree in Physics. Firstly I thank everyone who will congratulate me, but I would like to underline — hoping that the words hereafter will not appear ritual ones — that this goal has been reached only thanks to the irreplaceable contribution by many, fundamental people, who deserve my gratitude.

Firstly I want to thank, from the bottom of my heart, Valentino, with whom I shared this journey day by day, because he took me by the hand supporting my proposals and suggesting me always with wisdom, patience and huge competence. We faced the difficulties together and, still together, we defeated them. If the quality of the work contained in these pages will be considered good, I certainly owe it to him.

In the same way I would like to thank Marco, for having always advised and supported me, making me feel the neighborly of the University — but, much more importantly, his *personal* one — and believing in me in many occasions, to such an extent that he let me be at his side in the Physics I course and cultivate one of my biggest passions.

Fundamental thanks go to prof.ssa Maria Carla Gilardi, who has strongly willed that I would continue my path after the Master degree inside the same scientific group and who has supported me in many different ways.

I cannot and I do not want to drop the acknowledgements to the whole Nuclear Medicine Department of the Scientific Institute San Raffaele Hospital in Milano, where I have always found cooperating and welcoming atmosphere since my first day of the Master thesis. The collaboration and the professionalism by the physicians, the technicians and the secretaries has largely simplified my job and let me express my capabilities and my creativity at their best. A particular gratitude goes to dott.ssa Annarita Savi and dott.ssa Isabella Castiglioni, my two “adoptive mums”, for their huge esteem for me — shown in many occasions, e.g. by often asking my point of view and my help on different topics — and for

having supported and advised me in several moments.

My thanks go also to the secretaries (particularly to sig.ra Luisa Aquino) and to the entire personnel of IBFM-CNR in Segrate, for giving me help in the logistics and in the burocratic problems, always doing their best to give me what I needed.

Even if during this period I physically spent few time at the University, thanks must be warmly addressed (and it is a pleasant task) to prof. Stefano Ragazzi and all his colleagues of the Physics Department at the University of Milano-Bicocca, for having advised, supported and esteemed me in all these years.

At the base of all my modest successes there have always been the sacrifices made and the teachings given by my parents. I am aware that it is impossible to adequately express my gratitude and my affection to them, but I want to thank my dad and my mum from the bottom of my heart for all they were able to give me, in many different moments, both concerning the sustenance and the human point of view. I really hope you will have the fair reward for all the good you have done.

The biggest gift I have received from my parents is named Marila. From the very beginning I was, in her eyes, a reference figure, the "huge brother" for whom cherish infinite admiration and esteem and to whom give gratuitous affection. I thank her for all she has been able to transmit to me in these years, by many little and big actions which have always filled my heart with joy and gratitude. I hope life will give you the satisfactions a great person like you deserves. And I hope I will have the luck of being able to be always at your side to rejoice for this with you.

I would like also to thank my grandparents for the affection and huge esteem they have always conveyed by their behaviour. In their eyes I have always been a grandson to be proud of since the day I was born. My childhood memories are full of many pictures of which they are the protagonists. I hope I will be able, one day, to give my grandchildren a tiny part of the love I received from them. Similarly, I want to thank my uncles, aunts and all the other relatives for all the support they have always given me in different ways and for the universe of beautiful moments they have let me live with them.

During these years my journey has crossed the one of many colleagues and each of them has gifted me a smile, a teaching, a new friendship. Therefore I would like to thank, from the bottom of my heart and in strict alphabetical order: Alberto, Alessandro, Annalisa, Christian, Cristina, Daniele, Eleonora, Francesca, Gabriele, Ignazio, Luca, Miriam, Pietro, Richard, Salvatore. In particular to some of them (who know they are the recipients of these sentences) I am profoundly grateful for the marvellous moments spent together — during and outside work time — and for their personal and professional support they were able to give me

in many, many occasions. I wish all of you to walk a way which can give you the biggest satisfactions — human and personal, prior than professional: you really deserve it.

I would also like to send some greetings and thanks to my PhD colleagues (in particular the ones belonging to the XXIV cycle) who have shared with me, either partly or totally, this journey. We met now and then and mutually exchanged opinions on our works: I often received appreciation and esteem from them, which I reciprocate with sincerity. Particular thanks and greetings go to Arabella, Andrea and Federico.

In these years I had the luck of being able to cultivate one great passion, teaching. One of the biggest privileges I received was the possibility of meeting many girls and boys and seeing in their eyes their curiosity and joy of living. I therefore thank my students of the various university courses — at the University of the Studies of Milano-Bicocca and at the University Vita-Salute San Raffaele — and of the high school courses — at Liceo Scientifico Statale Paolo Frisi in Monza — for letting me reflect more and more deeply on my knowledge and constantly improve, but above all for having gifted me many pleasant rememberings and a bit of their endless energy.

A warm thought always goes to all my school teachers, who have been able, at every level, to shape me and get the best I could express. This goal has been reached also thanks to you.

To complete this journey I, obviously and normally, had to make sacrifices, too. Among them, I unfortunately had to reduce the time given to my friends. In particular, I would like to greet Francesca, Diego and Giuseppe, thanking them for the patience they had in waiting for me and shift with very little shreds of my time. You have given me several moments of happiness — which I have always with me — and the possibility of having someone to talk to, to confide in, have fun and reflect with. I am deeply grateful for all of this.

Now it is time to thank a very special person, the girl who gave me the possibility of having her at my side in many occasions and with whom I could share countless moments and experiences. We grew up together, my beloved Francy, and I hope we will be able to continue doing it and to make our relationship stronger and richer day by day. I thank you for all you have gifted me, for your irreplaceable help in these years of PhD course, for making each of the moments lived together unforgettable, even if often too brief. In few words, thanks for *you*.

A sincere gratitude, in the end, also to all people not cited here due to the limited room, but who have equally represented a fundamental element of these years.

Now a new season of my life begins. I do not know what the destination will

be, nor I am aware of the path to reach it. But, this is certain, I know I will never walk alone, but always in company of all the presents and rememberings each of you has left to me. The most beautiful luggage I could hope.

Ringraziamenti

Questa tesi presenta il lavoro iniziato tre anni fa e che mi porta, oggi, a conseguire la meta agognata del titolo di Dottore di Ricerca in Fisica. Ringrazio innanzitutto tutti coloro che si congratuleranno con me, ma tengo a sottolineare — sperando che le parole che seguono non appaiano rituali — che questo traguardo è stato raggiunto solo grazie all'insostituibile apporto di tante, fondamentali persone che meritano la mia gratitudine.

Innanzitutto voglio ringraziare dal profondo del mio cuore Valentino, con il quale ho condiviso giorno dopo giorno questo percorso, perché mi ha condotto per mano supportando le mie proposte e consigliandomi sempre con saggezza, pazienza e grande competenza. Abbiamo affrontato insieme le difficoltà ed altrettanto insieme le abbiamo vinte. Se la qualità del lavoro contenuto in queste pagine verrà considerata buona, lo devo certamente a lui.

Allo stesso modo voglio ringraziare Marco, per avermi consigliato e supportato sempre, facendomi sempre sentire la vicinanza dell'Università — ma soprattutto la sua *personale* vicinanza — e credendo in me in tante occasioni, fino al punto di permettermi di affiancarlo nell'insegnamento di Fisica 1 e di coltivare una mia grande passione.

Un ringraziamento fondamentale va alla professoressa Maria Carla Gilardi, che ha fortemente voluto che continuassi il mio percorso dopo la tesi specialistica presso lo stesso gruppo e che mi ha supportato in svariati modi.

Non posso e non voglio rinunciare ai ringraziamenti per tutto il Servizio di Medicina Nucleare dell'Istituto Scientifico Ospedale San Raffaele di Milano, presso il quale ho trovato sempre un'atmosfera collaborativa ed accogliente sin dal mio primo giorno di tesi specialistica. La collaborazione e la professionalità da parte dei medici, dei tecnici e delle segretarie ha notevolmente semplificato il mio lavoro e mi ha dato la possibilità di esprimere al meglio le mie capacità e la mia inventiva. Un ringraziamento particolare va alle dottoresse Annarita Savi ed Isabella Castiglioni, le mie due "mamme adottive", per la loro enorme stima nei miei confronti — dimostrata in svariate occasioni, ad esempio chiedendo spesso

il mio parere ed il mio aiuto su tante questioni — e per avermi supportato e consigliato in molti momenti.

Alle segretarie (in particolare alla Sig.ra Luisa Aquino) ed al personale vario dell'IBFM-CNR di Segrate va il mio ringraziamento per avermi fornito aiuto nella logistica e nelle questioni burocratiche, adoperandosi sempre per fornirmi ciò di cui necessitavo.

Pur se durante questo percorso ho fisicamente speso poco tempo tra le mura universitarie, è doveroso e piacevole ringraziare calorosamente il prof. Stefano Ragazzi e tutti i suoi colleghi del Dipartimento di Fisica dell'Università di Milano-Bicocca, per avermi consigliato, supportato e stimato in tutti questi anni.

Alla base di ogni mio modesto successo ci sono sempre stati i sacrifici compiuti e gli insegnamenti trasmessimi dai miei genitori. Sono conscio dell'impossibilità di esprimere adeguatamente la mia gratitudine ed il mio affetto verso di loro, ma voglio ringraziare dal profondo del cuore mio papà e mia mamma per tutto ciò che sono sempre stati in grado di darmi, nei più disparati momenti, sia dal punto di vista del sostentamento, sia dal punto di vista umano. Spero che possiate avere la giusta ricompensa per tutto il bene che avete compiuto.

Il più grande regalo ricevuto dai miei genitori si chiama Marila. Da sempre per lei ho rappresentato una figura di riferimento, il "fratellone" per il quale nutrire un'ammirazione ed una stima infiniti ed al quale donare un affetto gratuito. La ringrazio per tutto ciò che ha saputo trasmettermi in questi anni, con tanti piccoli e grandi gesti che mi hanno sempre riempito il cuore di gioia e gratitudine. Spero che la vita possa riservarti le soddisfazioni che una grande persona come te merita. E spero di avere la fortuna di poter essere sempre al tuo fianco per gioire di ciò insieme a te.

Vorrei anche ringraziare i miei nonni per l'affetto e la stima enorme che mi hanno sempre trasmesso con i loro gesti. Ai loro occhi sono stato un nipote di cui andare fieri sin dal giorno della mia nascita. I miei ricordi di infanzia sono pieni di tante immagini di cui loro sono protagonisti, ricordi che — ne sono certo — mi accompagneranno per tutta la vita. Spero che un giorno sarò in grado di regalare ai miei nipoti una piccolissima parte di tutto l'amore che ho ricevuto da loro. Allo stesso modo, voglio ringraziare i miei zii, zie e tutti gli altri familiari per tutto il supporto che mi hanno sempre dato in svariati modi e per le miriadi di bei momenti che mi hanno permesso di vivere insieme a loro.

Durante questi anni la mia strada si è incrociata con quella di tanti miei colleghi, ciascuno dei quali mi ha regalato un sorriso, un insegnamento, una nuova amicizia. Vorrei quindi ringraziare, dal profondo del cuore ed in rigoroso ordine alfabetico: Alberto, Alessandro, Annalisa, Christian, Cristina, Daniele, Eleonora, Francesca, Gabriele, Ignazio, Luca, Miriam, Pietro, Richard, Salvatore. Soprattutto verso alcuni (che sanno di essere i destinatari di queste frasi) sono

profondamente grato per i bellissimi momenti trascorsi assieme — dentro e fuori le mura lavorative — e per il supporto personale e professionale che hanno saputo darmi in tante, tante occasioni. Auguro a tutti voi di percorrere una strada che possa regalarvi le più grandi soddisfazioni, umane e personali prima che professionali: lo meritate davvero.

Vorrei mandare un pensiero ed un ringraziamento anche ai miei colleghi dottorandi (in particolare quelli del XXIV ciclo) che hanno condiviso con me, in parte o totalmente, questo cammino. Ci siamo incontrati di tanto in tanto e confrontati sui rispettivi lavori; ho spesso ricevuto apprezzamenti e stima da parte loro, che ricambio con sincerità. Un saluto ed un grazie particolare va ad Arabella, Andrea e Federico.

In questi anni ho avuto la fortuna di poter coltivare una mia grande passione, l'insegnamento. Uno dei più grandi privilegi ricevuti è stata la possibilità di incontrare tanti giovani ragazze e ragazzi e di vedere nei loro occhi la curiosità e la gioia di vivere. Ringrazio quindi i miei studenti degli svariati corsi universitari — all'Università degli Studi di Milano–Bicocca ed all'Università Vita-Salute San Raffaele — e dei corsi liceali — al Liceo Scientifico Statale Paolo Frisi di Monza — per avermi permesso di riflettere sempre più a fondo sulle mie conoscenze e migliorarle costantemente, ma soprattutto per avermi regalato tanti piacevoli ricordi ed un pizzico della loro energia sterminata.

Un pensiero affettuoso va sempre ai miei insegnanti di scuola, che ad ogni livello hanno saputo forgiarmi e trarre da me il meglio che potessi esprimere. Questo traguardo è anche merito vostro.

Per completare questo percorso ho dovuto, come è normale che sia, anche compiere sacrifici. Tra questi, ho purtroppo dovuto ridurre il tempo dedicato ai miei amici. In particolare vorrei salutare Francesca, Diego e Giuseppe, ringraziandoli della pazienza che hanno avuto nell'aspettarmi ed accontentarsi di piccolissime briciole del mio tempo. Mi avete regalato tanti momenti di felicità — che porto sempre dentro di me — e la possibilità di avere qualcuno con cui parlare, confidarsi, divertirsi, riflettere. Vi sono profondamente grato per tutto ciò.

A questo punto è giunto il momento di ringraziare una persona veramente speciale, la ragazza che mi ha dato la possibilità di averla accanto in tante occasioni e con la quale condividere tanti momenti ed esperienze. Siamo cresciuti insieme, mia amata Francy, e spero che potremo continuare a farlo ed a rendere il nostro rapporto sempre più forte e ricco. Ti ringrazio per tutto ciò che mi hai regalato, per l'aiuto insostituibile in questi anni di dottorato, per aver reso indimenticabile ognuno dei momenti vissuti insieme, anche se spesso troppo brevi. In breve, grazie di *te*.

Un grazie sincero, infine, anche a tutti coloro che non sono stati citati per motivi di spazio, ma che hanno ugualmente rappresentato un elemento fondamentale

di questi anni.

Inizia ora una nuova stagione della mia vita. Non so quale sarà la meta, né conosco la strada per arrivarci. Ma, è certo, so che non camminerò mai da solo, ma sempre in compagnia dei regali e dei ricordi che ciascuno di voi mi ha lasciato. Il più bel bagaglio che potessi sperare.

Contents

Contents i

Summary v

Riassunto ix

Part I Introduction

1 Scope of the work 3
- 1.1 The dawn of medical physics . . . 3
- 1.2 Digital 3D anatomy . . . 3
- 1.3 The other half of the sky: functional imaging . . . 5
- 1.4 Problems and potentialities of image reconstruction . . . 9
- 1.5 Structure of the work . . . 9

2 Positron Emission Tomography 11
- 2.1 Physical principles . . . 11
- 2.2 Effects superimposed to the signal due to emission . . . 19
  - 2.2.1 Poisson statistics . . . 19
  - 2.2.2 Source decay . . . 19
  - 2.2.3 Positron energy and range . . . 21
  - 2.2.4 Random coincidences . . . 22
  - 2.2.5 Scattered coincidences . . . 22
  - 2.2.6 Attenuation of photons in the matter . . . 22
- 2.3 Corrections to emission effects . . . 24
  - 2.3.1 Randoms . . . 24
  - 2.3.2 Scatter . . . 24
  - 2.3.3 Attenuation . . . 26
  - 2.3.4 Source decay . . . 28
- 2.4 Effects superimposed to the signal due to acquisition . . . 30

2.4.1 Energy resolution . . . 30
2.4.2 Dead–time . . . 31
2.4.3 Crystal efficiency . . . 32
2.4.4 Radial geometry . . . 32
2.4.5 Finite dimensions of crystals and depth of interaction . . . 33
2.4.6 Inter–crystal scatter . . . 35
2.5 Corrections to acquisition effects . . . 35
2.5.1 Radial repositioning . . . 35
2.5.2 Dead–time . . . 35
2.5.3 Detector normalization . . . 36
2.6 The Point Spread Function . . . 37

3 Image reconstruction **41**
3.1 Object representation . . . 41
3.2 Data representation . . . 42
3.3 Projector and backprojector . . . 44
3.4 Analytic methods . . . 44
3.4.1 Filtered back–projection . . . 46
3.5 Iterative methods . . . 49
3.5.1 Maximum Likelihood with Expectation Maximization . . . 50
3.5.2 Ordered Subsets Expectation Maximization . . . 55
3.6 Time-of-flight reconstruction . . . 57
3.6.1 Analytical algorithms . . . 58
3.6.2 Iterative algorithms . . . 60

Part II Performed work

4 PSF measurements **63**
4.1 PSF measurements . . . 63
4.2 The PSF model . . . 64
4.3 Accuracy of the method . . . 70

5 Implementation **75**
5.1 OSEM algorithm and PSF implementation . . . 75
5.2 Organization of the PSF kernel . . . 76
5.3 The edge effect . . . 79

6 Regularization **87**
6.1 Theoretical background . . . 87
6.1.1 Variational regularization . . . 89
6.2 Proposal of a new variational regularization prior . . . 92
6.3 Optimization of the regularization parameters . . . 97

6.4 Implementation of the variational regularization 104

Part III Results

7 Results — Non TOF imaging **107**
7.1 Description of the GE Discovery STE scanner 107
7.2 PSF measurements 108
7.3 Improvements thanks to PSF 110
7.3.1 Quantitative validation 111
7.3.2 Qualitative validation 120
7.4 Effects of the regularization 124
7.4.1 Optimization of the regularization parameters 124
7.4.2 The edge effect 125
7.4.3 Quantitative accuracy 127
7.4.4 Qualitative improvements 135

8 Results — TOF imaging **139**
8.1 Description of the GE Discovery 690 scanner 139
8.2 PSF measurements 140
8.3 Improvements thanks to PSF 142
8.3.1 Quantitative validation 142
8.3.2 Qualitative validation 151
8.4 Effects of the regularization 159
8.4.1 Optimization of the regularization parameters 159
8.4.2 Quantitative accuracy 160
8.4.3 Qualitative improvements 166

Part IV Conclusions

9 Discussion **175**

10 Future perspectives **181**

Mathematics **183**
M.1 Poisson statistics 183
M.2 Point and non–point sources with Gaussian PSF 185
M.3 The convolution operator 192
M.4 The transposed PSF 194
M.5 Derivation of the variational regularization term 201

Bibliography **203**

List of Figures **213**

List of Tables **217**

List of Symbols and Abbreviations **219**

Summary

Positron Emission Tomography and, more recently, hybrid PET/CT systems have become increasingly important in the clinical practice thanks to their capability of providing complementary, spatially coregistered anatomical and functional information. However, PET imaging is still characterized by a "poor" spatial resolution due to the multiple physical effects linked with the 511 keV photon-matter interaction and with the detection apparatus. Low spatial resolution leads to a spread in the activity distribution, resulting in a degradation of the image quality (e.g. reduction of the contrast), in an underestimation of the tracer uptake (activity concentration) and in an overestimation of the lesion volume. In principle, each effect could be estimated by experimental measurements or by simulation studies and then included in the reconstruction scheme to be compensated. Unfortunately, some of these effects are very difficult or nearly impossible to be measured. Consequently, another approach may consist in accounting for all these effects by a "global" Point Spread Function (PSF), which describes how a point source is rendered by the system itself. The knowledge of the PSF, in principle, allows the correction of the resolution degradation and the recovery of the correct quantitative information; on the other hand, this approach requires the knowledge of the PSF in each point of the PET Field Of View (FOV) since the response of the system is not uniform across the FOV of the scanner.

In this thesis it has been proposed a spatially variant PSF implementation in the image space of a 3D Ordered Subsets Expectation Maximization (OSEM) algorithm. Two different scanners from General Electric Medical Systems were considered, without (DSTE) and with (D690) Time Of Flight (TOF) capability. The PSF was chosen to be a 3D Gaussian function, which — thanks to the cylindrical symmetry of the scanners under study — was factorized into a transaxial two-dimensional Gaussian function and an axial one-dimensional Gaussian function. In particular, along the radial direction the Gaussian function has been chosen asymmetrical, with the larger tail directed towards the center of the scanner. Consequently, in each point of the FOV the knowledge of the PSF coincides

with the knowledge of four spread parameters, the *internal radial*, the *external radial*, the *tangential* and the *axial* ones.

Experimental measurements, therefore, were performed to determine the dependences of the PSF spread parameters on the position inside the scanner FOV. The way of measuring the response of the PET system is, in fact, an important point. Even in the simplest case, i.e. using a small point source in different positions of the scanner FOV, several factors have to be taken into account to obtain a well representative description of the system response, such as the dimensions of the source, the type of isotope used, the surrounding media (air, water, warm radioactive background), the number and the position of the measurements.

In the case here presented the cylindrical symmetry of the scanners suggested considering dependences on the radial and axial distances from the scanner centre. To measure the PSF of the system a small radioactive ^{22}Na point source (a cylinder with diameter and height of 1 mm encapsulated in Lucite) in air was used. Approximately 400 measurements uniformly distributed within the scanner FOV were performed per scanner. Finally, since the PSF should act at the image level of a 3D OSEM algorithm, it was considered important to determine the trend of the PSF parameters in the same conditions, to match, as much as possible, the PSF evaluations and the algorithm: therefore, the same algorithm was used to reconstruct the point source raw data.

From each reconstructed image three orthogonal two-dimensional planes, passing through the voxel with the maximum intensity, were extracted along three directions (originating the radial-tangential, radial-axial and tangential-axial planes): each of them was then fitted by a two-dimensional function to determine the corresponding spread parameters.

The fitting function took into account the post-filter applied on the images (to reduce the noise content), the *actual* position of the point source (in order to correct the unavoidable imprecisions in the source positioning), the source dimensions (to avoid an overestimation of the spread parameters due to the approximation of point source) and the intrinsic discretization along the axial direction due to the finite dimensions of the slices.

The entire set of spread results obtained with the ^{22}Na source were then fitted by a two-dimensional function of the radial and axial distances from the scanner center, to determine the required analytical dependences on the position inside the scanner FOV.

The proposed method of measurement was also validated and demonstrated its good accuracy in building the PSF model, justifying its use.

The implementation of the PSF consisted in a redefinition of the projector and backprojector of the 3D OSEM algorithm. The practical implementation has been performed by a factorization of the 3D PSF into a transaxial 2D PSF and an axial 1D PSF. Both the transaxial and the axial PSFs were stored into a three-

dimensional matrix to take their spatial dependence into account; the continuous model of the PSF has been discretized by calculating its *integral* for each voxel, allowing for a better adaptive implementation for each specific reconstruction FOV and pixel size (the much more common strategy of filling each voxel with the value of PSF in the *middle point* yields different shapes of the kernel, leading to possible inaccuracies). The dimension of the stored kernel was chosen to be about 4 times the spatial resolution of the scanners: this choice is conservative with respect to the possibility of generating artefacts due to the truncation of the PSF kernel. In this thesis the explicit expression for the transposed PSF operator was also derived, showing that — in the spatially variant case — this does not coincide with the transpose of the PSF kernel.

The PSF was tested on some phantom and clinical data. The results showed improved quantitative accuracy, spatial resolution and image quality. Furthermore, the combined use of TOF and PSF appeared to allow them to take advantage of each other, leading to the best results.

Unfortunately, a common effect of iterative reconstruction techniques is the increase of noise as iterations proceed, due to the ill-posed nature of the reconstruction problem. Usually in clinical practice the quality of the images is privileged over their quantitative accuracy by stopping the iterative algorithm after few iterations and, consequently, far from convergence. This is true all the more if PSF is included in the algorithm, since the speed of convergence results lower than in non–PSF algorithms. Another important effect observed in PSF-based reconstructions (and easily recognizable throughout the literature) is the enhancement of regions with sharp intensity transitions. In this thesis it was demonstrated, by means of 1D and 2D simulations, to be strongly related to the implementation of the spatial resolution recovery and, even in presence of a perfectly matched kernel, unavoidable unless an unpractical number of iterations is used.

Regularization techniques have been demonstrated to be useful for taking noise under control during the reconstruction and improving the benefits from the use of the PSF information by increasing the number of iterations used. In particular, in this thesis a Bayesian variational regularization strategy has been tested and employed. Two good candidates for the use in PET practice are the Huber (or Gauss-Total Variation) and the generalized p-Gaussian priors. The former provides good preservation of spatial resolution thanks to the Total Variation component for high gradients, but the Gaussian component for low gradients might be insufficient in controlling very noisy environments (as often encountered in PET, in particular when the number of iterations is increased to exploit the PSF action) unless the regularization strength is set to a very high value, obtaining an unnatural reconstructed image. The p-Gaussian prior provides a very strong smoothing on background regions, resulting in good noise

control, while it smoothes much less in signal regions.

In this thesis a modification of the p-Gaussian prior was proposed to maintain the smoothing effect for low gradients (i.e. in background regions) and to reduce the spatial resolution loss, while retaining "natural" transitions and appearance in the image. A 3D OSEM algorithm has been modified to include the proposed prior using a Maximum A Posteriori One Step Late multiplicative approach. The value for the parameter p was chosen by qualitatively evaluating the results obtained with different values and choosing the one yielding the best compromise between noise suppression, preservation of spatial resolution and natural appearance of the image.

The considered priors depend on some regularization parameters. In this thesis a figure of merit, taking into account both the qualitative and the quantitative content, was proposed to evaluate the global "detectability" of a lesion. The validation of this detectability index showed a very good correlation with the human response and, thus, justified its use to set the regularization parameters.

The regularization parameters were then determined by maximizing the detectability index for each prior. This optimization was performed for a sphere with diameter 10 mm and 10 OSEM iterations.

The validation of the proposed modifications was quantitative on data acquired with a NEMA IEC Body Phantom and qualitative on data relative to two oncological patients and consisted of a comparison between the standard reconstruction algorithms, the proposed algorithm, the results obtained with the p-Gaussian prior and with Gauss-Total Variation. This comparison showed an effective control of noise (but with natural appearance of the image) by the proposed prior with a contemporary good preservation of spatial resolution, contrast and definition of the activity distribution. Moreover, the proposed prior was shown to be able also to take the edge artefact under control, drastically reducing the overshoots originating at large transitions in the image. Positive results were obtained also when the regularization strategy was used in conjunction with the TOF information, suggesting a possible future employment in the PET reconstruction framework.

Riassunto

La Tomografia ad Emissione di Positrone (PET) e, più recentemente, i sistemi ibridi PET/TC sono diventati sempre più importanti nella pratica clinica grazie alla loro capacità di fornire informazioni anatomiche e funzionali complementari e spazialmente coregistrate. Tuttavia, l'imaging PET è caratterizzato da una scarsa risoluzione spaziale dovuta a molteplici effetti fisici collegati con l'interazione tra i fotoni a 511 keV e la materia e con l'apparato di rivelazione. La bassa risoluzione spaziale porta ad uno sparpagliamento della distribuzione di attività, che origina una degradazione della qualità dell'immagine (es. riduzione del contrasto), una sottostima dell'accumulo di tracciante (concentrazione di attività) ed una sovrastima del volume delle lesioni. In teoria, ogni effetto potrebbe essere valutato da misure sperimentali o da studi di simulazione e poi incluso nel processo di ricostruzione per essere compensato. Sfortunatamente, alcuni di questi effetti sono molto difficili o quasi impossibili da misurare. Di conseguenza, un altro approccio potrebbe consistere nel tenere conto di tutti questi effetti tramite una Point Spread Function (PSF) "globale", che descriva come una sorgente puntiforme viene rappresentata dal sistema. La conoscenza della PSF, in principio, consente di correggere la degradazione di risoluzione e di recuperare le corrette informazioni quantitative; d'altra parte, quest'approccio richiede la conoscenza della PSF in ogni punto del Campo di Vista (FOV) PET, poiché la risposta del sistema non è uniforme all'interno del FOV dello scanner.

In questa tesi è stata proposta l'implementazione di una PSF spazialmente variante nello spazio immagine di un algoritmo Ordered Subsets Expectation Maximization (OSEM) 3D. Due diversi scanner General Electric Medical Systems sono stati considerati, senza (DSTE) e con (D690) la tecnologia Time Of Flight (TOF). La PSF è stata descritta come una funzione Gaussiana, che — grazie alla simmetria cilindrica degli scanner in studio — è stata fattorizzata in una funzione bidimensionale transassiale ed una funzione assiale monodimensionale. In particolare, lungo la direzione radiale la Gaussiana è stata scelta asimmetrica, con una larghezza maggiore verso il centro dello scanner. Di conseguenza, in ogni punto del FOV la conoscenza della PSF coincide con la conoscenza di quattro

parametri di perdita di risoluzione (spread), quello *radiale interno*, quello *radiale esterno*, quello *tangenziale* e quello *assiale*.

Sono state quindi eseguite misure sperimentali per determinare le dipendenze dei parametri di spread della PSF dalla posizione nel FOV dello scanner. Il modo di misurare la risposta del sistema della PET è, infatti, un punto importante. Anche nel caso più semplice, cioè usando una piccola sorgente puntiforme in diverse posizioni del FOV dello scanner, molteplici fattori devono essere tenuti in conto per ottenere una descrizione rappresentativa della risposta del sistema, come le dimensioni della sorgente, il tipo di isotopo usato, i mezzi circostanti (aria, acqua, fondo radioattivo), il numero e la posizione delle misure.

Nel caso qui presentato la simmetria cilindrica degli scanner ha suggerito di considerare dipendenze dalle distanze radiale ed assiale dal centro dello scanner. Per misurare la PSF del sistema è stata usata una piccola sorgente di ^{22}Na (un cilindro con diametro ed altezza di 1 mm incapsulato in Lucite) in aria. Sono state effettuate circa 400 misure per ogni scanner, uniformemente distribuite dentro il FOV. Infine, poiché la PSF deve agire nello spazio immagine di un algoritmo 3D OSEM, è stato considerato importante determinare gli andamenti dei parametri della PSF nelle medesime condizioni, per far corrispondere il più possibile le stime della PSF con l'algoritmo: quindi, lo stesso algoritmo 3D OSEM è stato usato per ricostruire i dati grezzi acquisiti con la sorgente puntiforme.

Da ciascuna immagine ricostruita tre piani ortogonali bidimensionali, passanti attraverso il voxel con l'intensità massima, sono stati estratti lungo tre direzioni (originando i piani radiale-tangenziale, radiale-assiale e tangenziale-assiale): ciascuno di essi è stato poi fittato con una funzione bidimensionale per determinare i parametri di spread corrispondenti.

La funzione di fit ha tenuto conto del post-filtro applicato alle immagini (per ridurre il contenuto di rumore), della posizione *reale* della sorgente puntiforme (per correggere le inevitabili imprecisioni nel posizionamento della sorgente), delle dimensioni della sorgente (per evitare una sovrastima dei parametri di spread dovuta all'approssimazione di sorgente puntiforme) e della discretizzazione intrinseca lungo la direzione assiale a causa dello spessore finito delle fette.

Per ogni parametro di spread, i risultati ottenuti con la sorgente di ^{22}Na sono stati fittati con una funzione bidimensionale delle distanze radiale ed assiale dal centro dello scanner, per determinare le dipendenze analitiche richieste dalla posizione all'interno del FOV.

Il metodo di misura proposto è stato anche validato e ha mostrato la sua buona accuratezza nel costruire il modello di PSF, giustificando il suo uso.

L'implementazione della PSF è consistita in una ridefinizione del proiettore e del retroproiettore dell'algoritmo 3D OSEM. La realizzazione pratica è stata compiuta fattorizzando la PSF 3D in una PSF 2D transassiale ed una PSF 1D assiale. Ciascuna di esse è stata immagazzinata in una matrice tridimensionale,

per tenere in conto la dipendenza dalla posizione nello scanner; il modello continuo della PSF è stato discretizzato calcolando il suo *integrale* per ogni voxel, permettendo una rappresentazione più adattativa per ogni specifico FOV di ricostruzione e dimensione di pixel (la strategia molto più comune di riempire ogni voxel con il valore della PSF nel *punto medio* produce delle deformazioni della funzione contenuta nel kernel, portando a possiibli inesattezze). La dimensione del kernel memorizzato è stata scelta pari a circa 4 volte la risoluzione spaziale degli scanner: questa scelta è conservativa in riferimento alla possibilità di generare artefatti dovuti al troncamento del kernel della PSF. Inoltre, in questa tesi è stata derivata anche l'espressione esplicita per l'operatore trasposto della PSF, mostrando che — nel caso spazialmente variante — questo non coincide con il trasposto del kernel PSF.

La PSF è stata testata su dati da fantocci e clinici. I risultati hanno mostrato migliori accuratezza quantitativa, risoluzione spaziale e qualità di immagine. Inoltre, l'uso combinato di TOF e PSF sembra consentire una mutua interazione positiva, conducendo ai migliori risultati.

Sfortunatamente, un effetto comune delle tecniche di ricostruzione iterative è l'aumento di rumore al crescere del numero di iterazioni, dovuto al fatto che il problema della ricostruzione è intrinsecamente mal posto. Solitamente nella pratica clinica la qualità delle immagini è privilegiata rispetto alla loro accuratezza quantitativa, fermando l'algoritmo iterativo dopo poche iterazioni e di conseguenza lontano dalla convergenza. Questo è vero a maggior ragione se la PSF viene inclusa nell'algoritmo, dal momento che la velocità di convergenza risulta minore rispetto ad algoritmi senza PSF. Un'altra caratteristica importante osservata nelle ricostruzioni con PSF (e frequentemente riconoscibile in letteratura) è l'esaltazione di regioni con ripide transizioni di intensità. In questa tesi è stato dimostrato, per mezzo di simulazioni mono e bidimensionali, che tale effetto è fortemente collegato all'implementazione del recupero di risoluzione spaziale e, anche in presenza di un modello perfetto, inevitabile a meno che un numero di iterazioni elevatissimo (e quindi impratico) venga usato.

È stato dimostrato che le tecniche di regolarizzazione dell'immagine possono essere utili per mantenere il rumore sotto controllo durante la ricostruzione e migliorare i benefici ottenuti tramite l'uso dell'informazione PSF aumentando il numero di iterazioni usate. In particolare, in questa tesi una strategia di regolarizzazione variazionale Bayesiana è stata testata ed impiegata. Due buoni candidati per l'uso in PET sono gli a-priori Huber (o Gauss-Total Variation) e p-Gaussiano generalizzato. Il primo fornisce una buona preservazione di risoluzione spaziale grazie alla componente Total Variation per alti gradienti, ma la componente Gaussiana per i bassi gradienti potrebbe essere insufficiente a controllare zone molto rumorose (come spesso incontrato in PET, in particolare quando il numero di iterazioni viene aumentato per sfruttare l'azione della PSF) a meno che

l'intensità di regolarizzazione venga regolata ad un valore molto alto, ottenendo un'immagine ricostruita innaturale. L'a-priori p-Gaussiano fornisce un'azione di filtraggio molto forte nelle regioni di fondo, originando un buon controllo del rumore, mentre filtra molto meno nelle regioni di segnale.

In questa tesi una modifica dell'a-priori p-Gaussiano è stata proposta per mantenere l'effetto di filtraggio per bassi gradienti (cioè nelle regioni di fondo) e ridurre la perdita di risoluzione spaziale, ma mantenendo transizioni ed apparenza qualitativa "naturali". Un algoritmo 3D OSEM è stato modificato per includere l'a-priori proposto usando un approccio moltiplicativo Maximum A Posteriori One Step Late. Il valore per il parametro p è stato scelto valutando qualitativamente i risultati ottenuti con valori diversi e scegliendo quello che ha prodotto il migliore compromesso tra la soppressione di rumore, la preservazione di risoluzione spaziale ed un naturale aspetto dell'immagine.

Gli a-priori considerati dipendono da alcuni parametri di regolarizzazione. In questa tesi è stata proposta una cifra di merito — che tenesse conto del contenuto sia qualitativo sia quantitativo — per valutare la "rivelabilità" globale di una lesione. La validazione di questo indice di rivelabilità ha mostrato una correlazione molto buona con la risposta umana e ha quindi giustificato il suo uso per assegnare i parametri di regolarizzazione.

I parametri di regolarizzazione sono dunque stati determinati massimizzando l'indice di rivelabilità per ciascun a-priori. Quest'ottimizzazione è stata eseguita per una sfera di diametro 10 mm e 10 iterazioni OSEM.

La validazione delle modifiche proposte è stata quantitativa (su dati acquisiti con un fantoccio NEMA IEC Body Phantom) e qualitativa (su dati relativi a due pazienti oncologici) ed è consistita in un confronto tra gli algoritmi di ricostruzione standard, l'algoritmo proposto, i risultati ottenuti con l'a-priori p-Gaussiano e con la Gauss-Total Variation. Questo confronto ha mostrato un efficace controllo del rumore (ma mantenendo un aspetto naturale dell'immagine) da parte dell'a-priori proposto, con una contemporanea buona preservazione di risoluzione spaziale, contrasto e definizione della distribuzione di attività. Inoltre, l'a-priori proposto si è mostrato anche in grado di mantenere l'artefatto di bordo sotto controllo, riducendo drasticamente le esaltazioni delle transizioni intense nell'immagine. Risultati positivi sono stati anche ottenuti quando la strategia di regolarizzazione è stata usata in unione con le informazioni TOF, suggerendo quindi un possibile futuro utilizzo nell'ambito PET.

Part I

Introduction

Chapter 1

Scope of the work

1.1 The dawn of medical physics

Many among the most important physicians of all times — such as Hippocrat, Leonardo Da Vinci, Andrè van Wesele — used to study human anatomy on corpses, breaking the ethical barrier of death. For many centuries the study of body parts was reserved — as a sort of mystical secret — to medicine students around an autoptic table: *anatomy was strictly linked with death.* It is comprehensible, then, why in 1896 Wilhelm Röntgen's discovery of X-rays originated deep astonishment: for the first time it was possible *for many people* to see (figure 1.1) a bone inside an *alive* human being. It was born what nowadays is called "diagnostic imaging". Only few months, though, were necessary for Henri Becquerel to put another brick in the fundaments of modern medical physics: the discovery of radioactivity of uranium (and, two years later, of radium by Marie and Pierre Curie) paved the way to the analysis not only of the *anatomical* aspects, but also of the *functional* mechanisms of alive creatures.

1.2 Digital 3D anatomy

Although deep improvements were achieved during the first part of the twentieth century in X-ray imaging field (such as intensifying screens, rotating anodes, contrast agents), the "most significant single event in medical imaging since the discovery of X-rays" (as stated in [41]) was the introduction, in the early 1970s, of X-ray Computerized Tomography (CT) by G. Hounsfield and A. Cormack. Figure 1.2 shows the first clinical image acquired by Hounsfield at Atkinson Morley's Hospital on 1st October 1971.

This innovation completely filled the gaps between the postmortem examination anatomy and the "new" diagnostic imaging: indeed, until then all the

Figure 1.1: The first X-ray image [41]

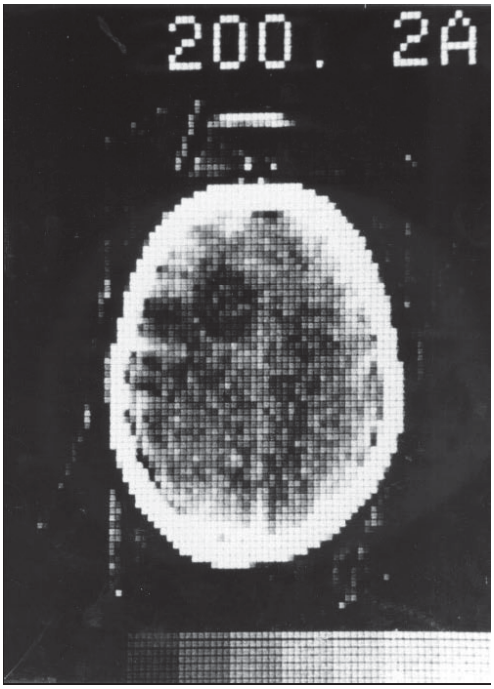


Figure 1.2: The first CT image [12]: note that for the first time in a X–ray exam of the head the skull is not the only visible part

examinations gave a *single projection* of the body on a definite plane, integrating the contribution of all the tissues between the source and the detector. For the first time, the idea of *reconstructing* a volume after acquiring *many* projections was introduced in medicine: the body could now be treated as a three-dimensional object, just like it was done in the ancient time on the corpse under study. But X-ray CT brought a revolution in the digitalisation of the signal, too: the new

informative content consists of some numbers, and the same set of numbers can yield different images depending on how it is represented (e.g. the choice regarding the color scale used to convert the numbers into an image, the so–called *lookup table* or LUT); furthermore, the digitalised images could be stored, sent, received in a much simpler way than before, for example via the Picture Archiving and Communication System (PACS) created in the 1980s.

1.3 The other half of the sky: functional imaging

The radiographic images obtained with X rays, as stated above, gave the physician a dramatically improved knowledge on anatomy. A complete diagnosis, however, needs also a complementary information concerning the *functionality* of the body under study. The origins of *physiology* could be dated back to the studies of William Harvey on blood circulation (1616), but in centuries of medicine the physiologic information could never be collected as an image.

The discovery of *artifical* radioactivity hinted a possible, revolutionary, application: if a molecule of biological interest is bonded to a proper radioisotope and administered to the patient, the distribution of the molecule could be inferred from the distribution of radioactivity in the patient. The production of *radionuclides* at Oak Ridge National Laboratory in 1946 was the first, fundamental step to concretize this epochal idea. The pioneering works by Benedict Cassen and Hal Oscar Anger (figure 1.3, to cite only the most important scientists) in the second half of 1950 completed the foundations of what nowadays is called *nuclear medicine*. The functional imaging was officially born. Since then, γ-emitting radioisotopes have been used in *scintigraphy* for many different applications and districts (some examples are provided in figure 1.4) thanks to its capability of providing a planar projection of the activity distribution.

The combination of different projections and a subsequent reconstruction step are the basis of the extension of scintigraphy to three-dimensional imaging, the Single Photon Emission Computerized Tomography (SPECT), developed by David E. Kuhl and Roy Edwards in the beginning of 1960s. For the first time it was possible to obtain a three-dimensional representation of in-vivo functionality in a human being: an astonishing result.

About ten years later, another milestone in nuclear medicine was to be set. In the 1970s, Tatsuo Ido at the Brookhaven National Laboratory was the first to describe the synthesis of ^{18}F-2-fluoro-2-deoxyglucose (FDG). The new compound — which had the peculiarity of β^+ decay, originating a couple of photons in coincidence and changing the characteristics of the radiation emitted — was first administered to two normal human volunteers by Abass Alavi in August 1976 at the University of Pennsylvania. Brain images obtained with an ordinary nuclear scanner demonstrated the concentration of FDG in that organ. This set the

Figure 1.3: Hal Anger and Benedict Cassen at the International Conference on Peaceful Uses of Atomic Energy in Geneva, Switzerland, in 1955 [93]

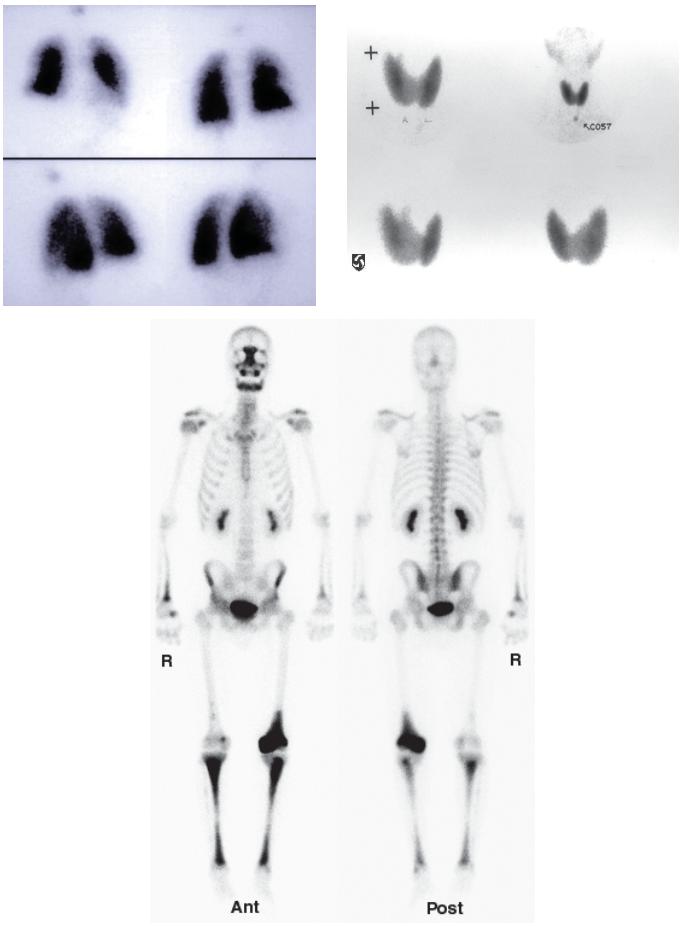


Figure 1.4: Some applications of scintigraphy: imaging of the lungs (top left), of the thyroid (top right) and of the bones (bottom)

groundwork for more in-depth research into using PET to diagnose and evaluate the effect of treatment on human disease.

The first primarily used commercial PET scanner was introduced in 1975 [94] (figure 1.5), even if for the entire 1980s PET was mainly used for research. During the early 1990s, PET expanded into hospitals, diagnostic clinics, mobile systems and physician practices as more and more of the medical community began to realize the utility of PET.

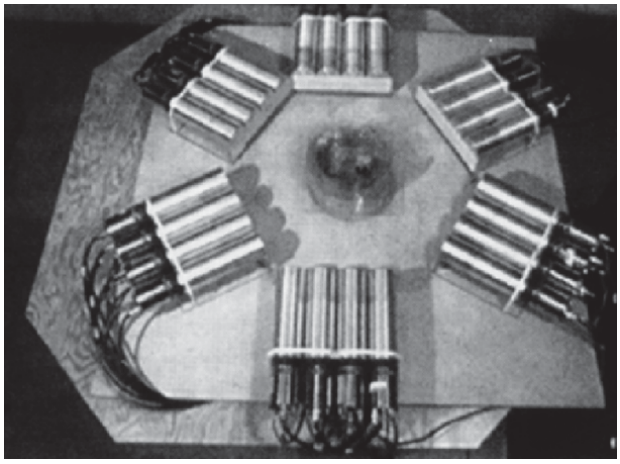


Figure 1.5: The PETT II prototype

For brain imaging — which received a strong improvement from X-ray CT, but which was revolutioned by PET — another important innovation was the introduction of Magnetic Resonance Imaging (MRI) in the early 1980s. Moreover, the implementation of new digital methods produced a terrific increase in sensitivity and quality of all exams.

The availability of different modalities on the same district enriched the quality of the clinical information. In particular, the advantages of complementary anatomical and functional images have long been appreciated, in particular if they are *fused* to represent, on the same image, both of them. Nevertheless, while generally successful for the brain, software approaches often encountered significant practical difficulties with the rest of the body, since the alignment procedures of the different images are generally labour intensive and uncertain of success. Consequently, it is understandable how the proposal of *integrating* an anatomical (CT) with a functional tomograph (PET or SPECT) into a single scanner, which could be able to produce natively coregistered images, was welcomed with great excitement. The proposal came to reality thanks to the work by Bruce Hasegawa from University of California San Francisco (UCSF) on the fusion imaging with SPECT and CT and the first PET-CT prototype by D. W. Townsend and R. Nutt from University of Pittsburgh in 1998 [14, 54, 100].

The relevance of these innovations can be understood by considering that Time Magazine nominated the PET-CT integrated scanner Invention of the Year 2000 [47]:

> *Even apart, PET and CT scanners are triumphs of technology, devices that have saved countless lives, prolonged others, and often made many exploratory operations unnecessary. Yet each has limitations that can lead to uncertainties in diagnosis. By successfully combining the two technologies, Ronald Nutt and David Townsend have eliminated those uncertainties and provided medicine with a powerful new diagnostic tool.*

Moreover, thanks to the revolutionary impact of their invention, Ronald Nutt and David Townsend were awarded the 2010 IEEE Medal for Innovations in Healthcare Technology. PET-CT imaging (see figure 1.6 for an example) is now an integral part of oncology for diagnosis, staging and treatment monitoring.

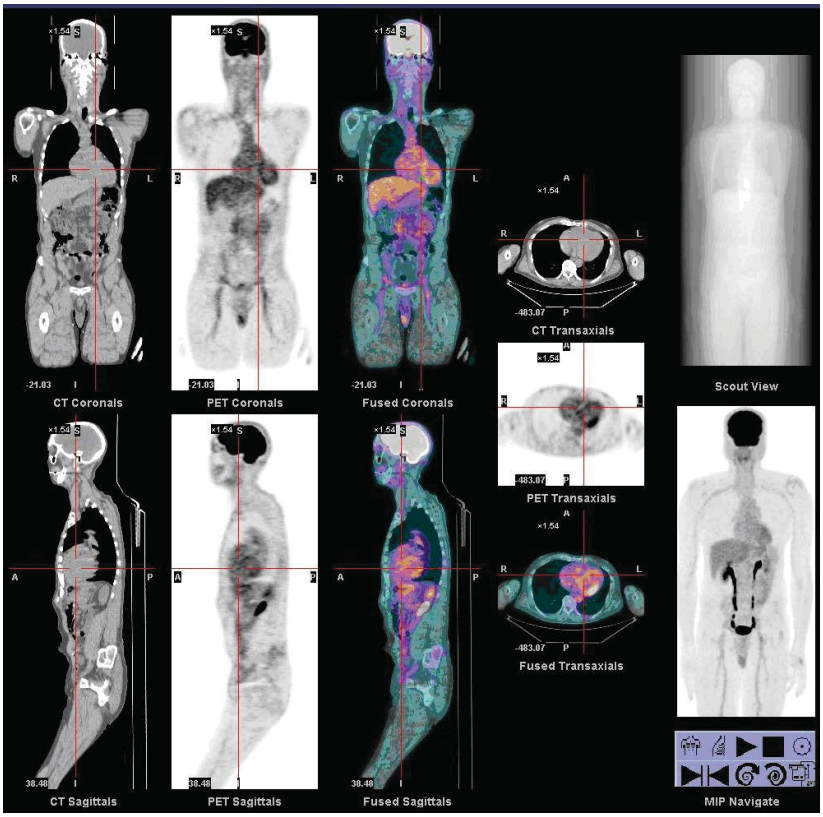


Figure 1.6: Example of PET-CT exam

1.4 Problems and potentialities of image reconstruction

There ain't no such thing as a free lunch, though. Apart from the increased radiation dose (which constitutes an important topic to be assessed when dealing with tomographic images), recostructing an image from its projections is not an easy task. Firstly, an exhaustive knowledge of projection processes is necessary to correctly correlate projections and object. This includes also the necessity to introduce corrections to various physical effects (e.g. scatter, randoms, dead time etc.) Secondly, a robust algorithm is required to avoid image degeneration because of noise: the reconstruction task is nearly always an *ill–posed problem*, i.e. a little fluctuation in the starting data may produce a deeply different image from the real one. Furthermore, the algorithm should not create spurious frequency contents with respect to the frequency spectrum of the collected data.

Since — passing to computerized reconstruction — there is not a unique final result, another key point is to be able to *optimize* the recostruction algorithm. It is not simple to define what optimize means, since this implies the definition of *performance* of the algorithm. Basically, mathematical estimators not always correlate with physicians' criteria of image interpretation.

Anyway, even if reconstructive approach originates the above mentioned problems, also its potentialities extend with respect to non-reconstructive approach. The most important and innovative characteristic is the *flexibility* of reconstruction, i.e. it is possible to define also an *image formation model* incorporating the physical effects linked with the detection of the information.

From this brief introduction it is apparent how the new approach to medical exams needs an interdisciplinary cooperation between different professional figures (physicians, medical physicists, technicians, engineers...) because of the increasing complexity of the hardware and software systems used.

1.5 Structure of the work

This work proposes some improvements to the quality and quantitative accuracy of PET images. In particular, these improvements are the effects of the implementation of the Point Spread Function (PSF) of the tomograph into an iterative reconstruction algorithm (to improve the spatial resolution of the resulting images) and the introduction of a regularization acting on the images (to control the noise increase as the number of iterations increase and to obtain more uniform images). Even if the basic concepts here exposed apply and could be adapted to a general iterative algorithm, in this work the implementations were performed inside a fully 3D PET iterative reconstruction algorithm by General Electric Medical System (GEMS^{TM}). Because of this, chapter 2 explains the main features of Positron Emission Tomography, including the origin of the system response function and of PSF.

Chapter 3 addresses the data acquisition and reconstruction processes. In particular, the difference between analytical and statistical approaches is underlined.

Chapter 4 describes the modelization of PSF, proposes a method to perform the physical measurements to analytically determine the PSF of a scanner across the transaxial and axial field of view and assesses the accuracy of this method.

Chapter 5 presents the computational implementation of the PSF in the reconstruction algorithm and addresses an important effect linked with this implementation, i.e. an enhancement of the sharp transitions in the image.

Chapter 6 introduces the theoretical background of the regularized reconstruction, with particular attention to the variational approach. It also proposes a new variational regularization prior depending on two parameters and a strategy to optimize them.

Chapter 7 reports the results of the proposed modifications obtained with a GE non Time-of-Flight scanner.

Chapter 8 reports the results of the proposed modifications obtained with a GE Time-of-Flight scanner.

Finally, chapter 9 discusses the results obtained and draws the conclusions of the work, while chapter 10 proposes some possible future developments.

In the appendix on page 183, some additional mathematical details are provided.

Chapter 2

Positron Emission Tomography

PET is a nuclear medicine exam relying on the annihilation of anti-electrons (positrons) with electrons and the consequent simultaneous emission of two 511 keV back-to-back photons

$$e^+ + e^- \rightarrow 2\gamma(511\,\mathrm{keV})$$

If it is possible to *coincidentally*[*] detect the two photons, this measure defines a *line of response* (LOR) along which the 2γ-emission event originated. Many LORs contribute to define, then, some *projections* of the object under study.

2.1 Physical principles

Neutron-deficient radionuclides are the possible candidates for PET use. Such isotopes, actually, have two decay possibilities:

$$\begin{array}{ll} \text{Positron decay} & p \rightarrow n + \beta^+ + \nu_e \\ \text{Electron capture} & p + e^- \rightarrow n + \nu_e \end{array}$$

The kinematic condition for the positron decay to happen is

$$\left[M\left({}^{A}_{Z+1}Y\right) - M\left({}^{A}_{Z}X\right)\right] c^2 \geq 2m_e c^2$$

otherwise the radionuclide will decay by electron capture. The more the free energy, the higher probability of the positron decay channel will occur.

The radioactive decay obeys to the decreasing exponential law

$$N(t) = N_0 e^{-\frac{t}{\tau}} = N_0 e^{-\frac{t \ln 2}{t_{1/2}}} \tag{2.1}$$

[*]Within a time interval so short to be approximated by a delta function.

where τ is called *mean lifetime*, while $t_{1/2}$ is the *half-life* of the emitter. From the decay law it is evident that not all the beta–emitters are apt to be used in PET exams: the radionuclide should decay neither too fast (to have enough time to accumulate in the required tissues and to be recorded by the scanner) nor too slow (to record enough counts within a moderate duration of the exam). If the radionuclide is not synthesized in the same facility of the PET exam, another constraint on the lifetime is represented by the required transportation time, forcing to choose longer lifetimes. For these reasons, the most used positron-emitting isotopes in PET imaging are listed in table 2.1.

	^{15}O	^{13}N	^{11}C	^{18}F
Half-life	2 min	11 min	20 min	110 min
Availability	On-site cyclotron	On-site cyclotron	On-site cyclotron	Cyclotron, regional distribution

	^{82}Rb	^{68}Ga	^{62}Cu
Half-life	75 s	68 min	9.7 min
Availability	Generator $^{82}Sr/^{82}Rb$	Generator $^{68}Ge/^{68}Ga$	Generator $^{62}Zn/^{62}Cu$

Table 2.1: Abbreviated list of β^+-emitting isotopes of potential interest in PET imaging [66]

Among these, ^{18}F is definitely the most used because of its "long" lifetime that allows also the transport to medical structures unable to produce it on their own.

All β^+-emitters are included in some pharmaceuticals, administered then — generally via intravenous injection — to the patient. The radiopharmaceutical, accumulating in the organ of interest inside the patient's body, creates a distribution of activity related to some characteristics of the biological process under study. As far as ^{18}F is concerned, it is inserted in a glucose analog, named **2-[^{18}F]fluoro-2-deoxy-D-glucose** (FDG), whose structure is represented in figure 2.1.

The first step of the FDG methabolism in the cell is the same as the one with standard glucose. The second step, instead, requires the presence of two OH groups in position 1 and 2, while in FDG the position 2 is occupied by ^{18}F: this prevents — temporarily, see below — FDG from being metabolized (figure 2.2b) and, consequently, it accumulates in the cells proportionally to the cellular activity (Warburg effect [104]). When ^{18}F decays, it transforms into ^{18}O: the resulting compound is *glucose-6-phosphate*, a product of the first part of normal glycolysis (figure 2.2a). The metabolism of former FDG, then, can proceed as

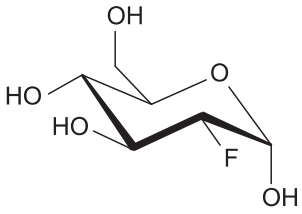
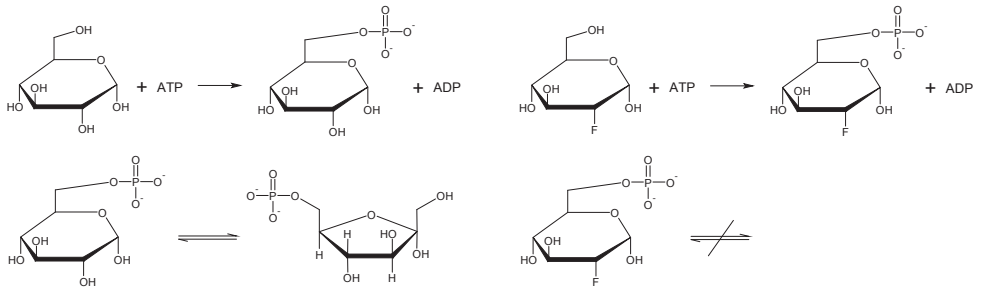


Figure 2.1: Chemical structure of FDG ($C_6H_{11}FO_5$)



(a) Glucose (b) FDG

Figure 2.2: Different metabolism between glucose and FDG

usual in the body.

As an example, in table 2.2 other important radiopharmaceuticals are reported.

$H_2{}^{15}O$	Functional brain studies (rCBF)
$C^{15}O_2$	Functional brain studies (rCBF)
CO	Cerebral blood volume studies
^{11}C-tyrosine	Amino acid for brain studies and oncology
$^{13}NH_3$	Ammonia for blood flow studies in the heart
^{11}C-raclopride	Dopamine receptor system, Parkinson's Disease
^{18}F-DOPA	Dopamine receptor system, Parkinson's Disease
^{11}C-acetate	Cardiological studies

Table 2.2: Some of the non-FDG radiopharmaceuticals used in PET imaging

A positron is emitted from the place of the radiopharmaceutical decay. This positron, travelling across the tissues, loses its kinetic energy in the electromagnetic collisions with the atomic electrons and, nearly at rest, either annihilates with an electron of the medium (approximately in $0.2 - 0.4$ ns [71]) or creates a bounded state e^+e^-, called *positronium*. In dependence on the antiparallel or parallel spin orientation, two possible atomic states are possible, respectively

called *parapositronium* and *orthopositronium*. While there is no energetic strong preference for one of the two configurations, the parapositronium — according to QED selection rules — annihilates into two photons with an intrinsic (i.e. in vacuum) mean lifetime $\tau = 125$ ps, while the orthopositronium can only annihilate into three photons with a lifetime $\tau = 140$ ns. This large difference in lifetimes creates a high probability for a positron in an ortho-state to fast para-annihilate with another electron of the medium, producing two γ rays and shortening the "effective" mean lifetime of the orthopositronium to $0.5 - 5$ ns in matter. Consequently, 3γ emission is negligible and all the detectable radiation can be attributed to the 2 annihilation γ quanta, which — according to the momentum and energy conservation laws — are emitted *back-to-back* (i.e. at 180° in opposite direction), each one carrying an energy of 511 keV (equal to the positron and electron rest masses).

The γ pairs are isotropically emitted in the full solid angle of 4π and can be detected by couples of opposite detectors operated in coincidence giving the LORs of the system. To reveal 511 keV photons, a material with high density and high atomic number is preferable. Moreover, since the detection must be performed in coincidence, also timing properties of the detector are important. For these reasons, the scintillating crystals are nowadays the most interesting devices. In table 2.3 the principal crystals used in PET are reported.

Crystal	I	λ_p [nm]	τ [ns]	Z	ρ [g/cm^3]	ε_E [%]	$1/\mu$ [mm]
$NaI(Tl)$	100	410	230	51	3.7	10	29.4
BGO	15	480	300	75	7.1	10	11.2
GSO	30	430	65	59	6.7	8.5	15.0
LSO	75	420	40	66	7.4	10	12.3
$LYSO$	80	420	41	60	7.1	11	12.6
$LaBr_3$	146	358	21	47	5.1	3	21.3

Table 2.3: Principal crystals used in PET: I is the relative intensity (on PMTs) of light emission, λ_p is the peak wavelength of the emitted light, τ is the time constant of the light emission process, Z is the effective atomic number, ρ is the crystal density, ε_E is the energy resolution at 511 keV and $1/\mu$ is the attenuation length at 511 keV [28, 48, 52, 63, 99, 107]

$NaI(Tl)$ was the scintillating material used in the first PET prototypes, thanks to its well-known and diffused production techniques and its high light yield. Having suboptimal effective atomic number and density and being highly hygroscopic, however, it was substituted by a new scintillator, the *bismuth germanate* $Bi_4Ge_3O_{12}$ (BGO), with a density nearly double than $NaI(Tl)$ and a strongly higher effective atomic number. Even if the light emission time is slightly longer than $NaI(Tl)$ and — more importantly — the light yield is sensibly lower,

its overall characteristics appear to fit better to PET: therefore, BGO has become the crystal used in the new generation PET scanners and has remained the traditional choice for years. An alternative to BGO is represented by the *cerium–doped gadolinium orthosilicate* $Gd_2SiO_5 : Ce$ (GSO). This scintillator presents a light yield nearly double with respect to BGO, it is five times faster and has a better energy resolution; on the other hand, its higher attenuation length provides, for equal dimensions of the crystals, lower sensitivity.

Few years ago, a new scintillator — the *cerium–doped lutetium orthosilicate* $Lu_2SiO_5 : Ce$ (LSO) — has been introduced: high density and effective atomic number comparable to BGO ones, light yield intermediate between $NaI(Tl)$ and BGO and a time constant less than a fifth of $NaI(Tl)$ one represent an important improvement in the detection. Unfortunately, it possesses a quite high level of intrinsic radioactivity (since it derives from a rare earth) which increases the random coincidence counts (see later). For production and commercial reasons a modified version of LSO, the *cerium–doped lutetium–yttrium orthosilicate* $Lu_{2(1-x)}Y_{2x}SiO_5 : Ce$ (LYSO), has been developed, substituting some atoms of lutetium with another rare earth, the yttrium. The resulting crystal mantains all the positive features of LSO, with only a slightly longer decay time: as a consequence, its employment in PET scanners is nearly interchangeable with LSO. The percentage of yttrium contained, however, influences its characteristics (in particular, its effective Z and density), originating different types of this scintillator. Both these new scintillators have been substituting the historical BGO and GSO, thanks to their faster response and, consequently, better capability of dealing with high count rates in the field of view (FOV). Moreover, they have renewed an old idea of PET, named Time of Flight (TOF) which is based on the measurements of the difference in the arrival time between the two γ involved in the annihilation process. Finally, other solutions have been proposed very recently and their characteristics are under study [56, 58].

In the current generation of PET scanners, the crystals have dimensions of about 4 mm in transversal direction and $4 - 8$ mm in the axial direction, while their thickness is about $20 - 30$ mm. Crystals are generally grouped into *blocks*, e.g. 6×6 or 8×8, and each block is read by a group (composed of two or four units) of photomultipliers tubes (PMT). Figure 2.3 presents a schematic representation of a crystal block and its relative PMTs. As a similar strategy, the set of individual crystals may be substituted by a single block of scintillator with specific designed cuts. A new perspective in the detector design is to replace PMTs with avalanche photodiodes (APDs) which are compact solid state semiconductor devices. The advantages of APDs over PMTs rely on their small size, lower cost of production and insensitivity to magnetic fields which is very important for the potential development of integrated PET/MRI systems. A disadvantage of the APDs is the limited electronic gain, lower (1000 to 10000 times) than PMTs, thus

requiring additional dedicated integrated preamplifier. To overcome this problem, silicon PMTs have been proposed. Silicon PMTs are types of APD running in Geiger mode allowing gains of the order of 10^6 (as for PMTs) to be obtained. Another solution under deep study consist in the use of *resistive plate chambers* (RPC), based on gas detectors with high temporal resolutions (≈ 300 ps), large dimensions and reduced costs. Their low sensitivity may be compensated by stacking them up, in order to obtain a columnar structure providing high spatial resolution [27].

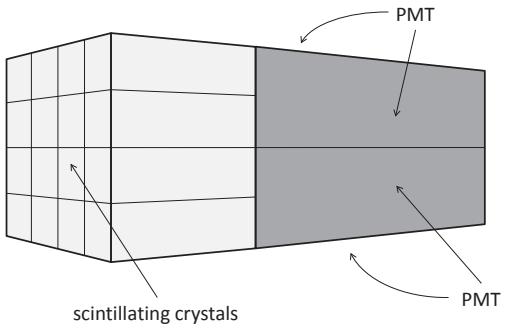


Figure 2.3: Scheme of a crystal block read by PMTs

The scintillator blocks are then disposed in complete rings with a diameter of about $70 - 80$ cm; the number of rings is nearly always 3 or 4 and this number determines the axial FOV — usually about 15 cm — and the total number of planes along z axis. In the axial gaps between two adjacent crystals it is possible to locate some absorbing septa (usually in tungsten) to axially define the image planes (used when performing 2D acquistion): these septa are automatically retracted if 3D acquisition is used (see figure 2.4 for comparison). The circular disposition of crystals leads to an increase in the efficiency in comparison with single-photon acquisition devices: each crystal records LORs with a relatively high number of other crystals, leading to a "fan beam" response. If 3D acquisition mode is used, this is true also for the axial response (changing the response to a "cone beam" one).

Another choice for the organization of the crystals and the readout has been used in some cases: the so–called *pixelated detector* (or *pixelar*) design. In this scheme the crystals are arranged (in a pseudo–continuous matrix) into slightly curved panels, which are optically connected via a continuous lightguide; the relatively big PMTs are cylindrical and disposed into an exagonal pattern (see figure 2.5). A similar organization allows using 5 or 7 PMTs to localize (with an improvement in the positioning of the event), to reduce the light collection dropoff only at the edges of the FOV (with the block scheme, the collection dropoff can happen at each block edge) and to limit the light collection variability across the

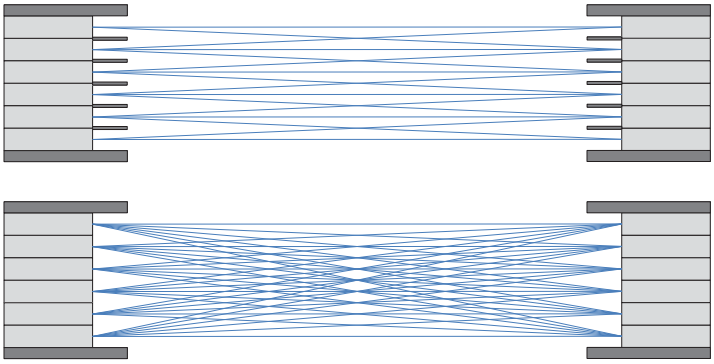


Figure 2.4: Difference between 2D (top) and 3D (bottom) acquisitions due to septa usage

FOV.

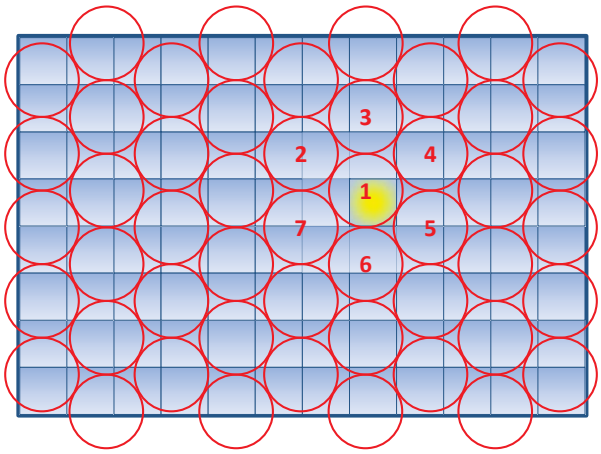


Figure 2.5: Scheme of pixelated detector design

Since in PET the acquisition of multiple peaks is not necessary, broadly[†] only one energy window is used in the readout electronics: typically, the low level is set at $375 - 425$ keV, while the high level is 650 keV. As far as temporal aperture is concerned, a good compromise to detect enough signal from the PMTs and to have a reasonable low contribution from randoms is to accept coincidence in a time window of $4.5 - 9.7$ ns[‡]. If the temporal response of the system is fast enough, it is possible to improve the data acquisition using also a *temporal information* on the photon detection. In fact, considering a LOR coincident with a diameter

[†]Two windows can be used to better analyse diffuse radiation, but their use is not widely spread.

[‡]For more details, see section 2.2.4.

of a crystal ring and assuming that the photon emission happens at a distance x from the center of the detector, the time difference between the arrival of the two photons on the crystals is

$$\Delta t = \frac{1}{c}\left[(d+x)-(d-x)\right] = \frac{2x}{c}$$

Using a scintillator with high timing resolution (such as LSO or LYSO) and adequate readout electronics, it is possible to estimate x from the measure of Δt. This new approach, called *time of flight PET* (TOF–PET), adds a further weight on each event: if in conventional PET every point of the detected LOR has the same probability of being the annihilation source, in TOF–PET it is possible to introduce a probability distribution along the LOR itself and restrict the uncertainty on the positioning of the event (see figure 2.6). This helps in reducing both scatter and, in particular, random contributions (see section 2.2), improving the image quality and the quantitative information [17, 23, 24, 60]. A figure of merit often used is the *signal-to-noise ratio* (SNR)

$$SNR = k\frac{1}{\sqrt{n}}\left[\frac{T^2}{T+S+R}\right]^{1/2}$$

where k is a constant, n is the number of volume elements influencing the noise, T, R and S are, respectively, the total true, random and scattered coincidences in the image. The main difference between conventional and TOF PET is in the factor n.

In conventional PET[§], since each detected event is evenly back-projected in all image elements along the LOR, all voxels contribute to the noise in each image element. Considering, for the sake of simplicity, a cylinder of diameter D and image elements of dimension d, it is possible to estimate $n_{conv} = D/d$.

In TOF, instead, each event is back-projected only in the position associated to such TOF information and into few volume elements adjacent to it, with a weight given by a TOF kernel or probability function of width Δx, representing the spatial uncertainty related to the time resolution Δt of the scanner by $\Delta x = c\Delta t/2$. In this case, a possible estimation of the number of voxel contributing to the noise may be $n_{TOF} = \Delta x/d$. Consequently, the SNR gain introduced by TOF may be estimated as

$$SNR_{TOF} = \sqrt{\frac{D}{\Delta x}}SNR_{conv}$$

It is important to note, however, that TOF–PET generates an about tenfold larger data set if compared with conventional PET, requiring efforts in optimizing the reconstruction task and reducing the computational load (for more details, see section 3.6).

[§]For more information, see chapter 3.

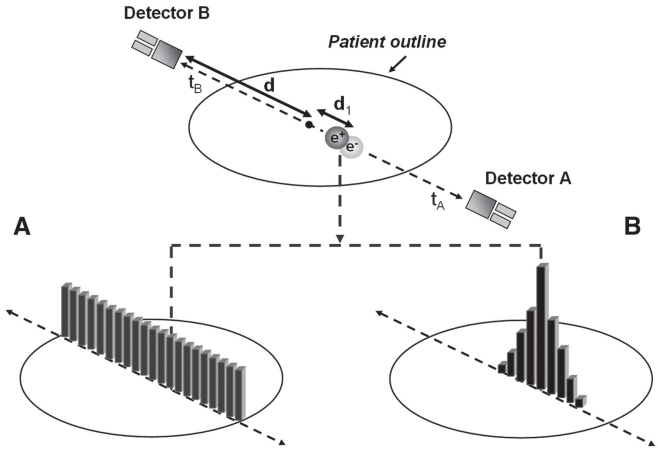


Figure 2.6: Comparison between the positioning of annihilation events in conventional (bottom left) and TOF (bottom right) PET, based on the times t_A and t_B of the detection events

Finally, table 2.4 presents a summary of the most recent PET systems available, with their main characteristics.

2.2 Effects superimposed to the signal due to emission

2.2.1 Poisson statistics

The object under PET study is divided into *pixels* (*picture elements*) — if 2D imaging is used — or *voxels* (*volume elements*) — if 3D imaging is used. Each element[¶] i is expected to emit μ_i events, which constitute the required unknown information. Because of Poisson statistics of the decay process, however, what is actually emitted is n_i events, with a probability

$$P(n_i) = \frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!} \tag{2.2}$$

This introduces an uncertainty source, in particular if quantitative information are needed.

2.2.2 Source decay

If quantitative information on radiopharmaceutical concentration is important, source decay itself is a source of uncertainty. Since the activity A depends exponentially on the time

$$A(t) = A_0 e^{-\frac{t}{\tau}} = A_0 e^{-\frac{t \ln 2}{t_{1/2}}} \tag{2.3}$$

[¶]In the whole text, when referring to voxels, a single index will be used instead of three different ones. This implicitly assumes the introduction of some ordering criteria.

	Det. Mat.	Rings	Crystals	Crystal dim. [mm]	Crystals/block	Axial FOV [cm]	Acq. Mode
GE Advance	BGO	18	12096	$4.0 \times 8.1 \times 30$	6×6	15.2	2D/3D
GE D–ST	BGO	24	10080	$6.3 \times 6.3 \times 30$	6×6	15.7	2D/3D
GE D–STE	BGO	24	13440	$4.7 \times 6.3 \times 30$	8×6	15.7	2D/3D
GE D–RX	LYSO	24	15120	$4.2 \times 6.3 \times 30$	9×6	15.7	2D/3D
GE D–690	LYSO	24	13824	$4.2 \times 6.3 \times 25$	9×6	15.7	3D
Siemens HR+	BGO	32	18432	$4.05 \times 4.39 \times 30$	8×8	15.5	2D/3D
Siemens ACCEL	LSO	24	9216	$6.45 \times 6.45 \times 25$	8×8	16.2	3D
Siemens HiReZ	LSO	39	24336	$4.0 \times 4.0 \times 20$	13×13	16.2	3D
Philips ALLEGRO	GSO	– (pixelar)	17864	$4.0 \times 6.0 \times 20$	– (pixelar)	18	3D
Philips Gemini GXL	GSO–Zr	– (pixelar)	17864	$4.0 \times 6.0 \times 30$	– (pixelar)	18	3D
Philips Gemini TF	LYSO	– (pixelar)	28336	$4.0 \times 4.0 \times 22$	– (pixelar)	18	3D

Table 2.4: Summary of the most recent PET systems available with their main characteristics

if the initial activity A_0 (or concentration of activity) in a *region of interest* (ROI) — or in a *volume of interest* (VOI) if three-dimensional information is used — should be found from the total counts inside it, equation 2.3 must necessarily be taken into account to obtain the correct result.

2.2.3 Positron energy and range

Positron energy is another factor of resolution worsening. Firstly, the positron is emitted with a non–zero initial energy. This implies that it will lose its energy in many collisions with the electrons of the traversed medium. Since the annihilation happens nearly at rest, the effective[‖] distance from the radioisotope decay — the *positron range* — introduces a shift between the real and the recorded origin of radiopharmaceutical decay, which is the one deduced from the annihilation place. The positron range obviously depends on the particle initial energy: since the positron energy spectrum is not a delta function — the β^+ emission is a three–body decay — also the positron range will follow a distribution generally peaked at low distances. Table 2.5 lists the principal PET radioisotopes, their end–point energies and the corresponding contributions of positron ranges in water to total resolution.

Isotope	Maximum positron energy [MeV]	Modal positron energy [MeV]	Positron range in water (FWHM) [mm]
^{11}C	0.96	0.386	1.11
^{13}N	1.19	0.492	1.4
^{15}O	1.72	0.735	1.5
^{18}F	0.64	0.250	1.0
^{68}Ga	1.89	0.836	1.7
^{82}Rb	3.36	1.524	1.7

Table 2.5: End–point energy, modal energy and positron ranges for commonly used PET radioisotopes [63, 72]

Moreover, the $\gamma-\gamma$ emission can happen not completely at rest, either because the positron annihilates with an electron before losing its entire kinetic energy or because the positronium center of mass is not at rest: in both cases, the conservation of energy and momentum forces a non perfect collinearity between the two photons (their separation angle is widely reported as $180° \pm 0.5°$). This leads to a wrong reconstruction of the LOR and, consequently, to an incorrect positioning of the β^+ decay.

[‖] Since positrons have the same mass of electrons, their path in the matter is tortuous: consequently, the "integrated" covered distance (the *mean free path*) does not correspond to the straight distance from the emission point to the rest point.

2.2.4 Random coincidences

If two photons, emitted by different nuclei, reach contemporarily (i.e. inside the considered temporal window Δt, usually of the order of some nanoseconds) two detectors, a fake LOR is created: these counts are consequently superimposed to the real count rate, originating more events than the real ones. Random coincidences are considered to be uniform in space and their rate is [52]

$$R_{rand} = 2\Delta t R_1 R_2 \tag{2.4}$$

while R_1 and R_2 are the singles count rate in the two detectors: as evident, then, the random rate has a quadratic dependence on the source activity. This relation is true if the singles rate is much larger than the rate of coincidence events and if dead–time (see section 2.4.2) can be ignored.

2.2.5 Scattered coincidences

Compton scatter is the predominant effect for 511 keV photons in water and light tissues. When a photon (with energy $E = h\nu$) impinges on an atomic electron, it is scattered at an angle θ (see figure 2.7) from its original flight direction. To fulfil energy and momentum conservation, the new energy of the photon is [22]

$$E' = \frac{E}{1 + \frac{E}{m_e c^2}(1 - \cos\theta)} \tag{2.5}$$

where m_e is the electron mass $m_e = 511$ keV$/c^2$. This dependence is plotted in figure 2.8. If a photon is scattered inside the patient's body, the two γ–rays will be obviously no more back–to–back. The scattered one, anyway, possesses an energy different from 511 keV: if its energy is correctly measured, this photon will be recognized as scattered and, consequently, the coincidence will not take place. Unfortunately, due to crystal energy resolution, some of these "wrong" photons may not be distinguished from the response to a 511 keV photon: in such a case, a rotated LOR will be generated. Scatter contributes, then, to loss of resolution and distortion of the images, particularly in case of 3D acquisition since the tungsten septa are not present.

2.2.6 Attenuation of photons in the matter

Photons interact with the matter via three processes: photoelectric effect, Compton scatter and pair production (absent for energies below $2m_e c^2 = 1022$ keV). Because of these interactions, an exponential attenuation of photon number takes place:

$$N(x) = N_0 e^{-\int_0^x \mu(t)dt} \tag{2.6}$$

where μ is called *linear attenuation coefficient*. Consequently, not all the emitted photons are detected, because of the interactions along the whole path from the

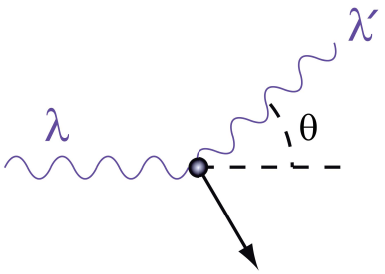


Figure 2.7: Compton effect: the incoming photon (with wavelength λ) is scattered by an electron at rest (circle): as a result, the photon changes its wavelength (λ') and direction and the electron gains kinetic energy (black arrow)

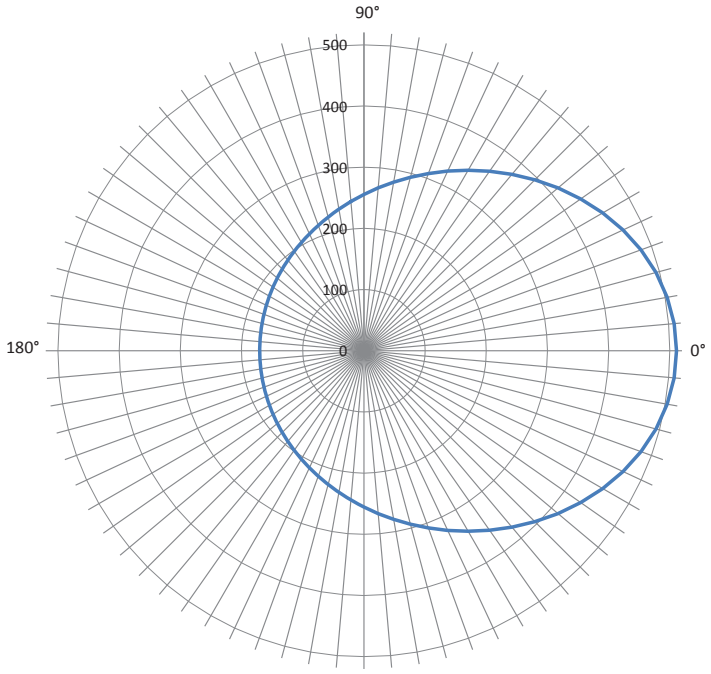


Figure 2.8: Energy (in keV) of a scattered photon (for an incident photon with energy 511 keV) as a function of the scattering angle — the angle between two adjacent angular divisions is 5°

decay place to the crystals. Moreover, since these effects are linear in the medium density, it is usually defined a *massive attenuation coefficient* $\mu' = \mu/\rho$ dependent only on the tissue characteristics (mainly effective atomic number). With this definition, the attenuation law changes into

$$N(x) = N_0 e^{-\int_0^x \mu'(t)\rho(t)dt}$$

showing that higher medium densities correspond to stronger attenuations.

2.3 Corrections to emission effects

Some of the effects related to emission are usually corrected directly during the acquisition phase or during the reconstruction phase.

2.3.1 Randoms

If the acquisition system is capable of registering the singles count rate on each detector, the effect of randoms can be corrected directly using equation 2.4. Otherwise (and more frequently) a second coincidence window is used: this temporal window, delayed — about 100 ns — in order to exclude the detection of real coincidences, allows the direct measurement of random events, which can be subtracted from the total registered counts.

These procedures, however, can lead to negative values in the image or, more importantly, can invalidate the description of the emission process in terms of Poisson statistics (since the difference of two Poisson distributions is no more a Poissonian). In some reconstruction algorithms, instead, the randoms are used *summed*, eliminating the negative value problems and bypassing the Poisson statistics spoiling (for more details, see section M.1). It is important, however, that the delayed window measurement has the lowest noise level possible, to minimize the noise contribution of random coincidences to the data: this can be achieved, for example, using a randoms variance reduction method [9].

2.3.2 Scatter

Scatter correction is definitely the most critical compensation, because Compton scatter in a clinical exam can reach $50 - 60\%$ of total data. In 2D mode the sensitivity to scattered coincidence is lower than in 3D mode and initially many workers ignored it completely. When quantitative information started to be required and, more importantly, when 3D–mode capable scanners were introduced, accurate scatter correction methods became fundamental. Many schemes have been proposed.

Convolution–subtraction techniques assume that the scatter can be modelled as the (generally two-dimensional) convolution of the unscattered (pure) component of data with an empirically derived scatter kernel function. Examples of

such a function are a spatially–invariant decreasing exponential $e^{-a|x-x_0|}$ [11] or its spatially–variant version $e^{-a(x_0)|x-x_0|}$ [13, 61]. Usually, the function parameters are measured on the sinogram (see section 3.2) only in the direct planes (i.e. coincidences generating a LOR with the extremes belonging to the same crystal ring) and the correction is subsequently applied to the whole data set, subtracting the scatter sinogram to the recorded one.

Montecarlo modelling techniques [57] utilize a 3D reconstructed image volume as the source intensity distribution for a photon–tracking Montecarlo simulation; each voxel content is assumed to represent the isotope concentration at that location of patient's body. The program, then, follows the history of each photon and its interactions in the scatter medium, allowing also to trace escaping photons to the detector gantry in a simulated 3D PET acquisition, provided that a model for the used scanner (dimensions, crystal disposition and characteristics etc.) has already been implemented. The interaction coefficient μ_{int}** is either assumed constant across the whole volume under study or, more recently, from the attenuation information on the volume (e.g. with a CT) a segmented image is derived and previously known values of *interaction* coefficients are used dependingly on the tissue characteristics (but this latter is a more computationally expensive approach).

Direct measurements have also been proposed [19]. Performing a short (2D) acquisition with septa in place prior to the normal 3D acquisition, correcting the 2D acquisition for the difference in efficiency with respect to 3D mode and subtracting this corrected measure from the complete acquisition, it is possible to estimate the scatter contribution. Since scatter varies slowly with space and quite independently from activity distribution, the approximation of extending the information from the direct planes also to the indirect planes is quite acceptable. Even if this approach allows taking into account also scatter from outside the axial FOV, this does not apply when multi–bed acquisitions are performed; more, it is crucial to have the attenuation and efficiency measured very precisely, otherwise spurious structures can lead to bad estimation of scatter.

Double energy window [39] approach consists in using an additional energetic window (the "Compton" window) at lower energies with respect to coincidental one. The counts inside the Compton window, corrected for the different efficiencies, are a measure of the scatter coincidences occuring in the measure. This method relies on the assumption that the spatial distribution of scattered coincidence is independent of energy. The correction is generally performed using a Montecarlo simulation, which allows "calibrating" the method and applying it on the patient's study. Also multiple Compton windows have been proposed [86] to better estimate the source distribution and object size dependences.

**As for attenuation, an equation identical to 2.6 describes the probability of the photon *interaction*: the only difference is that μ is substituted by the greater coefficient μ_{int}.

Although these and other methods have generally shown a good assessment of Compton effect, the most used methods nowadays are the *Gaussian fit* and, more frequently, the *model–based scatter correction* algorithms.

The Gaussian fit approach [90] assumes that the scatter distribution can be fitted with a Gaussian function. In each projection planes, the tails of the registered spatial distribution of counts are used to extrapolate a Gaussian distribution, which is then subtracted from the total counts (or used in a subsequent reconstruction step). This method works efficacely if the object is present in a small portion of the FOV (e.g. in brain studies), but fails if there is not enough statistics to robustly fit the tails. Additionally, it assumes a perfect and previous subtraction of randoms, which is not always assured (e.g. it is not compulsory to subtract randoms prior to reconstruct the image, but they can be considered during the reconstruction itself).

The model–based scatter correction algorithm [46, 69] uses the attenuation map (obtained from a PET transmission scan or from CT) and the emission data to calculate the percentage of photons falling on each detector, applying the indipendently known scanner geometric and acquisition model. This calculation is performed with the *Klein–Nishina* formula

$$\frac{d\sigma}{d\Omega} = \frac{{r_0}^2}{2}\left(\frac{E'}{E}\right)^2\left(\frac{E'}{E} + \frac{E}{E'} - \sin^2\theta\right) \tag{2.7}$$

(where E' is defined in equation 2.5), multiplied by the atomic number of the medium Z to account for the Z electrons per atom. A polar plot of the Klein–Nishina distribution is presented in figure 2.9.

The original data contain scatter, so this correction has to be applied iteratively. The calculation gives directly the number of *single* Compton scatterings, while *multiple* Compton events are modeled as a linear transform of the single scatter distribution. This method is highly accurate if the whole activity is in the FOV, but fails if scatter events enter the FOV from outside. Moreover, for a long time only planar scattering was considered due to computational costs and model complexity. Only in the last years, 3D (multi–planar) scattering corrections have started to be proposed and implemented.

2.3.3 Attenuation

It is possible to detect a pair of photons only if both of them are not absorbed along their paths from the emission point to the detectors. The probability for a photon not to be absorbed is described by $P(x) = e^{-\int_a^x \mu(t)dt}$: if a is the distance of the emission point from the scanner center and considering a circular (with radius r) attenuating medium, the probability that both photons will reach detectors is the product of the single probabilities (since the two events

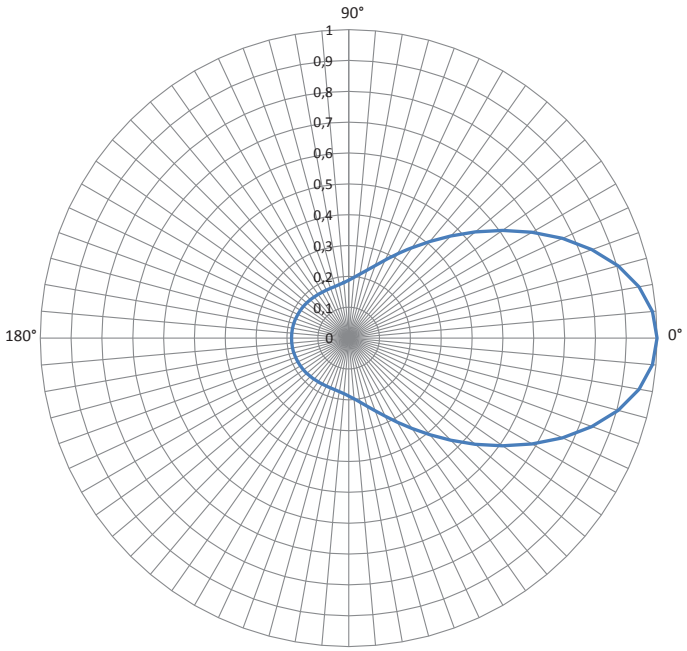


Figure 2.9: Polar plot of the Klein–Nishina distribution for an incident photon with energy 511 keV — the angle between two adjacent angular divisions is 5°

are independent):

$$P_a = e^{-\int_a^r \mu(t)dt} e^{-\int_{-a}^r \mu(t)dt} = e^{-\int_{-r}^r \mu(t)dt}$$

where $\mu(-t) = -\mu(t)$ has been used[††]. As evident from this relation, the probability that the photons reach the detectors is not dependent on the emission point, but only on the thickness of the traversed medium. This interesting property is the basis of the attenuation correction: if the (integrated) attenuation coefficient of the medium is measured along each possible LOR, it is then possible to introduce a correction for the entire set of recorded data.

In a first time, the attenuation coefficients were calculated via a transmissive scan with a β^+ source (usually ^{68}Ge) rotating in the FOV around the scanner axis and about half–way between the patient and the crystal blocks. This scan was composed of two sub–measurements: a *blank* scan (without the patient) and a *pure transmissive* scan with the patient already in position for the PET exam. The first measure provided I_0, while from the second measure the attenuted intensity I was obtained: their ratio $I/I_0 = e^{-\int \mu(t)dt}$ (calculated for each LOR) was the desired integrated coefficient. In 2D this method allowed an efficaciuos rejection of scatter and randoms by accepting coincidence only if the two detec-

[††]Here $(-t)$ represents the path inverse with respect to the attenuation direction: consequently, $\mu(-t)$ should produce an exponential *increase* in the number of photons.

tion points on the crystals and the position of the source itself were collinear. However, since these measurements were initially performed before injecting the radiopharmaceutical, the method implied a substantial increase in the duration and complexity of the exam (e.g. the patient had to be accurately repositioned after the injection and to remain steady for the whole duration of the attenuation scans and PET exam). Significative improvements were obtained when the attenuation scans could be taken after the injection (post–injection transmission scan) of the tracer.

In 3D the method of the trasmission measurements is not generally applicable, since the count rate is much higher and introduces an important amount of dead time. Then, ^{68}Ge was substituted with a collimated source of ^{137}Cs: this radioisotope emits single 662 keV photons[‡‡], making it possible to discriminate them against the annihilation radiation. The attenuation scan can be executed just before, during (alternating the two measures) or immediately after the PET exam, without the need of patient repositioning. Being in 3D acquisition mode, however, the detected scatter is significantly high: the best method to overcome this effect is to use a segmentation technique* to the images (not corrected for the scatter) and choose for the attenuation previously known values of μ factors depending on the different tissues.

Another approach is to use a X–ray tube and an adequate detector: this technique allows having, in a shorter time, more defined attenuation images. If these attenuation information are derived from CT, some empirical conversion functions (depending on CT energy) are used to translate Hounsfield numbers into μ coefficients at 511 keV: an example is presented in figure 2.10. With the introduction of the combined PET–CT systems, the anatomical and functional information are contemporarily available and automatically coregistered (i.e. for each point in one image it is possible to individuate the corresponding point in the other): therefore, the mismatch due to different positioning of the patient during CT and PET is avoided, with a consequent increase in the precision of the attenuation correction.

2.3.4 Source decay

After the image reconstruction, each voxel contains the number of decays originating from the corresponding point in the patient's body. If the source did not decay, from this information it would be possible to directly calculate the corresponding activity. Actually, for quantitative measurements it is necessary

[‡‡]This means that the attenuation coefficients have to be rescaled to the 511 keV energy.

*To segment means to categorize each part of the body under study in one of the tissue classes previously chosen and for which quantities of interest are known independently from the exam.

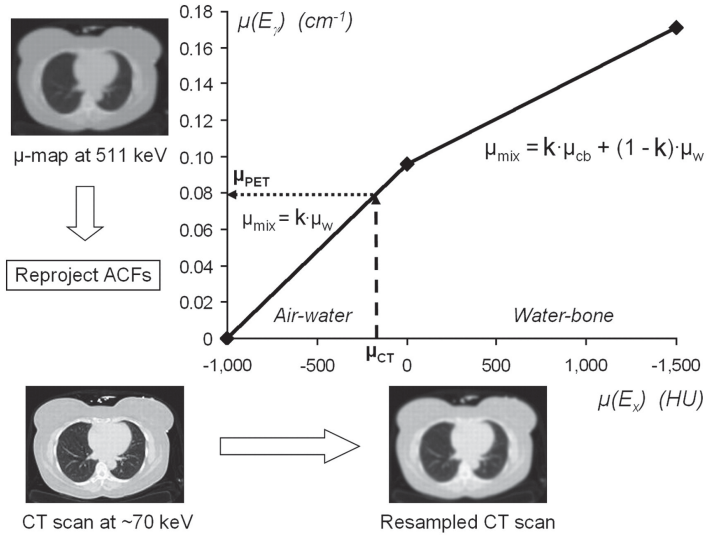


Figure 2.10: Example of convertion from Hounsfield units (HU) into attenuation coefficients at 511 keV energy ($\mu(E_\gamma)$)

to correct for the source decay to obtain the activity at a reference time, usually chosen as the time of the exam start.

The total number of decay events in voxel b (which — in the approximation of complete correction of all other effects — are also the reconstructed events in voxel b) is

$$N_{tot}^{(b)} = \int_{t_0}^{t_f} A^{(b)}(t)dt = A_{t_0}^{(b)} \int_{t_0}^{t_f} e^{-\lambda(t-t_0)}dt$$

where $\lambda = \ln 2/t_{1/2}$ is the time constant typical of the used radioisotope, $A_{t_0}^{(b)}$ is the activity in voxel b at the exam start time t_0, while t_f is the end time of the acquisition. By integration, it is then possible to express $A_{t_0}^{(b)}$ as a function of the reconstructed events $N_{tot}^{(b)}$:

$$A_{t_0}^{(b)} = \frac{\lambda}{1 - e^{-\lambda(t_f - t_0)}} N_{tot}^{(b)} \tag{2.8}$$

Quite often, however, the object of the study has a length higher than the axial FOV: in these cases, the acquisition is performed with more than one *bed*, i.e. dividing the object into consecutive parts as long as the axial FOV and partially superimposed one another. The complete image will consequently be the merging of the frames. If multi–bed acquisition is used, it is necessary to correct each bed dependingly on its own acquisition start moment. If t_0 is the *exam* start time and t_i (t_f) is the start (stop) time of the acquisition on the *bed*

to be corrected, the total number of events in voxel b is

$$N_{tot}^{(b)} = \int_{t_i}^{t_f} A^{(b)}(t)dt = A_{t_0}^{(b)} \int_{t_i}^{t_f} e^{-\lambda(t-t_0)}dt$$

and consequently equation 2.8 changes into

$$A_{t_0}^{(b)} = \frac{\lambda e^{\lambda(t_i-t_0)}}{1-e^{-\lambda(t_f-t_i)}} N_{tot}^{(b)} \tag{2.9}$$

Knowing t_i, t_f, t_0 (which are usually referred to the production time $t = 0$) and λ (which depends on the used radiopharmaceutical), it is therefore possible to correct the data as they were originated by a *temporally uniform* activity equal to the one at the start of the PET scan.

Similar considerations are performed for *dynamic* acquisitions, in which the same object is acquired at different times to study the temporal evolution of the activity distribution. This modality originates different *frames*, each corresponding to a different time interval. To obtain coherent information, each frame should be corrected for the source decay properly.

2.4 Effects superimposed to the signal due to acquisition

2.4.1 Energy resolution

Photons, being neutral particles, need to transfer their energy to charged particles (electrons) in order to be detected. At 511 keV two processes are allowed: photoelectric effect and Compton scatter. While in the first case the photon completely releases its energy to the electron[†], in the second case the electron receives only part of the incident photon energy E:

$$E_{e^-} = E - E'$$

where E', defined by equation 2.5, is the energy of the emerging photon. If the detector were very large, also this "missing" energy would be absorbed and the original energy would be correctly estimated: all the 511 keV photons would accumulate in the so called *full energy peak*, well distinguishable from the Compton shoulder. Unfortunately, both for logistic constraints and to obtain an adequate spatial resolution, single crystals are small enough to permit some photons to escape. Therefore, adding that the incident radiation is not monoenergetic due to the scatter events in the patient [112], a typical energy spectrum is shown in figure 2.11. The *energy resolution* of the system is defined as the ratio of the full width at half maximum (FWHM) of the full energy peak and the energy value at the full energy peak maximum.

[†]The binding energy of the electron remains in the atom and it is subsequently emitted via an Auger electron or de–excitation radiation, but in both cases this energy is likely to be completely collected.

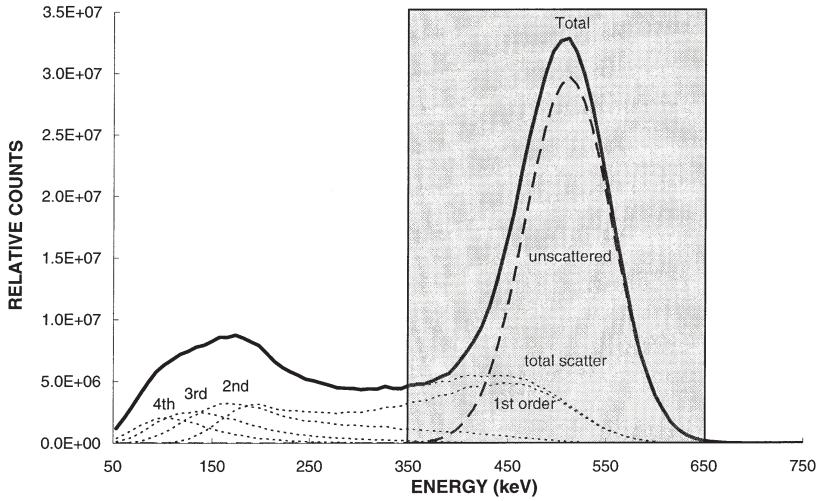


Figure 2.11: Example of a photon energy spectrum acquired in a PET scanner: the spectrum due to primary and scattered photons (solid line) is separated into different contributions (total scattering or different orders of photon scattering) shown by dashed lines

2.4.2 Dead–time

The light signal collected and transformed into an electrical signal by the photomultipliers (PMTs) is passed to a constant fraction discriminator (to extract temporal information to be used in the coincidence circuitry) and a differential discriminator (to sort the pulses according to their peak height). All these analysis steps introduce finite time of processing, during which the system is not able to distinguish an additional event: in fact, quite all digital parts have a *busy logic* able to skip any input during the processing time of an event. This lets the detector behave as *non–paralyzable*. Unfortunately, crystals cannot have a similar protection system. If two light pulses reach the PMT within its resolution time, they will be treated as a unique signal (the so-called *pile–up effect*) whose amplitude is the sum of the two original ones. Since the differential discriminator possesses an acceptance window, this sum event is likely to excess the window high level and, consequently, it will not be considered; unfortunately, each new scintillation event inside the PMT resolution time delays the recovery of full detection functionality. This behaviour is defined *paralyzable*.

At high count rates, then, the system can count much less events than the real ones. In practice, dead–time losses are dominated by pile–up within scintillators — electronics require definitely less time for the analysis than the response time of scintillators: this means that a higher number of crystals corresponds to the possibility of accepting higher rates.

Another effect related to processing time and count rate is the possibility of

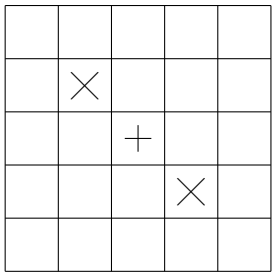


Figure 2.12: Possible wrong evaluation of pulse emission due to multicrystal reading: $\times$ are the crystals scintillating within resolution time of the PMTs, $+$ is the estimated crystal

assigning the recorded light pulse to a wrong crystal due to the block structure. In the case of high count rates, two photons can interact with two crystals inside the resolving time of the detector. Due to the estimation of the position as the centroid of the recorded pulse (*Anger logic*, [5]), the resulting crystal will be approximately equidistant to all the crystals giving response (figure 2.12): this introduces an enlargement of the real LOR.

2.4.3 Crystal efficiency

The scintillation efficiency of the crystals affects the measurement, since the number of emitted photons is proportional to the energy released (*Birks' formula*, [52]). If crystals were all identical, the number of emitted photons would be an exact evaluation of the energy deposition; unfortunately, each crystal (of a same scintillator material) has its own particular characteristics and produces light differently from other crystals. Additionally, ambience variations (such as temperature, humidity etc.) often make the crystal response vary. Also the PMTs affect the efficiency of the crystal block — e.g. if their windows drift — introducing an additional source of uncertainty, since the intrinsic response of crystals and PMTs is required to be as much uniform across the scanner as possible.

2.4.4 Radial geometry

The scanner registers the projections of the object on planes at different angles. However, because of circular disposition of crystals — assuming a perfect detection of photons — the projection bins have different amplitudes, leading to a wrong evaluation of counts towards the scanner transaxial edges, as evident from the figure 2.13. This introduces radial deformations, affecting both the qualitative and quantitative content.

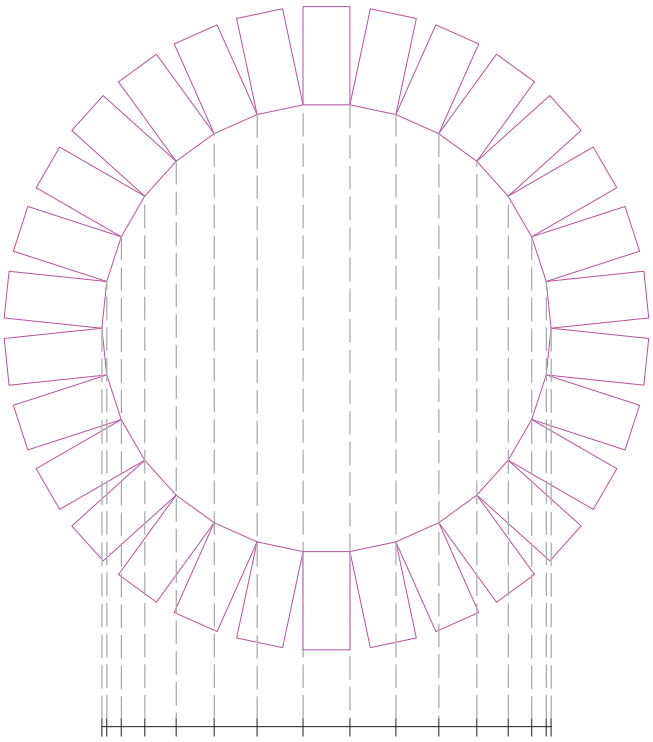


Figure 2.13: Effects on bin amplitude due to radial geometry

2.4.5 Finite dimensions of crystals and depth of interaction

Since the crystals have finite dimensions, what is actually measured is not a *line* of response, but a *tube* of response. As photons can arrive on the crystals following different directions, the attenuation thickness offered by the scintillator will vary consequently [59] — see figure 2.14. This leads to a conversion efficiency dependent on the detector width. Considering detectors in coincidence at a distance much greater than the dimensions of the crystals and a point source at the center of the detection system, the response can be described (in a monodimensional projection) by a triangular function with the vertex at the center of the crystal and null at the crystal edges (figure 2.15): its FWHM is consequently equal to half the crystal width. Moreover, a shorter path inside a crystal makes more probable the release of photon energy in other neighbouring crystals. This effect is related to the so–called *depth of interaction* (DOI) and is more important nearer the detector edges, since the angle between photons and crystals can be quite large (see figure 2.16). In addition, this effect enlarges the objects towards the *center* of the scanner, since the LOR connecting the interested neighbouring crystals (dashed line in figure 2.16) is always nearer to the scanner diameter than the original LOR (black arrow in figure 2.16).

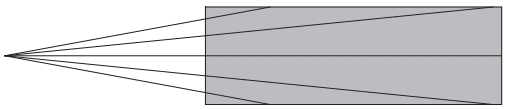


Figure 2.14: Different path lengths inside the scintillator crystals depending on incidence angle of the photon

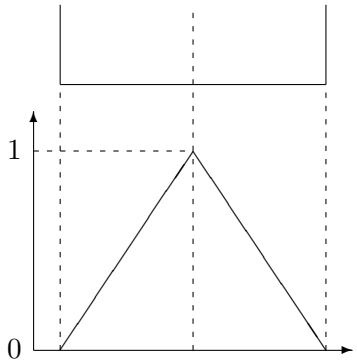


Figure 2.15: Relative response of a scintillator crystal to a point source located on the crystal axis

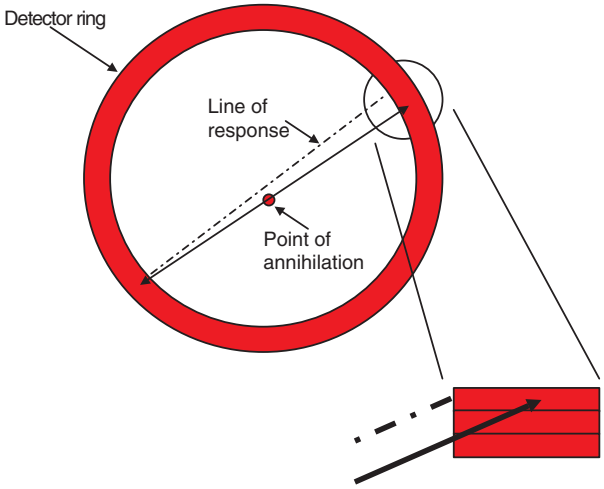


Figure 2.16: Depth–of–interaction effect: note that the spurious LOR (dashed line) falls always nearer the scanner center than the original LOR (black arrow)

2.4.6 Inter–crystal scatter

Although less frequent (because in high–Z materials the photoelectric effect is predominant), scatter events can happen also inside the crystal. If the emerging photon is not completely contained in the scintillator, it can release energy also in one or more neighbouring crystals (figure 2.17), leading to a mispositioning of the LOR and a broadening of the resolution. Nevertheless, the resulting distribution is narrow because scatter at 90° is disfavoured (see figure 2.9) and because the attenuation lengths of the most common scintillators are below 3 cm.

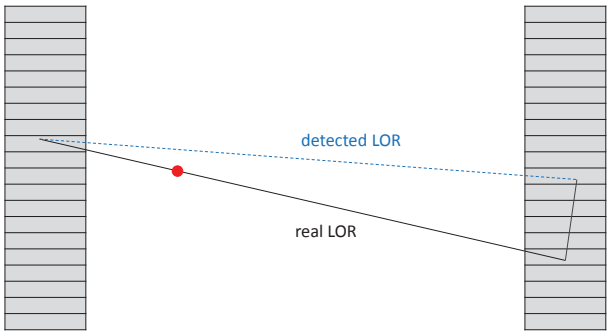


Figure 2.17: Scatter inside scintillator crystals

2.5 Corrections to acquisition effects

2.5.1 Radial repositioning

Each projection should represent the object under study without introducing deformations: consequently, the dependence of the bin width on radial position should be corrected before (or while) reconstructing the image. Basically, this is often performed by a redistribution of counts, filling each bin proportionally to the overlapping parts in the recorded and requested histograms (see figure 2.18). Each manufacturer, however, implements this correction in a different way, dependingly on the scanner geometry.

2.5.2 Dead–time

Dead time losses are usually parametrised as a combination of paralyzable and non–paralyzable contributions. The parameters necessary for the model are obtained by means of experiments involving repeated measurements of a decay source (e.g. see [18]). As far as the mispositioning due to high count rate and multicrystal reading is concerned, some schemes have been proposed (e.g. [8]) but they are generally implemented directly in the parametrization exposed above.

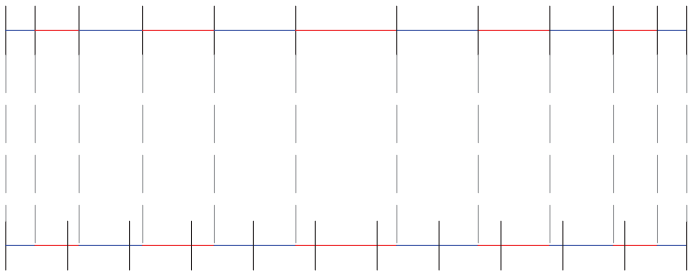


Figure 2.18: General radial repositioning scheme: the upper histogram is the recorded one due to the radial dependence of bin amplitude, the lower one has the bin amplitude corrected by redistribution of the original counts

2.5.3 Detector normalization

The reconstruction algorithms implicitly assume the same sensitivity for all LORs. Because of the motivations exposed in section 2.4, this supposition is not valid: a correction to *normalize* acquired data is then necessary. This correction consists in multiplying each LOR by a *normalization coefficient* (NC).

The most simple way to calculate the set of NCs is to perform a scan where every LOR is illuminated by the same coincidence source: NCs are then inversely proportional to the recorded counts. This *direct normalisation* is subject to some disadvantages. Firstly, scattered coincidences require a different normalisation with respect to true ones, but direct measurements do not allow this distinction. Secondly, 3D mode needs low count rates to reduce dead–time effects: consequently, to accumulate enough statistics a long measure time (tens of hours) is mandatory, introducing a significant practical problem since normalization should be frequently measured to ensure proper reconstruction.

A variance reduction method united to a factorized modelling of NCs [43] allows overcoming both previously highlighted problems. Each normalization factor is modelled as the product of intrinsic crystal efficiencies and some geometric factors accounting for various contributions, e.g. the photon incidence angle or the different gain and response of PMTs. The NCs, then, are no more related to the *LORs*, but to the *crystals*: considering that the number of LORs is generally more than a hundredfold the number of crystals, the factorization method allows a large reduction of measurement complexity, even if this can imply an increase of systematic errors since the NCs are not all independent.

Each of the above factors require calibration measurements (e.g. *geometric correction*, *photomultiplier gain*, *photomultiplier linearity* etc.), which are usually scheduled. If a detector part is substituted, however, it is wise to partially or totally recalibrate the system.

Even if component–based normalization accounts for most of the effects, some

artifacts and deformations are still present in the reconstructed images also after the normalization.

2.6 The Point Spread Function

Due to finite resolution, a point source is detected as if it were larger and with a lower intensity. The blurring of the edges and the reduced intensity of the activity distribution are usually referred to as *partial volume effect* (PVE). A generic region is considered to be subject to two quantitative effects:

spill–out i.e. a reduction in the activity distribution due to counts belonging to the region, but assigned (for resolution effects) to the surroundings;

spill–in i.e. an increase in the activity distribution due to counts belonging to the surroundings, but assigned (for resolution effects) to the region.

The balance of these two effects depends on the lesion-to-background ratio (LBR) — i.e. the ratio between the activity concentrations in the region and in the surroundings — and affects the final result for the activity concentration.

In a general optical system, the Point Spread Function (PSF) is a function describing how a point source is rendered, i.e. it is equal to the impulse response of the optical system. This PSF is generally representable — thanks to the Central Limit Theorem (CLT) — as a two-dimensional (or three-dimensional, if the optical system is able to record 3D data) Gaussian with sigmas related to the detector characteristics.

In a PET system, all the (partially or totally) not corrected effects — see section 2.4 — contribute to the PSF, even if essentially four factors are predominant [63]:

- finite dimensions of crystals, contributing with $FWHM_D \approx 4$ mm
- non collinearity of photons, contributing with $FWHM_N \approx 2$ mm
- positron range, contributing with $FWHM_P \approx 0.5$ mm
- block effects (including both the "packing" effects and the contribution of the single crystals) contributing with $FWHM_B \approx 2$ mm

Additionally, the reconstruction algorithms introduce a multiplicative factor $k_R \approx 1.25$ for the resolution leakage [65]. Consequently, a general formula for the resolution in a PET system is

$$FWHM_{tot} = k_R \sqrt{{FWHM_D}^2 + {FWHM_N}^2 + {FWHM_P}^2 + {FWHM_B}^2}$$

If the reported values are used, the PET system PSF can be described as a Gaussian Function (GF) with $FWHM_{tot} \approx 5 - 6$ mm. Figure 2.19 shows the

trend of the main contributions to PET resolution as a function of the spatial coordinate for a typical configuration of the detector.

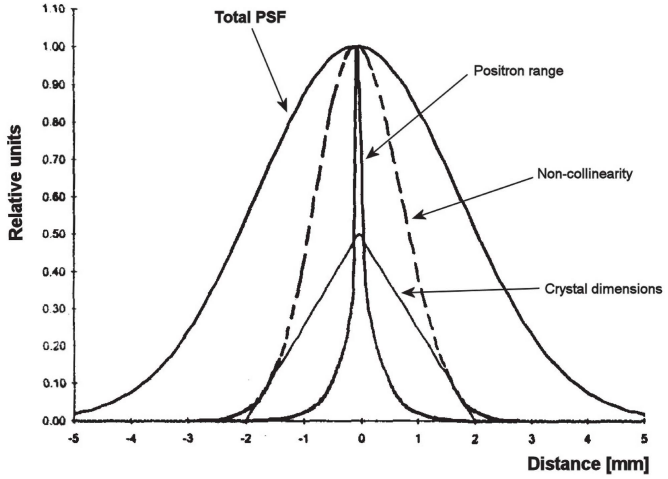


Figure 2.19: Principal components of PET resolution for a system with diameter 80 cm, crystal dimension of 4 mm and using ^{18}F as the radionuclide

The resolution is not uniform across the field of view: nearer the detectors the degradation effects are more likely and stronger and, consequently, the deterioration of resolution is more pronounced. Additionally, if a more accurate analysis is performed, along the radial direction the PSF appears to be progressively asymmetric getting closer to the detector edges, with the larger tail directed towards the scanner center: this deformation is mainly caused by circular geometry. The source radially divides the scanner into two portions (see figure 2.20), whose dimensions depend on the source position: if the source is farther from the center, more crystals will contribute to the internal part of the PSF, increasing the asymmetry. The tangential profile of the PSF, instead, is a symmetric — the source always divides the scanner into two halves — Gaussian with distance dependent width; also the axial contribution is well modelled as a symmetric Gaussian, especially if 3D acquisition mode is used. As an example, figure 2.21a shows the reconstructed image of a source located 20 cm from the transaxial FOV center, while an asymmetric two-dimensional Gaussian is presented in figure 2.21b for comparison.

Signal analysis teaches that the resulting image I will consequently be the convolution[‡] of the PSF with the real activity distribution A:

$$I(\vec{x}) = A(\vec{x}) \star PSF(\vec{x}) = \int_V A(\vec{x})\, PSF\left(\vec{t} - \vec{x}\right)\, d\vec{t}$$

[‡]Strictly speaking, the term *convolution* may be used only in case of spatially-invariant PSF. For spatially-variant PSF, some modifications should be taken into account — e.g. see sections 5.1 and M.3.

If it were possible to *deconvolve* the point spread function from the resulting image, the correct information would be *completely* recovered. Unfortunately, this is not feasible in a classical way (e.g. via Fourier transforms), in particular if 3D mode is adopted, mainly due to the computational requirements of a similar process.

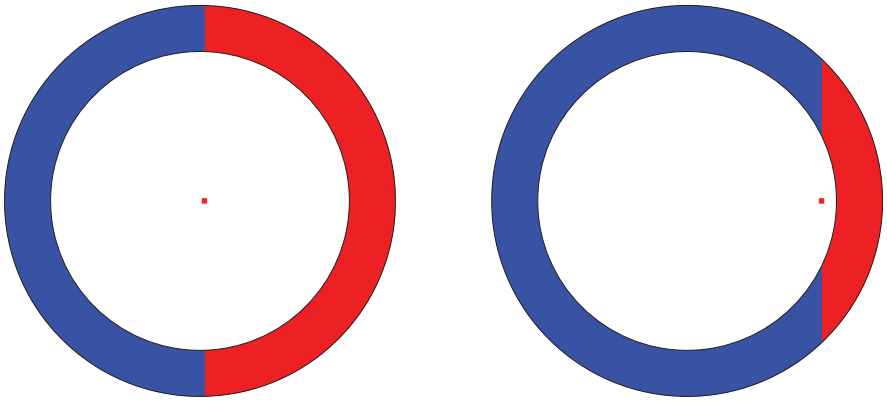


Figure 2.20: Increasing radial asymmetry with higher distance from FOV center: the blue parts contribute to the radial internal part of PSF, the red parts to the radial external one

(a) Source located 20 cm off detector center (b) Two-dimensional asymmetric Gaussian

Figure 2.21: Comparison of a source located off-center and a radially asymmetric two-dimensional Gaussian function

Chapter 3

Image reconstruction

3.1 Object representation

The unknown space–dependent* radiotracer distribution can be described by a continuous function $f(\mathbf{r})$ specifying the radiotracer concentration in the point of space described by the vector $\mathbf{r} = [x\ y\ z]^T$. This continuous function, though, is completely impractical since it would require an infinite sampling: it is necessary to choose a resolution–limited representation using a set of $j = 1 \dots J$ spatial basis functions $\alpha_j(\mathbf{r})$:

$$f(\mathbf{r}) \approx \sum_{j=1}^{J} c_j \alpha_j(\mathbf{r})$$

where $\mathbf{c} = \{c_j\}$ is a J–dimensional vector containing the coefficients relative to the spatial basis functions. The most common choice for $\alpha_j(\mathbf{r})$ is the *pixel* (picture element, if two-dimensional information is available) or the *voxel* (volume element, if three-dimensional information is acquired), for which the vector $\mathbf{c}$ — simply holding the pixel or voxel values — can be directly interpreted as a two-dimensional or three-dimensional image. Generally speaking, if $\{\alpha_j(\mathbf{r})\}$ is an orthogonal basis, c_j can be expressed as

$$c_j = \int f(\mathbf{r}) \alpha_j(\mathbf{r}) d\mathbf{r}$$

Instead, if non–orthogonal functions are used, even for an *a–priori known* function $f(\mathbf{r})$, finding $\mathbf{c}$ is a tough task. In any case, however, the image reconstruction process has the scope of determining the unknown vector $\mathbf{c}$ from the acquired data.

*The eventual additional dependence on time is omitted for simplicity.

3.2 Data representation

As stated in section 2.1, a coincidental (i.e. inside the temporal resolution of the system) detection of photons defines a LOR. The entire set of LORs constitutes the acquired data.

Assume a pair of photons being detected at positions $\mathbf{r}_1$ and $\mathbf{r}_2$[†]. The line connecting these two points forms an angle ϕ with respect to a previously fixed reference direction; moreover, this line has a distance s from the origin of the system, conventionally chosen as the scanner center — see figure 3.1. The point (s, ϕ) uniquely defines an event in the ring plane: the entire set of data will be represented by a cloud of points in a cartesian plane with s as the abscissa and ϕ as the ordinate: this cartesian plane is called *sinogram*, from the sine–function appearance of a point source. The events are consequently binned dependingly on the flight direction and distance from center.

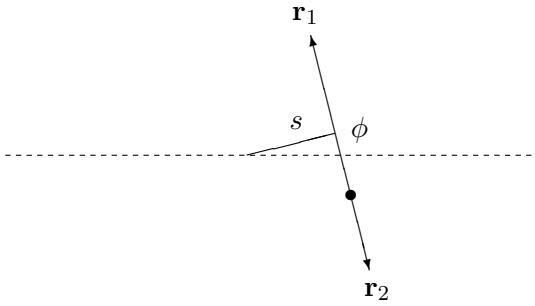


Figure 3.1: Definition of the variables used in the data representation

To retain the complete information available, another format of data representation is the so–called *list–mode*. Each event in this case is stored by recording $\mathbf{r}_1$, $\mathbf{r}_2$ and additional information such as t_1, t_2 (the arrival time on the detectors with respect to a reference time, usually chosen coincident with the start time of the exam), E_1, E_2 (the detected energies) and so on. Obviously, the increase of details reduces the information loss but causes a significant increment of the required disk space. In general, list–mode data are used to archive an exam if further data processing is expected.

A third mode (less used) is to *backproject* the sinogram, translating each point to a line in a 2D or 3D image. This mode allows a great reduction in the storage space required, although it is not suited for further elaboration of data: in this case, in fact, it would be necessary to firstly reproject the image to the sinogram, with increasing computational time and potential errors introduced by the projection–backprojection steps.

[†]For the sake of simplicity, consider a unique crystal ring, defining a plane orthogonal to the scanner axis in the three-dimensional space.

Figure 3.2 presents an example of the three representation methods applied to a point source. Generally, the most used starting point for image reconstruction is the sinogram storage.

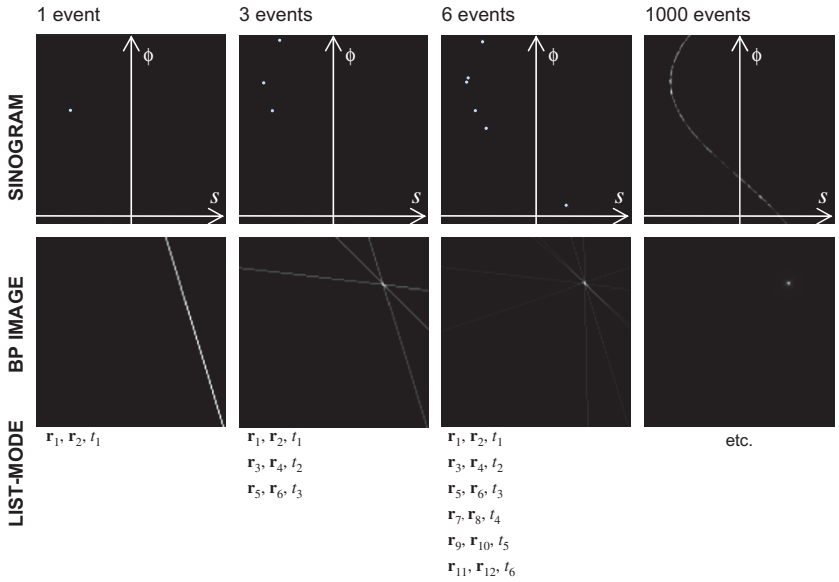


Figure 3.2: Three methods to store data — an example of a point source [80]

The majority of the reconstruction algorithms works on the recorded sinograms (which contain the so–called *raw data*). The sinogram storage is different between two-dimensional and three-dimensional acquisition.

If 2D acquisition is performed (i.e. the scanner uses the septa), the 3D volume in the FOV is divided into $2N_R + 1$ axial slices (where N_R is the number of crystal rings in the scanner). The complete sinogram is then a set of $2N_R + 1$ uncorrelated two-dimensional histograms (s, ϕ) as described above. In the case of 3D acquisition, instead, the storage is generally based on the *projection planes*: the complete information is organized into sinograms taken on many planes, not only perpendicular to the axial direction, but oblique, too. In both cases, however, the mean counts in the sinogram bin (s, ϕ) is expressable as

$$\left\langle m_{(s,\phi)} \right\rangle = A_{(s,\phi)} N_{(s,\phi)} \int\limits_{\mathrm{LOR}(s,\phi)} f(\mathbf{r}) d\mathbf{r} + b_{(s,\phi)}$$

where $A_{(s,\phi)}$ takes into account the attenuation of the photons on $\mathrm{LOR}(s, \phi)$ due to scatter and absorption, $N_{(s,\phi)}$ describes the losses due to detection imperfections and the additive term $b_{(s,\phi)}$ represents the contribution of random coincidences and scattered photons external to $\mathrm{LOR}(s, \phi)$. Consequently, the relation between the image I and the sinogram S can be resumed as

$$S = \mathrm{P}I + B \tag{3.1}$$

where B is a sinogram of background events and the operator P — the *projector* — translates a datum in the image space to an information in the sinogram space. This relation is theoretically invertible, giving

$$I = \mathrm{P}^{-1}(S - B) = \mathrm{B}(S - B) \tag{3.2}$$

where $\mathrm{P}^{-1} = \mathrm{B}$ is the *backprojector*.

3.3 Projector and backprojector

The core of image reconstruction is the modelization of the projector and backprojector. Firstly, it is important to note that these operators act on *discretized* data: consequently, the backprojector is not exactly the inverse operator of the projector (due to computer finite algebra), so two different operators are usually implemented. Generally, the three–dimensional projectors and backprojectors rely on linear monodimensional interpolation and they are the generalization of the two-dimensional ones, here summarised for the sake of simplicity. The most diffuse *discrete* two-dimensional implementations are presented in figure 3.3.

Pixel–driven or *voxel–driven* projector (figure 3.3a), fixed a *focal point*, draws a line connecting the focal point with the center of the examined pixel (voxel) and finds the intersection between this line and the detector. The two bins adjacent to this intersection are filled using a linear interpolation based on overlapping parts of the pixel and the bins. The corresponding backprojector works in a similar way.

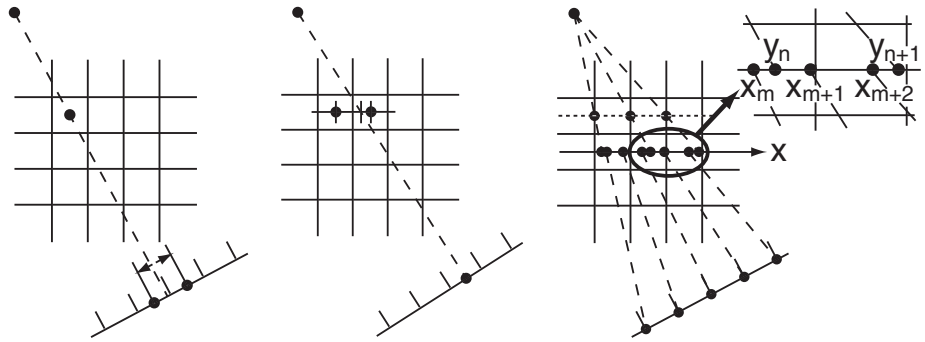
Ray–driven projector (figure 3.3b) is similar to the pixel–driven one, but the line connects the focal point and the center of the sinogram bin. Apart from this different choice, the interpolation works in the same way as pixel–driven projector/backprojector.

Distance–driven projector (figure 3.3c), fixed a *focal point*, maps the boundaries of each pixel and detector cell onto a common axis. The length of overlap on this axis is used as a weight to project or back–project.

This latter method is generally the most used, being fast and because the pixel– and ray–driven projectors often result in high–frequency artifacts. Its three-dimensional version maps the two-dimensional boundaries of the image voxels and detectors onto a common plane: the required weight is then represented by the *area* of overlap on this plane.

3.4 Analytic methods

The most straightforward way of reconstructing an image from its projections is to exploit the real detection process and try to revert it. Many analytic algorithms have been proposed (e.g. [80, 81]). If 3D acquisition mode is employed — even if



(a) Pixel–driven (b) Ray–driven (c) Distance–driven

Figure 3.3: Some of the projectors used in tomography [30]

native three-dimensional analytical algorithms exist (e.g. [51]) — the information is usually reorganized as it were acquired in 2D, translating the projections on the oblique planes into projections on the planes orthogonal to the scanner axis — this method is called *Fourier rebinning* (FORE) (e.g. [31]), since this reorganization is performed in the frequency space. For this motivation and for the sake of simplicity, only 2D inversion is here described.

In the two-dimensional case, the tomographic data acquisition is usually modeled as a Radon transform $\mathcal{R}$[‡]

$$m(t,\theta) = \mathcal{R}\{f\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta\left(x\cos\theta + y\sin\theta - t\right)\,dx\,dy \tag{3.3}$$

So starting from the measured 2D projections (corrected for all the superimposed physical effects), the image reconstruction coincides with the inversion of the Radon transform. The fundamental *Fourier slice theorem* creates a connection between the monodimensional Fourier transform of a projection $\mathcal{F}_1\{m(t,\theta)\}$ and the two-dimensional transform of the original object $\mathcal{F}_2(\omega,k)$:

$$\begin{aligned}
\mathcal{F}_1\{m(t,\theta)\} &= \int_{-\infty}^{+\infty} m(t,\theta)e^{-i2\pi\omega t}\,dt = \\
&= \left[\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{-i2\pi\omega t}\,dx\,dy\right]_{t=x\cos\theta+y\sin\theta} = \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{-i2\pi\omega(x\cos\theta+y\sin\theta)}\,dx\,dy = \\
&= F(\omega\cos\theta,\omega\sin\theta) = \mathcal{F}_2\{f(x,y)\}|_{(\omega\cos\theta,\omega\sin\theta)}
\end{aligned} \tag{3.4}$$

[‡]In the following mathematical steps, the discrete sinogram variables s and ϕ are substituted with the continuous variables t and θ, respectively, to underline the difference between the mathematical theory and the practical implementation.

So, the theorem states that the monodimensional Fourier transform of the projection at angle θ is equal to the two-dimensional Fourier transform of the image evaluated along a line forming an angle θ with respect to the horizontal axis in the 2D frequency space (see figure 3.4). Consequently, if the projections allowed filling the entire two-dimensional Fourier space (i.e. if all the angles $\theta \in [0, \pi]$ were analysed), the inverse two-dimensional Fourier transform would give the original object. Unfortunately, the discreteness of the projections forces interpolations in the two-dimensional Fourier space, especially at high frequencies, generating artifacts. Direct Fourier methods are consequently not popular to invert the Radon transform.

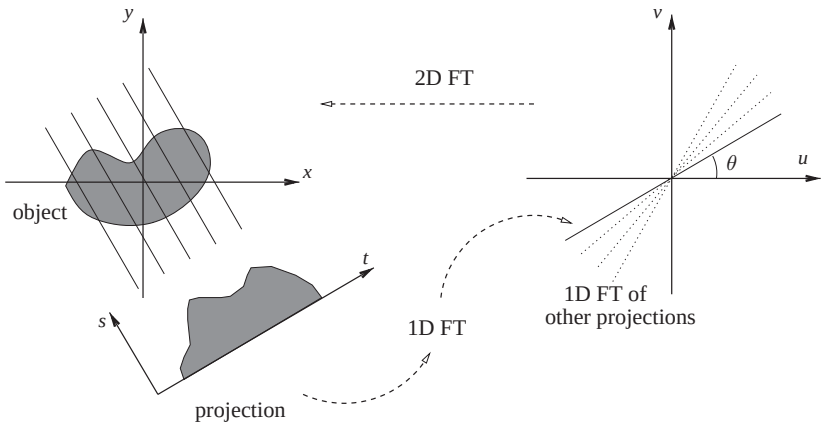


Figure 3.4: Radon transform and Fourier slice theorem [1]

3.4.1 Filtered back–projection

Inverting the Fourier slice theorem (equation 3.4) with the frequency variables expressed in the polar form $u = \omega \cos\theta$, $v = \omega \sin\theta$ and using relation 3.3

$$\begin{aligned}
f(x,y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(u,v)e^{i2\pi(ux+vy)}dudv = \\
&= \int_{0}^{2\pi}\int_{0}^{+\infty} F(\omega\cos\theta,\omega\sin\theta)e^{i2\pi\omega(x\cos\theta+y\sin\theta)}\left|\begin{array}{cc}\frac{\partial u}{\partial\omega} & \frac{\partial v}{\partial\omega}\\ \frac{\partial u}{\partial\theta} & \frac{\partial v}{\partial\theta}\end{array}\right| d\omega d\theta = \\
&= \int_{0}^{\pi}\left[\int_{-\infty}^{+\infty} F(\omega\cos\theta,\omega\sin\theta)|\omega|e^{i2\pi\omega(x\cos\theta+y\sin\theta)}d\omega\right] d\theta = \\
&= \int_{0}^{\pi} \hat{m}(x\cos\theta + y\sin\theta,\theta)d\theta = \mathcal{B}\left\{\hat{m}(t,\theta)\right\} \qquad (3.5)
\end{aligned}$$

it is then possible to obtain the original image as the *backprojection* $\mathcal{B}$ of the *modified* projection $\hat{m}(t,\theta)$:

$$\hat{m}(t,\theta) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} m(t,\theta) e^{-i2\pi\omega t} dt \right] |\omega| e^{i2\pi\omega t} d\omega$$

The process of backprojection can then be synthesised as follows:

$$f(x,y) \xrightarrow{Scan} m(t,\theta) \xrightarrow{FT} M(\omega,\theta) \xrightarrow{\times H(\omega)} \hat{M}(\omega,\theta) \xrightarrow{FT^{-1}} \hat{m}(t,\theta) \xrightarrow{\mathcal{B}} f(x,y)$$

where FT (FT^{-1}) indicates the direct (inverse) monodimensional Fourier transform and $H(\omega) = |\omega|$.

If discrete image and projections are considered, the integrals are substituted by summations and Fast Fourier Transforms (FFT) algorithms can be applied. In this case, however, $H(\omega)$ becomes

$$H(\omega) = \begin{cases} |\omega| & |\omega| < \omega_c \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

which is called the *ramp filter* (represented in figure 3.5): its *cut–off frequency* ω_c is half the maximum spatial frequency representable in an image with pixel size a:

$$\omega_c = \frac{1}{2a}$$

This algorithm is called *filtered backprojection* (FBP) due to the multiplication by the filter $H(\omega)$ before the appliance of the discrete backprojection operator $\mathcal{B}$.

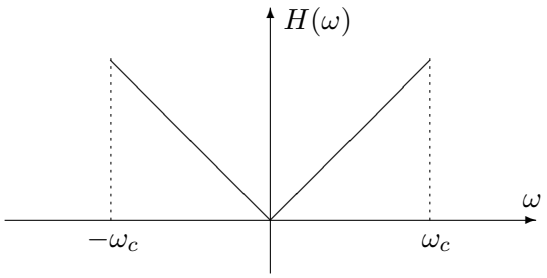


Figure 3.5: Ramp filter

Because the acquisition process is not an ideal one, also noise will be present. Since noise is typically characterized by high frequencies — emphasized by $H(\omega)$ — it is wise to additionally introduce a low–pass filter, multiplying its transfer function $G(\omega)$ before inverting the monodimensional FT:

$$f(x,y) \xrightarrow{Scan} m(t,\theta) \xrightarrow{FT} M(\omega,\theta) \xrightarrow{\times H(\omega)G(\omega)} \xrightarrow{\times H(\omega)G(\omega)} \hat{M}_F(\omega,\theta) \xrightarrow{FT^{-1}} \hat{m}_F(t,\theta) \xrightarrow{\mathcal{B}} f_F(x,y) \tag{3.7}$$

Some filter examples are the *Hamming* filter, the *Butterworth* filter and the *Wiener* filter [33, 37]: for these and others filters, $H(\omega)G(\omega)$ versus $a\omega_c$ is represented in figure 3.6.

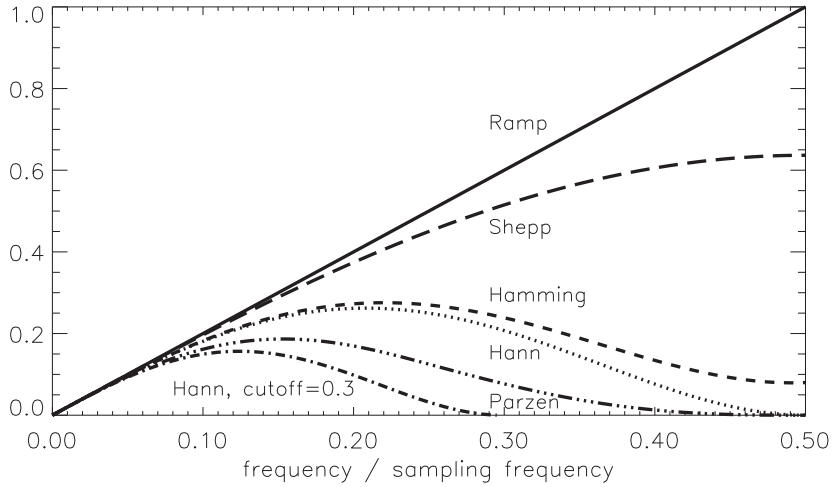


Figure 3.6: Common filters used in FBP [1]

FBP is a very fast and simple algorithm because of the exploitation of FFT computational capabilities. Unfortunately, two main drawbacks should be emphasized. Firstly, FBP requires a precorrection of all the factors superimposed to the real data (see equation 3.2): this means that, before applying the algorithm, it is necessary to perform a subtraction on the sinogram. A similar process very often yields negative values in the final image, which do not represent physical values and thus cause errors in the quantitative evaluations of the results[§]. Secondly, since FBP is based on plain backprojection of counts, local noise effects are spread over a large area, generating typical "star" artifacts (see figure 3.7). These artifacts are reduced with a lower cut–off frequency, but since noise and signal spectra overlap, this results in a contemporary reduction of the signal. A trade–off between artifacts and image sharpness is then required. Finally, the simpleness and speed of FBP implies also a *rigidity* of the algorithm: the backprojection of what is considered to be *entirely* and *solely* signal data is the only performed task. There is neither the possibility to recover wrong or incoherent data contribution to the final image (e.g. not completely exact precorrections introduce unavoidable errors in the results) nor the potentiality of refining the final image by introducing additional information or methods (e.g. to use some

[§]It is useful to underline that an image with negative values can be, from a mathematical point of view, acceptable, since it can still generate non-negative projections — the unique data obtained from the scanner. The negativity of the values in the image becomes obviously unacceptable since these values are linked with the activity concentration in the different voxels.

pre–known characteristics expected to occur in the reconstructed image).

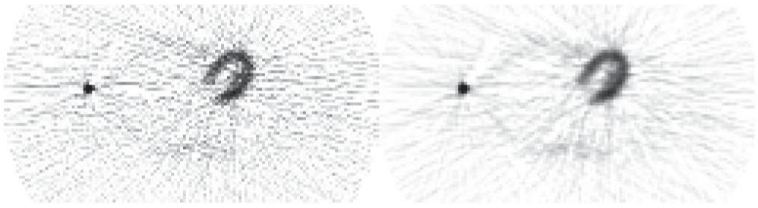


Figure 3.7: Example of FBP artifacts using ramp filter (left) and Hann filter (right) [1]: the pixels with negative values are not shown

3.5 Iterative methods

While analytic algorithms are generally based on the *geometric* aspects of the acquisition phase, another approach to image reconstruction relies on the *statistical* characteristics of the image formation, assuming that the real image — represented by a certain set of *parameters* to be estimated — is the one maximizing a certain objective function: the solution is reached via an *iterative* approach, during which the present approximation of the solution is compared with the acquired data and proper corrections are consequently introduced. The stop condition can be represented by the difference between two consecutive approximations being under a predefined threshold; otherwise, another method is to compare the acquired data with the projections of the present approximation and to stop the algorithm when this difference is below a fixed value. Iterative methods allow using a modelization of the emission and detection processes more accurate than analytic approaches, since physical corrections are generally used *during* the reconstruction and not *before* — and this prevents also the generation of negative values in the final image — and statistical models of emission and detection reflect more correctly the real acquisition mechanism.

The central elements of iterative methods are the projector and backprojector (generally, as stated in section 3.3, they are not one the inverse operator of the other) and the so–called *system matrix*, a matrix which translates the information contained in the image space to the projection space as described in equation 3.1. It is important to note, however, that the system matrix hardly coincides with the pure geometric projection matrix, of which instead is usually a refinement with the addition of some physical corrections. An important peculiarity of a similar implementation is the possibility to *factorize* this matrix (e.g. see [81]), allowing the separation of the different contributions to better experimentally determine them or to improve the computational implementation.

Some of the main characteristics required to an iterative algorithm are:

convergence i.e. to reach a result corresponding to the posed conditions and improving with the increase of iterations — in the case of inconsistent data, the algorithm should enter a *limit cycle* and the final image should remain confined inside it

independence from start condition i.e. to converge both if the starting image is chosen uniform and if a different initial condition is applied — for example, if the result of a (fast but inaccurate) FBP reconstruction is used to start the (long but accurate) iterative reconstruction (IR) it is possible that the inaccuracies of FBP are amplified instead of being corrected

speed i.e. to involve few calculations per iteration and to require a low number of iterations to converge

simplicity i.e. to be easily implementable — also on different computing devices — and user–friendly — simply understandable and requiring few tuning interventions

Obviously no algorithm possesses all these requisites; in addition, their importance varies dependingly on the application under study. For example, higher spatial frequencies require more iterations to be correctly reconstructed: in these cases, an algorithm with a faster convergence can be very useful. Consequently, also taking into account the object under study is important in the choice of an iterative algorithm.

Presently, the most used iterative algorithms are the *maximum likelihood with expectation maximization* (MLEM) [64, 87] and the *ordered subsets expectation maximization* (OSEM) [45]. They represent a good compromise between the quality of the results and the simplicity of implementation and, for these reasons, are among the best candidates for medical applications.

3.5.1 Maximum Likelihood with Expectation Maximization

The general problem of data acquisition can be statistically described by introducing the sample space of observations Y, in which the vector of performed observations $\mathbf{y} \in \mathbb{R}^m$ is contained. Calling χ the space of all possible data, let $\mathbf{x} \in \mathbb{R}^n$ be an element of χ (with $m < n$) called the *complete data*. The vector $\mathbf{x}$ is not observed directly, but only by means of $\mathbf{y} = \mathbf{y}(\mathbf{x})$, where $\mathbf{y}(\mathbf{x})$ is a many–to–one mapping. An observation $\mathbf{y}$ consequently determines a subset $\chi(\mathbf{y})$ of χ.

The *probability density function* (pdf) of the complete data is $f_x(\mathbf{x}|\theta) = f(\mathbf{x}|\theta)$, where $\theta \in \mathbb{R}^r$ is a set of parameters to be estimated assuming that the maximum likelihood estimate of θ lies in the region Θ ($\theta \in \Theta \subset \mathbb{R}^r$). The pdf f is assumed to be a continuous function of θ and appropriately differentiable.

Since $\mathbf{y}(\mathbf{x})$ is a many–to–one map, the pdf of the incomplete data $\mathbf{y}$ is

$$g(\mathbf{y}|\theta) = \int_{\chi(\mathbf{y})} f(\mathbf{x}|\theta) d\mathbf{x} = l_y(\theta)$$

where the *likelihood function* l_y has been introduced. For convenience, another function is actually used in the derivation of the algorithm, the *log–likelihood function* $L_y(\theta) = \log l_y(\theta)$.

The MLEM algorithm is supposed to find θ maximizing $\log f(\mathbf{x}|\theta)$, but unfortunately $\mathbf{x}$ is not available. Therefore, what is actually performed is maximizing the *expectation* of $\log f(\mathbf{x}|\theta)$ given the measured data $\mathbf{y}$ and the *present* estimate of θ. The algorithm is consequently composed of two steps:

- the E (expectation) step, consisting in calculating

$$Q\left(\theta \middle| \theta^{[k]}\right) = E\left[\log f(\mathbf{x}|\theta) \middle| \mathbf{y}, \theta^{[k]}\right]$$

  where $\theta^{[k]}$ is the estimation at step k of θ, which is then assumed to be known at *every* step of the iterative algorithm and represents the conditioner of the complete data likelihood;

- the M (maximization) step, consisting in calculating the *new* estimation $\theta^{[k+1]}$

$$\theta^{[k+1]} = \arg\max_{\theta} Q\left(\theta \middle| \theta^{[k]}\right)$$

  where it is important to note that the maximization is with respect to θ, i.e. the conditioner of the complete data likelihood.

Starting from the initial guess $\theta^{[0]}$, the MLEM can be diagrammed as follows:

$$\theta^{[0]} \xrightarrow{E} Q^{[1]} \xrightarrow{M} \theta^{[1]} \xrightarrow{E} Q^{[2]} \xrightarrow{M} \theta^{[2]} \xrightarrow{E} \ldots \xrightarrow{M} \theta^{[i]} = \tilde{\theta}$$

where $\tilde{\theta}$ satisfies the chosen stop condition.

When the complete data has a pdf belonging to the *exponential family*[¶], i.e. of the form

$$f(\mathbf{x}|\theta) = \frac{b(\mathbf{x})}{a(\theta)} e^{[\mathbf{c}(\theta)]^T \mathbf{t}(\mathbf{x})}$$

the EM algorithm may be simplified. The function $\mathbf{t}(\mathbf{x})$ is called the *sufficient statistic* of the family, since it provides all of the information necessary to estimate the parameters of the distribution from the data. Consequently, for the exponential family the expectation step can be written as

$$Q\left(\theta \middle| \theta^{[k]}\right) = E\left[\log b(\mathbf{x}) \middle| \mathbf{y}, \theta^{[k]}\right] + [\mathbf{c}(\theta)]^T E\left[\mathbf{t}(\mathbf{x}) \middle| \mathbf{y}, \theta^{[k]}\right] - \log a(\theta)$$

[¶]Members of the exponential family are for example Gaussian, binomial, uniform, and Poisson distributions.

Since this quantity should be then maximized with respect to θ, the first term in the expression can be omitted. Therefore, $Q\left(\theta \middle| \theta^{[k]}\right)$ contains only one expectation estimation: the E and M steps can be redefined as

$$\begin{array}{lrcl} \textbf{E} \text{ step} & \mathbf{t}^{[k+1]} & = & E\left[\mathbf{t}(\mathbf{x}) \middle| \mathbf{y}, \theta^{[k]}\right] \\ \textbf{M} \text{ step} & \theta^{[k+1]} & = & \arg\max_{\theta} \left[\left[\mathbf{c}(\theta)\right]^T \mathbf{t}^{[k+1]} - \log a(\theta)\right] \end{array} \tag{3.8}$$

In emission tomography, the body is divided into B boxes and each of them generates $n(b)$ events, $b = 1, 2, \ldots, B$. The set of D detectors[‖] gives the *vector of observations*

$$\mathbf{y} = \{y(d)\} = [y(1), y(2), \ldots, y(D)]$$

The generation of events from box b can be described as a Poisson process with mean $\lambda(b)$

$$f(n|\lambda(b)) = P\left(n(b) = n|\lambda(b)\right) = e^{-\lambda(b)} \frac{[\lambda(b)]^n}{n!}$$

The vector of unknown parameters $\lambda = [\lambda(1), \lambda(2), \ldots, \lambda(B)]$ is the goal of image reconstruction and corresponds to the vector θ of the above theoretical introduction. An event from box b has a probability $p(b, d)$ — here assumed known — of being observed in detector d and, as an initial simplification, it may be assumed that all events are detected:

$$\sum_{d=1}^{D} p(b, d) = 1 \qquad \forall b = 1 \ldots B \tag{3.9}$$

Since $y(d) = \sum_{b=1}^{B} n(b)p(b, d)$, the detector variable $y(d)$ is Poisson distributed (see proof in section M.1)

$$f(y|\lambda(d)) = P\left(y(d) = y|\lambda(d)\right) = e^{-\lambda(d)} \frac{[\lambda(d)]^y}{y!}$$

with

$$\lambda(d) = E\left[y(d)\right] = \sum_{b=1}^{B} \lambda(b)p(b, d)$$

Calling $x(b, d) = n(b)p(b, d)$ the number of events from box b detected in detector d, the vector

$$\mathbf{x} = \{x(b, d),\ b = 1 \ldots B,\ d = 1 \ldots D\}$$

represents the *complete data*, since there is a many–to–one mapping from $x(b, d)$ to $y(d) = \sum_{b=1}^{B} x(b, d)$. Since each variable $x(b, d)$ is Poissonian with mean $\lambda(b, d) = \lambda(b)p(b, d)$, assuming that each box generates events independently of

[‖] In PET, the detectors are actually replaced by the LORs, but for the sake of generality they will be referred as detectors, anyway.

every other box and that the detectors operate independently, too, it is possible to express the log–likelihood as

$$L(\mathbf{x},\lambda) = \log l(\mathbf{x},\lambda) = \log \prod_{\substack{b=1\dots B\\ d=1\dots D}} e^{-\lambda(b,d)} \frac{[\lambda(b,d)]^{x(b,d)}}{x(b,d)!} =$$
$$= \sum_{\substack{b=1\dots B\\ d=1\dots D}} \Big[-\lambda(b)p(b,d) + x(b,d)\log\lambda(b) + x(b,d)\log p(b,d) - \log x(b,d)!\Big] \tag{3.10}$$

Remembering that the Poisson distribution belongs to the exponential family and noting that the sufficient statistic for this distribution is represented by the data themselves ($\mathbf{t}(\mathbf{x}) = \mathbf{x}$), it is possible to use the simplified EM algorithm (expression 3.8). The E step is then represented by

$$x^{[k+1]}(b,d) = E\left[x(b,d)\,\middle|\,\mathbf{y},\lambda^{[k]}\right] = E\left[x(b,d)\,\middle|\,y(d),\lambda^{[k]}\right]$$

where in the last equality only the detector d is considered in the maximization since each detector is independent of the others. Since $x(b,d)$ is Poissonian with mean $\lambda^{[k]}(b,d)$ and $y(d) = \sum_{b=1}^{B} x(b,d)$ is Poissonian with mean $\lambda^{[k]}(d) = \sum_{b=1}^{B}\lambda^{[k]}(b,d)$, this expectation value can be computed as (see section M.1)

$$x^{[k+1]}(b,d) = \frac{y(d)\lambda^{[k]}(b,d)}{\sum_{b'=1}^{B}\lambda^{[k]}(b',d)} = \frac{y(d)\lambda^{[k]}(b)p(b,d)}{\sum_{b'=1}^{B}\lambda^{[k]}(b')p(b',d)} \tag{3.11}$$

The M step, instead, is performed by maximizing the log–likelihood (equation 3.10) with respect to $\lambda(b)$:

$$0 = \frac{\partial L(\lambda)}{\partial\lambda(b)} = \frac{\partial}{\partial\lambda(b)} \sum_{\substack{b=1\dots B\\ d=1\dots D}} \Big[-\lambda(b)p(b,d) + x(b,d)\log\lambda(b) +$$
$$+ x(b,d)\log p(b,d) - \log x(b,d)!\Big] \tag{3.12}$$

To maximize this relation, the new estimation $\lambda^{[k+1]}(b)$ should then satisfy

$$\sum_{d=1}^{D}\left[-p(b,d) + \frac{x^{[k+1]}(b,d)}{\lambda^{[k+1]}(b)}\right] = 0$$

Introducing condition 3.9 and substituting the expression derived in E step (equation 3.11), it is possible to express the updating rule

$$\lambda^{[k+1]}(b) = \lambda^{[k]}(b)\sum_{d=1}^{D}\frac{y(d)p(b,d)}{\sum_{b'=1}^{B}\lambda^{[k]}(b')p(b',d)}$$

Finally, if condition 3.9 is relaxed, the EM updating rule becomes

$$\lambda_b^{[k+1]} = \frac{\lambda_b^{[k]}}{\sum_{d=1}^{D} p_{bd}} \sum_{d=1}^{D} \frac{y_d p_{bd}}{\sum_{b'=1}^{B} \lambda_{b'}^{[k]} p_{b'd}} \tag{3.13}$$

where, as often chosen for typographical simplicity, the arguments have been transformed into subscripts. This will be the convention used in this thesis.

The key point of the EM algorithm resides in the two summations over $b = 1 \dots B$ and over $d = 1 \dots D$:

- $\sum_{b'=1}^{B} [(\cdot) p_{b'd}]$ is usually called the *projector* $\mathrm{P}(\cdot)$, since it translates the image λ into the projections $\mathbf{z} = \mathrm{P}(\lambda)$
- $\sum_{d=1}^{D} [(\cdot) p_{bd}]$ is usually called the *backprojector* $\mathrm{B}(\cdot)$, since it translates the projections $\mathbf{z}'$ into the image $\lambda' = \mathrm{B}(\mathbf{z}')$

These summations contain the *weight factors* p_{bd}, which are fundamental since they represent the system model and allow introducing known information about geometry, distance–dependent resolution and — in some cases — also attenuation and scatter**. From a computational point of view, the nested summations are the toughest part of the implementation, due to high request of resources. Therefore, it is wise to reduce the operations inside the summations at the highest degree, to lighten the calculus load: e.g., it is possible to calculate all the weights before starting the iterations or to extract as many elements as possible from the summations. In an efficiently designed algorithm the computational efficiency is [63]

$$n_{operations/iteration} \propto (n_{pixels})^{2 \sim 3}$$

where n_{pixels} is the dimension of the reconstructed image: hence, halving the pixel size at least quadruples the elaboration time.

Often the EM algorithms are explained saying

> *the image is corrected by multiplying it for the backprojection of the ratio of the real counts and the expected counts from the present approximation†† and — if condition 3.9 is not assumed — by dividing for the backprojection of the unitary matrix.*

Sometimes in this thesis similar terminology will be used.

It is apparent that the main drawback of iterative algorithms — and the reason that their use in image reconstruction started only less than twenty years

**Other way of implementing the physical effects are described in the next section, since they are common both to MLEM and OSEM.

††The expected counts from the present approximations results from the projection of the image approximation, eventually corrected for the superimposed physical effects — see later.

ago, when the power of computing devices became high enough — is the computational time required to reach an adequate level of convergence. Every scheme to reduce the convergence time is therefore very useful.

3.5.2 Ordered Subsets Expectation Maximization

A generally used strategy to reduce the computational load is to divide the task to be performed into more simple groups of operations: this scheme is usually referred to as *divide et impera* (divide and conquer), from the Romans' behaviour in the control of their empire. Two typical examples are a class of sorting algorithms and the FFT.

Also in iterative algorithms this method has been succesfully implemented in the so–called *ordered subsets (maximum likelihood) expectation maximization* (OSEM). It is a MLEM algorithm in which the entire set $\Omega = P(i)$, $i = 1 \dots n_P$ of projections originating from detectors $d = 1 \dots D$ is divided into equipotent (and usually disjoint) *subsets* S_m, $m = 1 \dots M$

$$\Omega = \bigcup_{m=1}^{M} S_m$$

where M is a divisor of the number of angular projections n_P. Many different criteria for filling these subsets are present (e.g. see [45]), but the most used is to choose equispaced projections

$$S_m = \left\{ P\left(m + jM\right), \quad j = 0 \dots \frac{n_P}{M} - 1 \right\}$$

Each MLEM iteration k on the entire set Ω is then divided into M sub–iterations (k, m) in which the approximation $\lambda^{[k,m]}$ — computed using *only the data in subset* S_m — is updated using the data in subset S_{m+1}. The resulting rule is consequently

$$\lambda_b^{[k'+1]} = \frac{\lambda_b^{[k']}}{\sum\limits_{d \in S_m} p_{bd}} \sum_{d \in S_m} \frac{y_d p_{bd}}{\sum\limits_{b'=1}^{B} \lambda_{b'}^{[k']} p_{b'd}} \tag{3.14}$$

where the sub–iteration index $k' = kM + m - 1$ has been introduced for analogy with the MLEM rule. It is important to note that a *complete* iteration $k \to k+1$ in the OSEM algorithm corresponds to M effective iterations ($k' \to k' + M$), while the time required for a sub–iteration is approximately $1/M$ the time for a MLEM iteration due to the reduced terms in the summations. Consequently, the OSEM algorithm presents an about M–fold faster convergence with respect to MLEM (for OSEM convergence proof, see for example [45]). It is necessary to underline, however, that a trade–off between speed and dimension of the subsets is necessary: if each subset has too few projections, the method generally looses

robustness since the image is compared with insufficient data. As a rule of thumb, a balanced choice is to have about $10 \sim 20$ projection angles per subset.

What has been presented until now is valid if no physical processes superimposed to the signal are considered. The corrections[‡‡] presented in sections 2.3 and 2.5 can be incorporated — as anticipated in the previous section — inside the weights p_{bd} using a factorization scheme as presented for example in [67, 78, 82, 85]. On the one hand, this method allows simplifying the algorithm, but it requires quite large memory to store the complete matrix. Two other methods are to precorrect the recorded data or to introduce the corrections *inside* the algorithm itself (the so–called *in–loop corrections*). Precorrecting allows using the rule described in equation 3.14 as it is, but it has got the disadvantage of possible negative values in the final image (as stated for FBP in section 3.4.1). Using in–loop corrections, on the contrary, requires more calculations — and, consequently, higher reconstruction time — but yields more accurate results and avoids negative values. Generally, the trade–off is represented by precorrecting all the *multiplicative* effects (e.g. detector normalization, dead time etc.) and to correct in the loop all the *additive* effects (e.g. randoms, scatter etc.), since only these latter can originate negative values. To introduce in–loop corrections, it is important to note that physical processes influence the recorded data $\mathbf{y}$ and, consequently, the projection of the image present approximation should be corrected adequately before calculating the ratio to be backprojected. For example, a possible implementation is

$$\lambda_b^{[k'+1]} = \frac{\lambda_b^{[k']}}{\sum\limits_{d \in S_m} p_{bd}} \sum_{d \in S_m} \frac{y_d p_{bd}}{\sum\limits_{b'=1}^{B} \lambda_{b'}^{[k']} p_{b'd} + R_d + S_d} \tag{3.15}$$

where R_d and S_d are the contributions of the random and scatter coincidences (respectively) to the data recorded by detector d. If attenuation effects are precorrected, y_d already contains attenuation information. Anyway, in recent algorithms — thanks to higher computational resources — also attenuation is included in the loop. In this case, the updating rule becomes

$$\lambda_b^{[k'+1]} = \frac{\lambda_b^{[k']}}{\sum\limits_{d \in S_m} A_d p_{bd}} \sum_{d \in S_m} \frac{A_d y_d p_{bd}}{A_d \sum\limits_{b'=1}^{B} \lambda_{b'}^{[k']} p_{b'd} + R_d + S_d} \tag{3.16}$$

where A_d is a coefficient — called *attenuation correction factor* (ACF) — describing the attenuation along LOR d. For the sake of clarity, it may be convenient to introduce the definitions of projector and backprojector presented above: in

[‡‡] These corrections are nearly always computed in the sinogram space, i.e. the superimposed effects are calculated for each *detector* $d = 1 \dots D$.

this case, the updating rule becomes

$$\lambda_b^{[k'+1]} = \frac{\lambda_b^{[k']}}{\mathrm{B}\,(A_d)}\mathrm{B}\left\{\frac{A_d y_d}{A_d\,\mathrm{P}\left(\lambda_{b'}^{[k']}\right) + R_d + S_d}\right\} \tag{3.17}$$

It is important to note that all the physically–corrected equations, here explained from a physical point of view, can be mathematically derived following the same strategy presented in section 3.5.1 and remembering that R_d and S_d are the expectation values of random processes described by Poisson distribution and the sum of Poisson distributions is a Poisson distribution, too, as stated in section M.1.

If attenuation is taken into account (with the same notations used in deriving the MLEM updating rule), the variables $x(b,d)$ follow a Poisson distribution with mean $\lambda(b)p(b,d)A_d$, since the expected mean $\lambda(b)p(b,d)$ should be lessened to take into account the attenuation A_d. Random and scatter coincidences change the expression for the recorded events in LOR d:

$$y(d) = \sum_{b=1}^{B} x(b,d) + \hat{R}_d + \hat{S}_d = \sum_{b=1}^{B} A_d\, n(b)p(b,d) + \hat{R}_d + \hat{S}_d$$

where $\hat{R}_d$ and $\hat{S}_d$ are the random and scattered coincidences, respectively, actually recorded in LOR d, both following a Poisson distribution of mean R_d and S_d, respectively. Consequently, equation 3.11 should be changed into

$$x^{[k+1]}(b,d) = \frac{A_d y(d)\lambda^{[k]}(b)p(b,d)}{\sum\limits_{b'=1}^{B} A_d\,\lambda^{[k]}(b')p(b',d) + R_d + S_d}$$

and, by using the same strategy as in equation 3.12 and the division of the LORs in the different subsets, the updating rule presented in equation 3.16 is obtained.

3.6 Time-of-flight reconstruction

As described in chapter 2, in TOF modality also the difference in the arrival time of each couple of photons on the detectors is recorded for each event. This results in a more detailed information about the body under study, but it leads to an increase in the complexity of the reconstruction process. The most important challenge is to make reconstruction time clinically viable, since TOF brings a non-negligible additional computation burden. The choice of data organization (sinogram or listmode) and the choice of algorithm (analytical or iterative) are only parts of this equation, together with other elements, such as computer architecture or software optimization. A second issue is the correctness of the

reconstruction in terms of validating the individual components that must now properly include the TOF information; for example, the scatter correction.

While in conventional PET the forward and back projection is done along the LORs and all points along the LOR have equal probability, with TOF forward and back projection is still done along the lines but all points along the line have a *different* probability. The maximum probability is given to the point determined by the measured TOF difference. The width of the probability distribution is determined by the expected accuracy of the time difference measurement (mostly Gaussian distributions are assumed). All the different reconstruction strategies share this common element.

As stated above, data organization is one of the most important topics in TOF modality. Even if the most natural scheme is the list-mode representation (since each event is individually stored, independent of the others), it leads to very large datasets and is difficult to be used with analytical algorithms. For these reasons, in the following only sinogram-based reconstructions will be presented, with just a brief reference to list-mode only for iterative algorithms.

The sinogram-based approach consists in dividing the maximum allowed time difference between the arrival of the photons into n_{TOF} intervals (usually chosen with equal amplitude) and in generating, from the entire recorded dataset, n_{TOF} sinograms, each corresponding to a specific time interval (see figure 3.8)*. Each sinogram is then used in the reconstruction by introducing a proper weight to take the temporal information into account.

3.6.1 Analytical algorithms

Following the same approach of equation 3.5, the two-dimensional image may be written as

$$f(x,y) = \int_0^{\pi} d\theta \sum_{e=1}^{n_{TOF}} FT^{-1}\left[M(\omega,\theta,e)L(\omega)\right](s)h(t-t_e)\Big|_{\substack{t=x\cos\theta+y\sin\theta \\ s=-x\sin\theta+y\cos\theta}}$$

where s is the projection coordinate and t is the coordinate (relative to time information) perpendicular to s, ω is the frequency space coordinate associated to s, FT^{-1} indicates the inverse one-dimensional Fourier transform operator, e is the index of TOF bin and t_e is the position of the spatial center of the bin e along the direction t. $M(\omega,\theta,e)$ is the one-dimensional Fourier transform of the projection $m(s,\theta,e)$ (i.e. for angle θ and TOF bin e) and $h(t-t_e) = e^{-\frac{(t-t_e)^2}{2\sigma^2}}$ is the TOF resolution function, where

$$\sigma = \frac{FWHM_x}{2\sqrt{2\ln 2}} = \frac{c\Delta t}{4\sqrt{2\ln 2}}$$

*In practice, this approach adds a 4^{th} dimension to the 3D conventional sinogram representation.

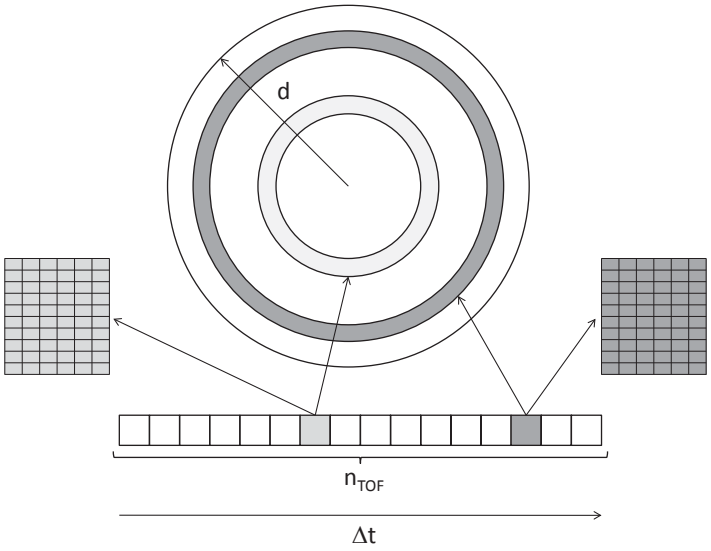


Figure 3.8: Time-binned sinograms for TOF reconstruction: the entire dataset is divided into sinograms obtained from different temporal bins (corresponding to different positions of the maximum of the probability distribution)

is the localization uncertainty due to the finite time resolution Δt of the system. Finally, $L(\omega)$ is the filter in the frequency space (see also $H(\omega)G(\omega)$ in expression 3.7): different proposals have been presented (e.g. see [62, 88, 98]), but a common choice (proposed in [25]) is to obtain it as the convolution (in the frequency space) between the ramp filter ($H(\omega)$, expression 3.6) and the square of the Fourier transform of the Gaussian TOF response kernel (directed along the projection direction s):

$$L(\omega) = |\omega| \otimes F\left[e^{-\frac{s^2}{2\sigma^2}}\right]^2 \propto |\omega| \otimes e^{-4\pi^2\sigma^2\omega^2}$$

It is also possible to generalize what exposed above to the three-dimensional case, by introducing also proper coordinates along the axial direction and writing again the recorded projections as a line integral weighted by the TOF kernel h [21]†:

$$p_{3D}(s,\phi,z,\delta;t) = \sqrt{1+\delta^2}\int_{-\infty}^{\infty} f(s\cos\phi - l\sin\phi, s\sin\phi + l\cos\phi, z + l\delta) \\ h(t - l\sqrt{1+\delta^2})\,dl \quad (3.18)$$

where f represents the 3D object, s and ϕ are the radial and angular coordinates, respectively, z is the axial midpoint of each LOR, δ is the tangent of the oblique

†A change in notation with respect to the above one has been preferred to be consistent with literature references.

angle θ and t is the TOF variable. Figure 3.9 helps with the definition of these variables.

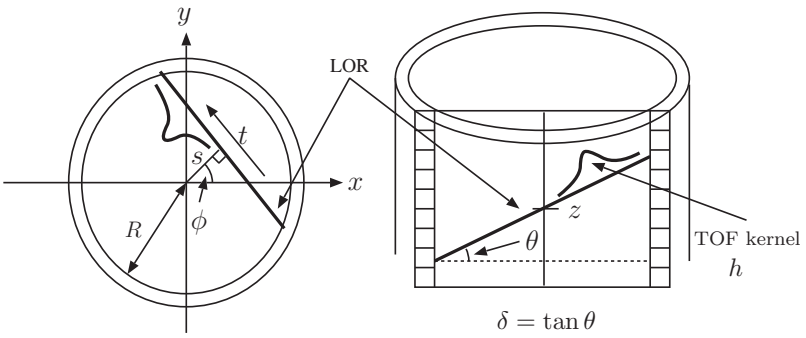


Figure 3.9: Definition of the variables for the presented 3D TOF analytical reconstruction: transaxial (left) and axial (right) views of a cylindrical scanner

3.6.2 Iterative algorithms

Using the OSEM approach as a reference for iterative algorithms, the extension to the TOF case needs the introduction of the time-of-flight sinograms (as described above) y_{dt} (with $y_d = \sum_t y_{dt}$) and of the *time spread function* (TSF) c_{bdt} along LOR d, representing the probability of detecting, in LOR d, an event with TOF=t originated in voxel b. As before, the TSF is assumed to be Gaussian with FWHM equal to the time resolution of the system. The most straightforward implementation is represented by a direct generalization of equation 3.14:

$$\lambda_b^{[k'+1]} = \frac{\lambda_b^{[k']}}{\sum\limits_{d \in S_m, t} c_{bdt}} \sum_{d \in S_m} \sum_t \frac{y_{dt} c_{bdt}}{\sum\limits_{b'=1}^{B} \lambda_{b'}^{[k']} c_{b'dt}} \tag{3.19}$$

The inclusion of the corrections for the physical effects must take into account the different statistical properties of TOF reconstruction (e.g. due to the space-dependent probability along the LOR). For this reason, different methods (in particular for attenuation and scatter corrections) have been proposed (e.g. [20, 23–25, 105]).

Finally, as far as list-mode data are concerned, they may be reconstructed with iterative algorithms by following similar strategies (see e.g. [74, 101]), all based on the introduction of an additional (usually Gaussian) TOF function.

Part II

Performed work

Chapter 4

PSF measurements

Since the PSF describes how a point source is rendered by the scanner, the method used to characterize the response of the PET system consisted in acquiring (in 3D mode) a β^+ source located in different positions within the entire FOV, in order to map the spatial variance of the response (section 4.1). For each position, the acquired data were then reconstructed (with the same algorithm — OSEM — in which the PSF would be implemented) and the resulting images were used to build the required PSF model by estimating the specific parameters described in section 4.2.

4.1 PSF measurements

The PSF of the PET system was experimentally measured using an encapsulated, non-collimated radioactive point source, composed of an active ^{22}Na core (a small cylinder with diameter of 1.65 mm and height of 1 mm) surrounded by a disk of lucite (for more details, see [53]) — as proposed in [2] — with an activity of 100 μCi at the time of the measurements.

Using the same source for all the measurements allows having difference of statistics due only to sensitivity effects and not to variable emission count rate, since the source activity can be considered constant in time — the half–life of ^{22}Na is $\tau_{^{22}Na} \approx 951$ days — when the measurements are performed in few hours. Moreover, by using the same point source for all the measurements in the FOV, a possible source of errors in building the PSF model is removed. Unfortunately, ^{22}Na possesses different physical characteristics with respect to FDG, but the equivalence between FDG in water and ^{22}Na in lucite has been demonstrated in [2].

The β^+ source was acquired at discrete positions along a diameter of the transaxial circular FOV and along the axial direction of the scanner: to simplify

the task, the chosen diameter was the horizontal one and the axis of the cylindrical source was oriented along the vertical direction. Since the count rate varies over the FOV due to transaxial and axial sensitivity effects, for each sampled point in the 3D FOV the acquisition time was chosen accordingly to the detected event count rate in order to obtain high enough statistics (approximately 1.5×10^8 counts) to guarantee robust estimation of the parameters.

4.2 The PSF model

The PSF kernel should be constructed such that the measurement model used in the reconstruction approximates the system response as close as possible. For a particular point source measurement, this amounts to creating the PSF kernel such that its forward projection matches the measured data. As noted in [97] for the sinogram PSF approach, the corresponding Maximum Likelihood (ML) estimate can be obtained by using MLEM. If the point source is small (compared to the Point Spread Function), the ML estimate for the image-based PSF of the corresponding voxel is therefore identical to the image reconstructed using MLEM (with the same projectors as will be used in the reconstruction of the clinical data).

However, this approach has the following problems:

- it is difficult to interpolate the PSF for locations where the point source was not measured;
- different voxel sizes should be taken into account to change the description of the PSF accordingly;
- for small voxel sizes, the ML estimate of the kernel is noisy.

Therefore, a different approach was followed.

The point source raw data were precorrected for detector efficiencies, dead-time and geometric effects and then reconstructed using a 3D OSEM algorithm on a FOV of 128 mm centred on the source position. The reconstruction parameters were 10 iterations, image matrix 256 pixels $\times$ 256 pixels, pixel size 0.5 mm. The number of subsets was set to 28 (for DSTE measurements, see chapter 7) or 18 (for D690 measurements, see chapter 8). The OSEM implementation used the distance-driven projectors proposed in [30].

As stated above, when reconstructing in this way, the resulting image appeared very noisy and difficult to analyse. This effect was overcome by filtering the reconstructed images with a *transaxial* Gaussian post–filter with $\mathrm{FWHM}_{PF} =$ 4 mm, while along the axial direction no filter was applied. An example of images without and with transaxial postfilter is provided in figure 4.1. In principle, the

addition of a post-filter artificially enlarges the PSF and might lead to overestimation of the required parameters. The method, however, took into account also the postfilter in the PSF estimation as detailed below.

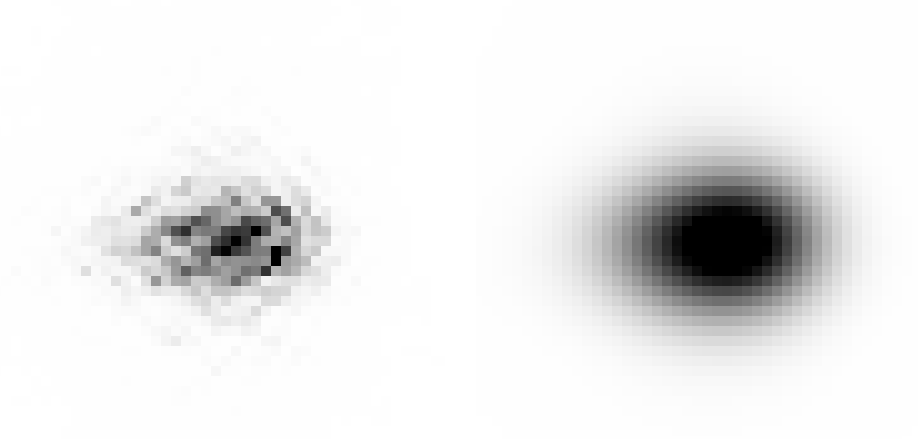


Figure 4.1: Example of OSEM reconstruction (FOV=128 mm on 256 pixels × 256 pixels) of a point source located 20 cm from the scanner center in absence (left) and in presence (right) of Gaussian post–filter

The PSF was modelled as a three-dimensional (3D) Gaussian function with space-dependent widths. Moreover, the PSF was factorized into a symmetric one-dimensional Gaussian function along the scanner axis (called z) and a two-dimensional (2D) Gaussian function in the transaxial plane. The transaxial 2D Gaussian function was then modelled by a symmetric function along the tangential axis and by an *asymmetric* function along the radial direction, with the wider half of the asymmetric function towards the centre of the scanner.

In each point of the scanner FOV, the 3D PSF was therefore characterized by four spread parameters (the *internal radial* σ_i, the *external radial* σ_e, the *tangential* σ_t and the *axial* σ_a ones), which – exploiting the cylindrical symmetry of the scanners – were assumed to depend only on the *radial* and *axial* distances from the scanner FOV centre. The PSF (normalized to unitary volume) in the generic point P can be then expressed as

$$PSF_{(P)}(x,y,z) = \frac{2}{(2\pi)^{3/2}\,\sigma_t\sigma_a\,(\sigma_i+\sigma_e)}\left[\theta(-x)e^{-\frac{x^2}{2\sigma_i^2}} + \theta(x)e^{-\frac{x^2}{2\sigma_e^2}}\right] e^{-\frac{y^2}{2\sigma_t^2}}\, e^{-\frac{z^2}{2\sigma_a^2}} \tag{4.1}$$

where $\theta(x)$ is the Heaviside step-function and the coordinates of the Cartesian system (x,y,z) (whose origin is P) correspond, respectively, to the radial, tangential and axial directions, as stated before. In the equation above, all the σ parameters are intended as $\sigma = \sigma\left(\vec{P}\right)$, where $\vec{P}$ indicates the position of the point P with respect to the scanner centre.

Each three-dimensional image (corresponding to a definite position of the ^{22}Na source inside the scanner FOV) was considered as a matrix, whose indexes were referred to as *row number*, *column number* and *slice number*; after rotating the transaxial images to have the radial direction coincident with the positive x axis (along the row direction) and selecting the voxel with the maximum intensity, three two-dimensional planes passing through this voxel were extracted; each of them was characterized by fixing one out of row number, column number and slice number (see figure 4.2), as follows:

1. fixing the slice number, the *radial-tangential* plane was obtained;

2. fixing the row number, the *radial-axial* plane was obtained;

3. fixing the column number, the *tangential-axial* plane was obtained.

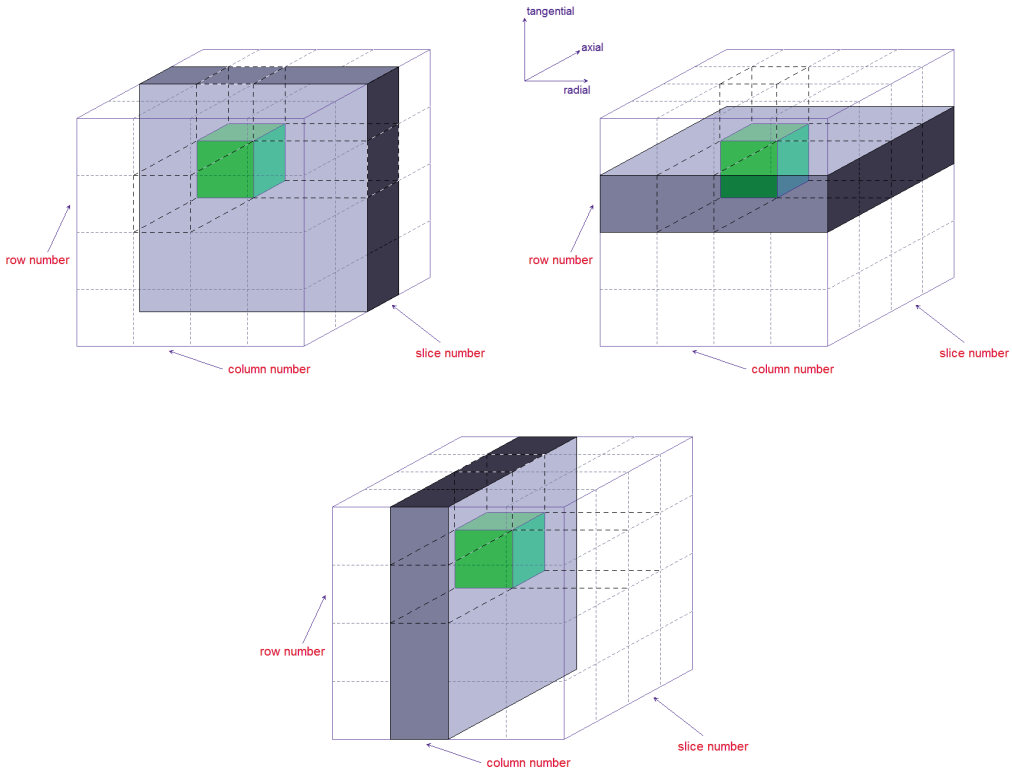


Figure 4.2: Extraction of the three planes passing through the voxel of maximum intensity to measure the PSF: the radial–tangential plane (top left, fixed slice number), radial–axial plane (top right, fixed row number), tangential–axial plane (bottom, fixed column number)

Each plane originated a 2D image, which was then fitted* with the proper function to obtain an estimation of the corresponding spread parameters.

Since the point source was a cylinder of radius r and height h with axis parallel to the tangential direction, after introducing the parameter

$$\gamma = \frac{\text{Transaxial pixel size}}{\text{Axial pixel size}} \tag{4.2}$$

the fit functions† used were

1. for the radial-tangential plane,

$$\begin{aligned} f_{rt}(x,y) = \frac{A_1}{C_1+D_1}\left[\operatorname{erf}\left(\frac{y-F+h/2}{\sqrt{2}\sqrt{E_1^2+\sigma_c^2}}\right) - \operatorname{erf}\left(\frac{y-F-h/2}{\sqrt{2}\sqrt{E_1^2+\sigma_c^2}}\right)\right]\cdot \\ \cdot \int_{-r}^{r}\left\{\frac{C_1}{\sqrt{C_1^2+\sigma_c^2}} e^{-\frac{(x-B-t)^2}{2\left(C_1^2+\sigma_c^2\right)}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}}\frac{C_1\,(x-B-t)}{\sigma_c\sqrt{C_1^2+\sigma_c^2}}\right] + \right. \\ \left. + \frac{D_1}{\sqrt{D_1^2+\sigma_c^2}} e^{-\frac{(x-B-t)^2}{2\left(D_1^2+\sigma_c^2\right)}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}}\frac{D_1\,(x-B-t)}{\sigma_c\sqrt{D_1^2+\sigma_c^2}}\right]\right\} dt \end{aligned} \tag{4.3}$$

2. for the radial-axial plane,

$$\begin{aligned} f_{ra}(x,z) = \frac{A_2}{C_2+D_2}\cdot \\ \cdot \int_{-r}^{r}\left\{\frac{C_2}{\sqrt{C_2^2+\sigma_c^2}} e^{-\frac{(x-B-t)^2}{2\left(C_2^2+\sigma_c^2\right)}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}}\frac{C_2\,(x-B-t)}{\sigma_c\sqrt{C_2^2+\sigma_c^2}}\right] + \right. \\ \left. + \frac{D_2}{\sqrt{D_2^2+\sigma_c^2}} e^{-\frac{(x-B-t)^2}{2\left(D_2^2+\sigma_c^2\right)}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}}\frac{D_2\,(x-B-t)}{\sigma_c\sqrt{D_2^2+\sigma_c^2}}\right]\right\}\cdot \\ \cdot \left[\psi_t\left(\langle z\rangle - H + 1\right) - \psi_t\left(\langle z\rangle - H - 1\right)\right] dt \end{aligned} \tag{4.4}$$

$$\begin{aligned} \psi_t(s) = \left(s+\gamma\sqrt{r^2-t^2}\right)\operatorname{erf}\left(\frac{s+\gamma\sqrt{r^2-t^2}}{\sqrt{2}G_1}\right) - \\ - \left(s-\gamma\sqrt{r^2-t^2}\right)\operatorname{erf}\left(\frac{s-\gamma\sqrt{r^2-t^2}}{\sqrt{2}G_1}\right) + \\ + \sqrt{\frac{2}{\pi}}G_1\left[e^{-\frac{\left(s+\gamma\sqrt{r^2-t^2}\right)^2}{2G_1^2}} - e^{-\frac{\left(s-\gamma\sqrt{r^2-t^2}\right)^2}{2G_1^2}}\right] \end{aligned} \tag{4.5}$$

*All the fits presented in this work have been performed using the software ROOT (http://root.cern.ch) developed at European Organization for Nuclear Research (CERN).

†In the equations, all the variables — but the ones regarding axial direction — plus r and h are intended in "transaxial pixels", while the variables for axial direction are intended in "axial pixels".

3. for the tangential-axial plane,

$$f_{ta}(y,z) = A_3 \left[\operatorname{erf}\left(\frac{1}{\sqrt{2}} \frac{y - F + h}{\sqrt{E_2^2 + \sigma_c^2}} \right) - \operatorname{erf}\left(\frac{1}{\sqrt{2}} \frac{y - F - h}{\sqrt{E_2^2 + \sigma_c^2}} \right) \right] \cdot \\ \cdot \left[\omega\left(\langle z \rangle - H + 1 \right) - \omega\left(\langle z \rangle - H - 1 \right) \right] \quad (4.6)$$

$$\omega(s) = (s + \gamma r) \operatorname{erf}\left(\frac{s + \gamma r}{\sqrt{2} G_2} \right) - (s - \gamma r) \operatorname{erf}\left(\frac{s - \gamma r}{\sqrt{2} G_2} \right) + \\ + \sqrt{\frac{2}{\pi}} G_2 \left[e^{-\frac{(s+\gamma r)^2}{2G_2^2}} - e^{-\frac{(s-\gamma r)^2}{2G_2^2}} \right] \quad (4.7)$$

In the expressions above, the rounding operator

$$\langle z \rangle = \arg\min_n |z - n|$$

was introduced.

The fit functions took into account different elements.

Post–filter

1. The applied post–filter

$$f_{PF} = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2\sigma_c^2} \right) \qquad \sigma_c = \frac{\mathrm{FWHM}_{PF}}{\sqrt{8 \ln 2}}$$

is included in the analytic expression of the fit function. This is important since the post–filter both enlarges the Gaussian spread and changes the shape of a non–point source (for details, see section M.2).

Positioning

2. Furthermore, also the coordinates of the source position were obtained from the fit (respectively, the x coordinate from parameter B, the y coordinate from parameter F and the z coordinate from parameter H) and they were used to estimate the true radial and axial distances from the scanner FOV centre. This should help compensate for the imprecision in the source positioning.

Source dimensions

3. Since the fitted images have a transaxial pixel size of 0.5 mm and the active volume of ^{22}Na is a cylinder with radius of 0.825 mm and height of 1 mm, the condition of "point source" is not rigorously fulfilled. Therefore, also its physical dimensions (i.e. the radius r and the height h) should be taken into account to avoid an overestimation of the spread parameters. The physical dimensions in the fitting functions were fixed to the known values for r and h.

Discretization along the scanner axis and effective radius

4. Since a PET scanner typically produces images with slicewidths of 2–4 mm, the PSF kernel along the axial direction should take into account this coarseness. For this reason, expressions 4.4 and 4.6 were derived from the functions M.19 and M.20 by integrating along the axial direction, by recalling that

$$\int \operatorname{erf}(ax+b)\,dx = \frac{1}{a}\left[(ax+b)\operatorname{erf}(ax+b) + \frac{e^{-(ax+b)^2}}{\sqrt{\pi}}\right] + c$$

Moreover, given the orientation of the cylindrical radioactive source, in f_{ta} the radius is substituted with an "effective" radius r_{eff} (to partially compensate the effect[‡] of the circular shape on the final result) defined as

$$r_{eff} = \frac{A_{cs}}{2\,ps}$$

where A_{cs} is the area of the circular segment comprised in the slice and ps is the transaxial pixel size. This corresponds to considering the comprised area as a rectangle of sides ps and $2r_{eff}$. In figure 4.3 a scheme of this procedure is presented. Finally, as detailed in section 5.2, the discretization along the scanner axis is taken into account by using, in the reconstruction algorithm, the *integral* of the PSF along the axial direction.

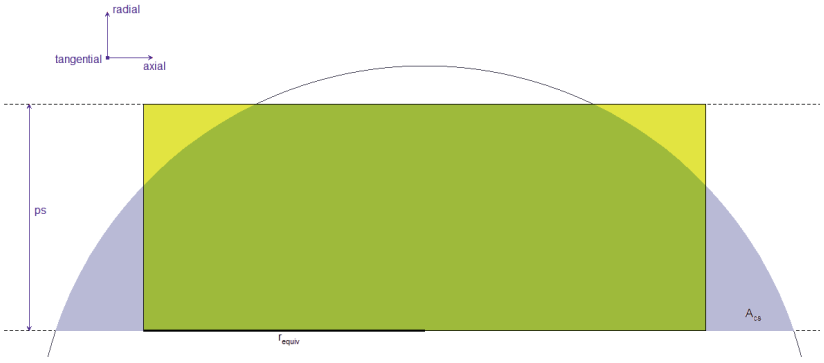


Figure 4.3: Introduction of the equivalent radius in the fitting procedure: when the tangential-axial plane is selected, due to the spatial extent ps of the pixel along the radial direction (comparable to the source dimensions), the circular shape is not negligible; here the real area A_{cs} (grey) is assumed to be equivalent to a *rectangular* area (yellow) of sides ps and $2r_{equiv}$.

Each spread parameter was then estimated twice:

[‡]The "recorded" PSF in the tangential-axial plane is the superposition of elementary contributions in which the PSF blurs different rectangular cross sections of the cylinder, with equal height — i.e. equal vertical side — and different radii — i.e. different horizontal side.

- internal radial spread parameter was represented by parameters C_1 and C_2;
- external radial spread parameter was represented by parameters D_1 and D_2;
- tangential spread parameter was represented by parameters E_1 and E_2;
- axial spread parameter was represented by parameters G_1 and G_2.

The final estimation of each spread parameter was then computed as the mean of the corresponding two values. The parameter A_1, A_2 and A_3 represented normalization factors and were not used in the following steps.

All the obtained values for each spread parameter were then graphed as a function of both radial and axial distances from the scanner FOV centre: it was then possible to fit each set of data with a polynomial quadratic in both distances, symmetric with respect to the centre of axial FOV:

$$f(r,z) = \left(R_0 + R_1 r + R_2 r^2\right)\left(A_0 + A_2 z^2\right) \tag{4.8}$$

where r indicates the radial distance and z the axial distance from the scanner center.

By knowing the coefficients R_0, R_1, R_2, A_0 and A_2 it was then possible to calculate the four spread parameters for each point inside the whole scanner FOV.

4.3 Accuracy of the method

To check whether the fitting procedure is consistent and unbiased, a simulated set of PSF measurements were generated. Each PSF measurement consisted in considering a three-dimensional image, selecting a position inside the scanner (identified by its radial and axial distances from the center) and generating a cylinder of known radius (0.825 mm) and height (1 mm), with axis parallel to the vertical direction in the transaxial images (corresponding to tangential direction in the scanner). The dimensions of the images were $256 \times 256 \times 47$ with simulated voxel size of $0.5 \times 0.5 \times 3.27$ mm^3. The image was centred on the ideal position set before. The center of the cylinder, instead, was positioned by introducing random fluctuations (with respect to the position selected at the beginning) along both the radial (x axis) and axial (z axis) directions — to simulate random mispositioning of the source — while it was centred along the tangential direction. Each voxel of the image was filled with the area of the cylinder internal to the voxel itself.

The resulting image was then convolved with the spatially–variant PSF defined as in equation 4.1 with

$$\begin{aligned}
\sigma_i &= A_i r^2 + B_i r + C_i \\
\sigma_e &= A_e r^2 + B_e r + C_e \\
\sigma_t &= A_t r^2 + B_t r + C_t \\
\sigma_a &= (A_a r^2 + B_a r + C_a)(Da^2 + E)
\end{aligned}$$

where r and a are, respectively, the radial and axial distances from the scanner centre of each voxel in the image, and the coefficients $A \ldots E$ were set by choosing the values indicated in table 4.1. This simulation did not take into account any superimposed noise (since the noise model in the resulting image was unknown): even if this could appear to consitute a limitation and a more favourable condition than the real one, it should be remarked that the additional post-filter — introduced to remove the noisy patterns — generates a very smooth image, which appears very similar to a non-noisy image. Consequently, this limitation is not considered to reduce the reliability of the validation results.

	$r = 0$ cm		$r = 10$ cm		$r = 20$ cm	
	$a = 0$ cm	$a = 8$ cm	$a = 0$ cm	$a = 8$ cm	$a = 0$ cm	$a = 8$ cm
σ_i	1.40		2.00		3.30	
σ_e	1.40		1.50		1.80	
σ_t	1.40		1.50		1.62	
σ_a	1.52	1.63	1.55	1.66	1.65	1.77

Table 4.1: Choice of sigma parameters (in mm) for the simulation of PSF fitting

The transaxial spread parameters σ_i, σ_e and σ_t were assumed to depend only on the *radial* distance from the scanner centre, since the axial dependence was expected to be weak (and later confirmed in real data, as shown in chapters 7 and 8). Finally, a two–dimensional Gaussian post–filter with FWHM = 4 mm was applied to the image, while along the axial direction no filter was applied.

The images were then fitted as in the clinical practice and by fixing both the radius and the height parameters to the known values.

In figure 4.4 a synthetic image of the cylinder and the corresponding PSF image are shown, while in figure 4.5 the fit results for two planes (radial–tangential and tangential–axial) on the above example are presented. As can be seen from the figures, the fitted images appear to resemble well the original images (e.g. compare the statistics relative to the histograms in the top-right part of each figure).

By graphing all the evaluated spread parameters as a function of both radial and axial distances from the scanner FOV centre and fitting the data with a polynomial quadratic in both distances, symmetric with respect to the centre of axial FOV, as in equation 4.8, the spatial dependences of the parameters were evaluated (figure 4.6) and can then be compared to the known ones. For the transaxial parameters a direct comparison could be performed, since these parameters depend only on the radial distance. For the axial spread parameter, since it depends on both radial and axial distances, the comparison was performed only along the axial direction, by fixing the radial distance at 5 cm[§].

[§]This distance is arbitrary, but it was considered a reference point, common also to some

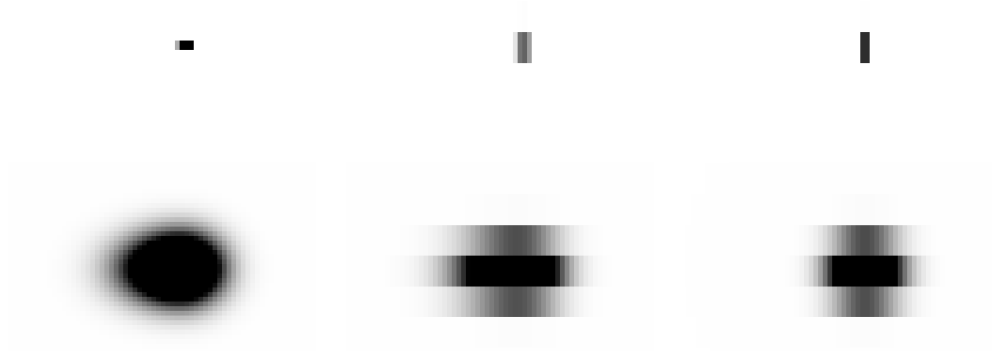


Figure 4.4: Example of simulated PSF acquisition: cylinder without PSF and post–filter contributions (top) and final simulated image (bottom); transaxial (left), coronal (center) and sagittal (right) views.

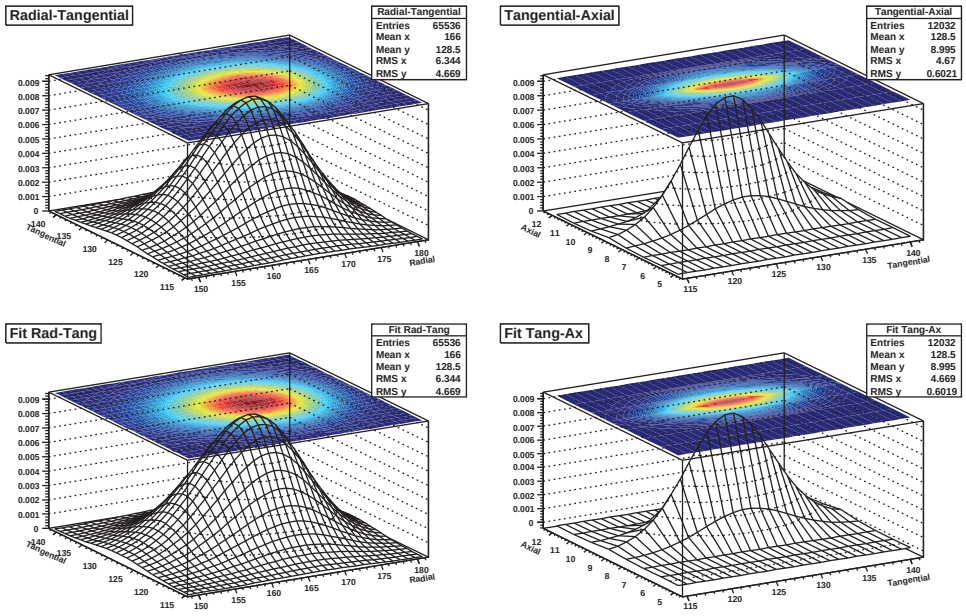


Figure 4.5: Example of fits on simulated PSF acquisition. Top: radial–tangential and tangential–axial images; bottom: corresponding fits

In figure 4.7 this comparison is presented. The parameters appear to be evaluated with global good accuracy. In the axial direction a slight overestimation can be noticed. Anyway, this overestimation is always lower than 5%: if the difficulties in applying this method to a very coarse grid (as in the axial spacing case) are taken into account, the level of accuracy is considered to be satisfying. For the transaxial parameters the accuracy appears to be very good. For all these considerations, the fitting procedure is believed to be accurate enough to support the building of the PSF models described in sections 7.2 and 8.2.

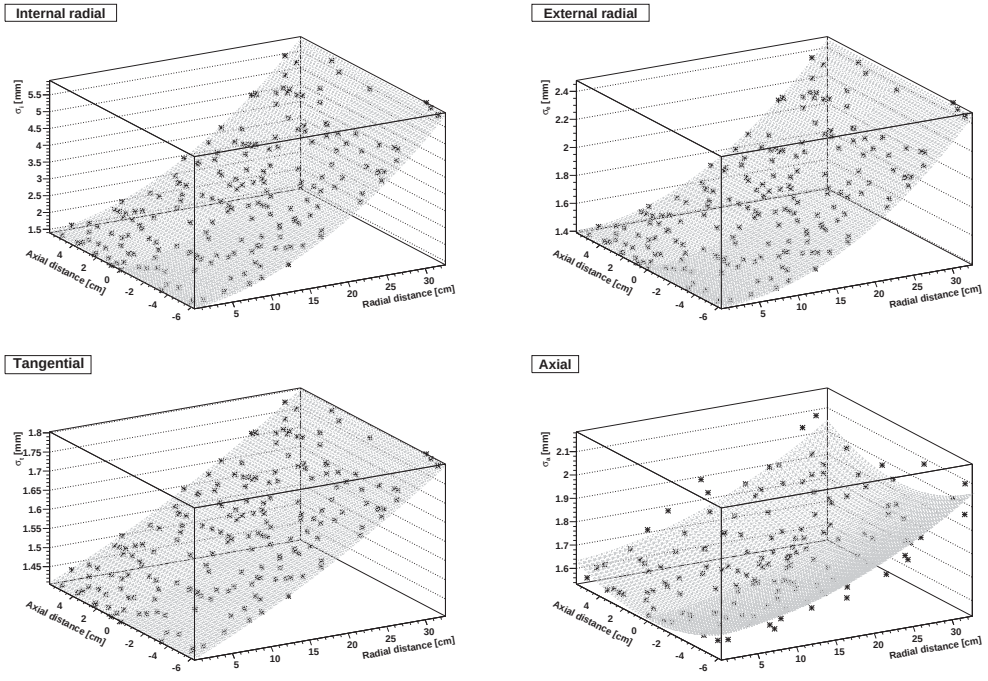


Figure 4.6: Fitting of spread parameters in simulated PSF: σ_i (top left), σ_e (top right), σ_t (bottom left) and σ_a (bottom right)

validation results presented in chapters 7 and 8 (e.g. the NEMA IEC Body Phantom).

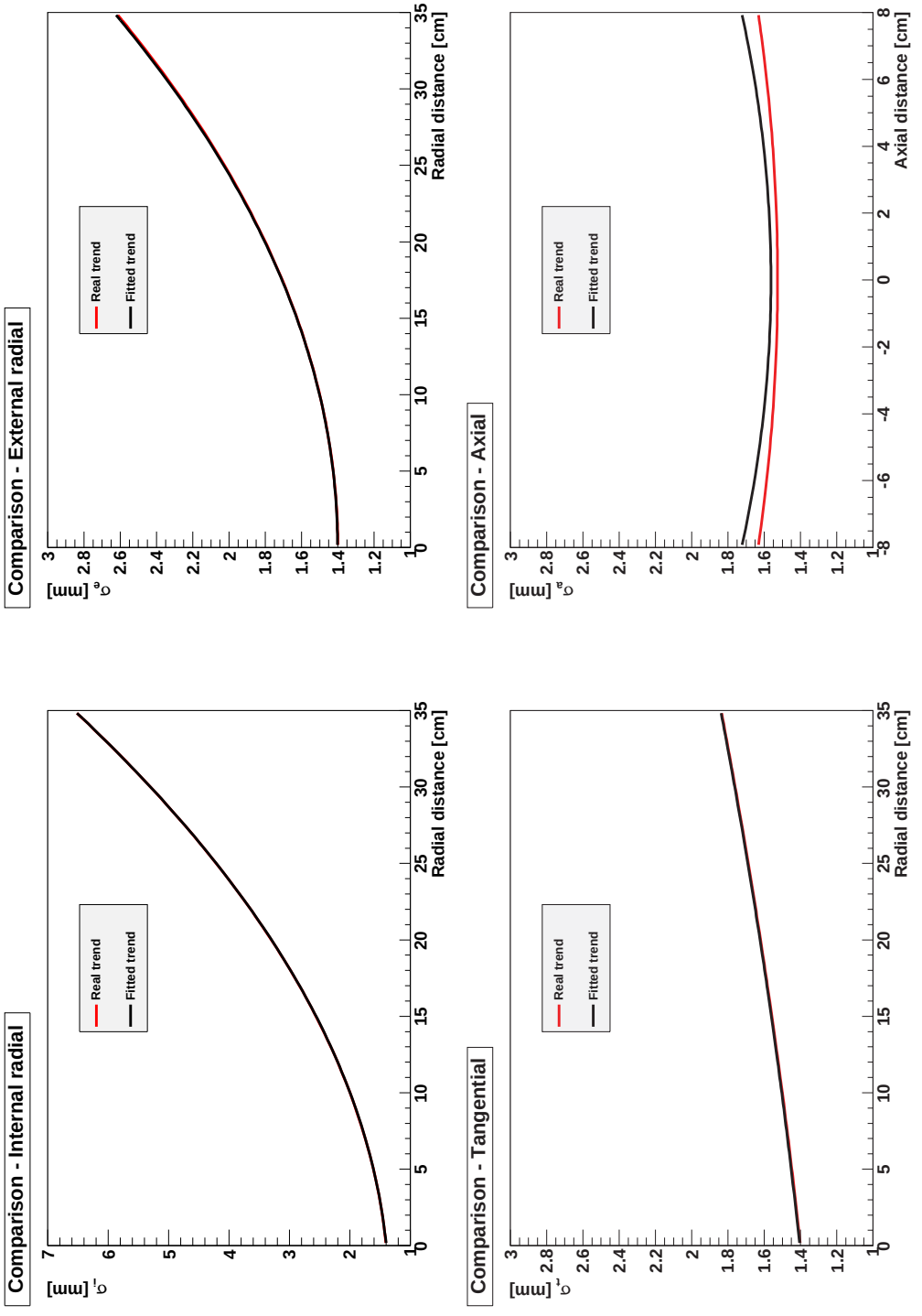


Figure 4.7: Comparison of evaluated (black) and known (red) spatial dependences of the spread parameters in simulated PSF: σ_i (top left), σ_e (top right), σ_t (bottom left) and σ_a (bottom right)

Chapter 5

Implementation

5.1 OSEM algorithm and PSF implementation

To reconstruct the data, the GE scanners used in this thesis employ an OSEM algorithm [45, 87] as described by the equation:

$$\lambda_b^{[k+1]} = \frac{\lambda_b^{[k]}}{\mathrm{B}_b A_d} \mathrm{B}_b \frac{A_d y_d}{A_d \mathrm{P}_d \lambda_{b'}^{[k]} + R_d + S_d} \tag{5.1}$$

where $\lambda_b^{[k]}$ indicates the counts in the voxel b of the image λ at iteration k, A_d is the attenuation correction factor relative to the LOR d, y_d are the counts recorded along LOR d (pre-corrected for normalization, decay, dead time and geometric effects), while R_d and S_d are the estimations of random and scattered coincidences relative to LOR d, respectively; finally, the projector P and the backprojector B are defined as

$$\begin{aligned} \mathrm{P}_d(\cdot)_b &= \sum_b (\cdot)_b \, p_{bd} \\ \mathrm{B}_b(\cdot)_d &= \sum_{d \in K_m} (\cdot)_d \, p_{bd} \end{aligned} \tag{5.2}$$

where p_{bd} is the weight factor linking voxel b and LOR d (i.e. the probability of detecting, in LOR d, an event coming from voxel b) and K_m indicates the m^{th} subset of projection angles.

The algorithm was implemented in MATLAB©*.

Starting from it, the measured PSF has been implemented as a "convolution" with the image during the reconstruction process: this approach assumes that the degradation of the spatial resolution due to the PSF is accounted for in the *image*

*MATLAB© is a product of The MathWorks™, http://www.mathworks.com.

space and not in the raw data or *sinogram* space[†]. This latter approach has often been used by other groups (e.g. [3, 70, 78]), while image–based implementation is less frequently found in literature (some examples are [16, 82, 83]). Both methods can only approximate the actual resolution model, which is complex due to the superposition of several different effects. Moreover, some of these effects are characteristic of the sinogram level (e.g. the detector blurring), while others are more relative to the image level (e.g. the positron range). On the other hand, from a theoretical point of view the two strategies should convey the same information about the response of the tomograph, especially in cases where the system response varies slowly across the FOV. As far as the estimation of the PSF is concerned, the sinogram approach is usually more involved than the image one, due to the complex organization of 3D-planogram data, whereas the image approach requires an additional reconstruction step (as described in section 4.1).

In this work the PSF was introduced by redefining the projector and back-projector as follows:

$$\begin{aligned} \mathrm{P}_d(\cdot)_b &= \sum_b \left[(\cdot) * PSF\right]_b \, p_{bd} \\ \mathrm{B}_b(\cdot)_d &= \left[PSF^T * \left(\sum_{d \in K_m} (\cdot)_d \, p_{bd} \right) \right]_b \end{aligned} \tag{5.3}$$

where $*$ indicates a redefinition of the convolution operator (see section M.3) and PSF^T indicates the transposed PSF. For a spatially-variant kernel this is not simply the transposed of the PSF kernel (as in the spatially-invariant case): for each voxel, instead, it contains contributions from neighbouring voxels, too (see section M.4).

5.2 Organization of the PSF kernel

The PSF was stored into two matrices: one for the transaxial two-dimensional Gaussian function, the other for the axial one-dimensional Gaussian function.

The transaxial PSF component was stored in a three-dimensional matrix composed of $N \times N \times M^2$, where N is the number of pixels of the reconstructed single slice and M is an odd number representing the *kernel size* of the spread. It is important to note that M is not a measure of the spread — it should not be confused with σ_i, σ_e and σ_t — since it represents the side (in pixels) of the square

[†]Please note that, in principle, a PSF determined on the sinogram could be used in the image space and a PSF determined on the image could be used in the sinogram space: however, this would obviously require some additional processing after the measurement procedure, which requires more time and, more importantly, introduces new sources of uncertainty. For this reason, when referring to either image or sinogram space for the implementation, it will be implicitly assumed that the PSF has also been *determined* in the corresponding space.

kernel containing the spatially variant two-dimensional transaxial PSF. M can be adjusted depending on the transaxial FOV width, to let about 4 FWHM (in each direction) be contained in the kernel in order to obtain a complete sampling in the tails of the PSF itself, avoiding a leakage in the PSF representation.

For each pixel P in the two-dimensional matrix sized $N \times N$, the position of the pixel centre inside the scanner FOV was determined (by knowing the chosen coordinates for the image centre): then, the distance from the centre of the transaxial FOV was calculated and consequently $\sigma_i(P)$, $\sigma_e(P)$ and $\sigma_t(P)$ were known from the fits (equation 4.8). Moreover, also the angle between the horizontal direction and the radial direction in P was obtained, in order to have the analytic expression for the properly rotated PSF (see section M.3). Finally, after filling the square kernel, this was reorganized (with leading row direction) into a $(M^2 \times 1)$ vector to gain higher computational performance. In figure 5.1 an example of a square kernel and the corresponding one-dimensional vector is provided.

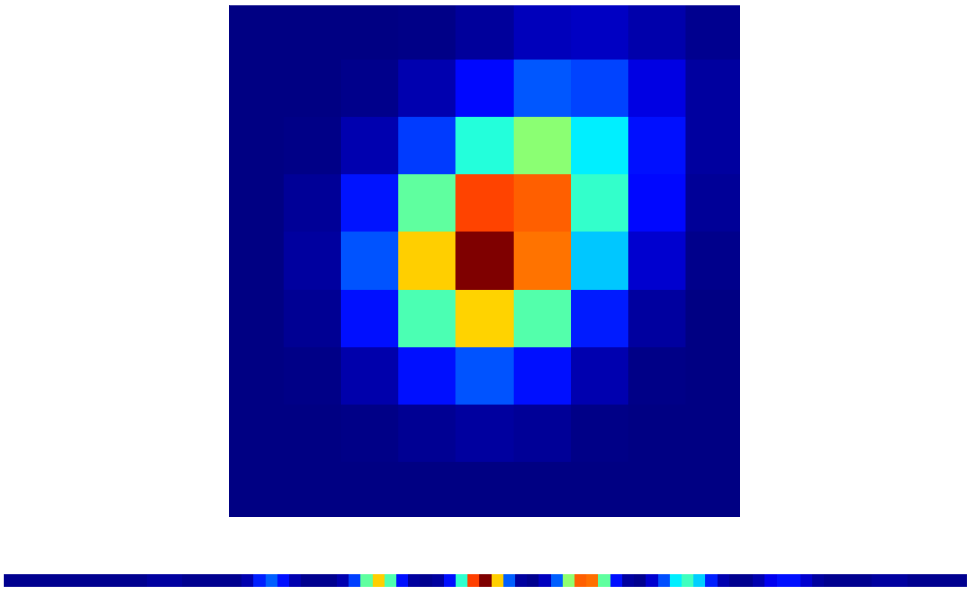


Figure 5.1: Organization of the transaxial PSF storage for a particular transaxial position: square kernel (top) and its organization into a one-dimensional vector (bottom)

The axial PSF was stored in a three-dimensional matrix $N \times N \times (M_A \times N_{SLICES})$, where N is the above-mentioned number of pixels of the reconstructed transaxial images, M_A is the length of the one-dimensional kernel vector along the axial direction of the scanner and N_{SLICES} is the number of slices in the reconstructed image. Since the axial PSF depends on both the transaxial and axial distances from the scanner centre, for each couple of row and column indexes

(identifying a point P in the transaxial FOV) a series of N_{SLICES} vectors of length M_A had to be stored, each containing the axial PSF vector relative to one slice. This strategy was not necessary for transaxial PSF, since σ_i, σ_e and σ_t were assumed (and experimentally verified, see chapters 7 and 8) to be independent of the axial position in the scanner. In figure 5.2 an example of the axial PSF organization is provided. Also for the axial case, the information relative to a fixed transaxial point was reorganised into a single one-dimensional vector, to improve the computational performance.

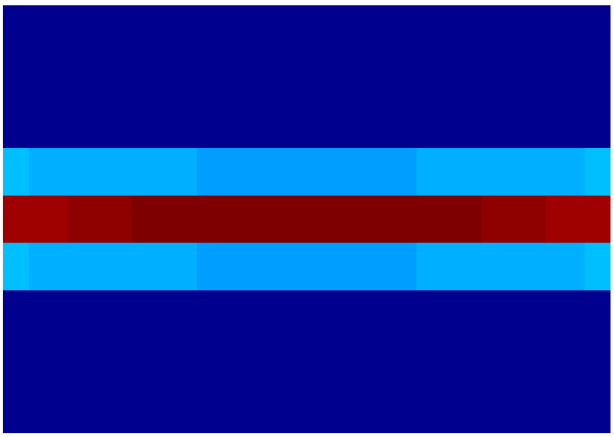


Figure 5.2: Organization of the axial PSF storage for a particular transaxial position: each slice is coded into a single column and the one-dimensional axial kernel relative to a particular slice is stored in the corresponding column

Integral vs. mean

To obtain the PSF for a particular voxel size, the value of each kernel coefficient was calculated as the *integral* of the PSF over the corresponding voxel, instead of the value of the PSF in the middle point of the voxel, strategy that should provide a description of the PSF more representative of the real one, due to the continuous behaviour of the resolution degradation. This can be understood by recalling that each voxel in the image contains the sum of counts assumed to have been originated from it: consequently, the resolution degradation effect is "integrated" over the voxel. This may be neglected for small pixels, where the "integration" is over a small region; instead, when the pixel size increases, the incidence of this behaviour enlarges. As an example, in figure 5.3 a one-dimensional comparison between the two different methods of calculation is performed for small pixel size (0.5 mm) and large pixel size (2.73 mm, the largest — and most used — pixel size available on the systems used for the validation in chapters 7 and 8). As expected, in the former case the approximation of calculating the PSF in the middle points of the voxels is good, while for large pixel sizes the approximated method changes both the height and, more importantly, the *shape* of the PSF.

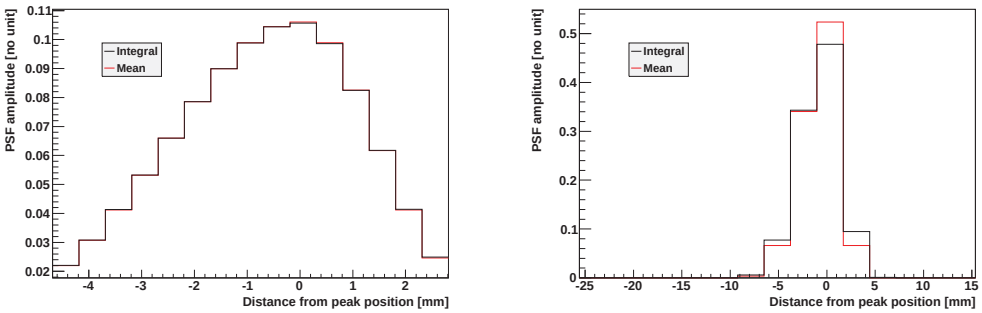


Figure 5.3: Comparison between PSF kernels calculated in the middle point of each voxel (red) or as the integral over the voxel (black) for small pixel size (0.5 mm, left) and large pixel size (2.73 mm, right).

5.3 The edge effect

As will be shown in the following chapters, the incorporation of PSF modelling in the PET reconstruction algorithm provides improved spatial resolution and image quality. However, an important effect observed in PSF-based reconstructions is the enhancement of regions with sharp intensity transitions, with the edges of the objects being strongly (and innaturally) emphasized if PSF recovery is used. An example of this *edge effect* or *edge artefact* is provided in figure 5.4, which represents three spheres (whose diameters are 30 mm, 23 mm and 15 mm) filled with a *homogeneous* FDG solution reconstructed in absence and in presence of PSF recovery.

This effect is recognizable throughout the literature — for some examples, see [4, 32, 67, 70, 78, 83, 89, 91] — and it is usually justified with a kernel not matched (in particular, overestimated) with the actual spread.

(a) Original reconstruction (b) Reconstruction with PSF

Figure 5.4: Edge effect in a reconstruction of three uniform spheres

To investigate this effect more deeply, two simple simulations of one-dimensional and two-dimensional data iterative reconstruction were set-up.

One-dimensional simulation

A box of given amplitude superimposed to a fixed background (blue line in figures 5.5 and 5.6) was used as the test image to be reconstructed. The box was convolved with a symmetric GF having $\sigma = \sigma_r$ and the resulting array simulated the recorded data $\mathbf{y}$ read on the detectors (green line in figure 5.6). The PSF model in the reconstruction algorithm was a symmetric GF with $\sigma = \sigma_u$. In the examples here provided, the box had a width of 3 cm and an amplitude of 25, while the chosen[‡] background was equal to 5; the simulated signal-to-background (SBR) ratio was then 5:1. The entire FOV of 70 cm was reconstructed on $N = 256$ pixels and the spread GF had $\sigma_r = 3$ mm. The number of iterations used was — where not differently specified — 280, corresponding to 10 full iterations at 28 subsets.

No attenuation and noise effects were introduced and the projector and back-projector operators were the identity operators ($p_{bd} = \delta_{bd}$). The updating rule in the simulated model was then

$$\lambda_b^{[k+1]} = \frac{\lambda_b^{[k]}}{(PSF^T \star \mathbf{1})_b} \left\{ PSF^T \star \frac{\mathbf{y}}{\left(\lambda^{[k]} \star PSF\right) + \mathbf{B}} \right\}_b \tag{5.4}$$

where λ_b represents the profile to be reconstructed, $\mathbf{1}$ is a N–dimensional array of ones, $\mathbf{B}$ is an array containing the estimation of background for each pixel and the division between arrays is intended element–by–element.

If no PSF recovery is applied, the reconstructed image coincides with the recorded data (figure 5.5). If PSF recovery is used, instead, the artifacts appear regardless of the recovery PSF used. Some examples are provided for a GF with $\sigma_u = 3$ mm (corresponding to the correct PSF model, figure 5.6a), a GF with $\sigma_u = 3.2$ mm (corresponding to a slightly overestimated PSF model, figure 5.6c) and a GF with $\sigma_u = 2.8$ mm (corresponding to a slightly underestimated PSF model, figure 5.6e). It is apparent that an underestimation of the PSF mitigates the effect, but obviously at the expense of the recovery of the spatial resolution (the "most underestimated" PSF corresponds to the absence of PSF recovery — figure 5.5 — which shows no edge artifacts). It is interesting to note that at a very high number of iterations (here 28000) the similar behaviours shared by correct and uncorrect PSF models differentiate: the edge artifact disappears from the image obtained with the exact PSF (figure 5.6b), while in the case of the overestimated PSF model the edge effect is amplified (figure 5.6d). As stated above, if the PSF is underestimated, the effect is still nearly absent, but the algorithm converges to an image with worse spatial resolution (figure 5.6f).

[‡]In the simulation, the background estimation used in the reconstruction can be different from the real one: however, as it will be seen, the major problem is not represented by this difference. Consequently, for the sake of simplicity, the correct background is used in the presented examples.

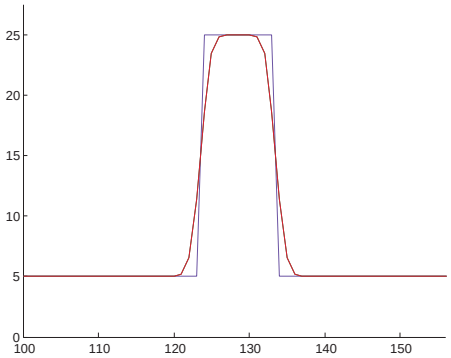


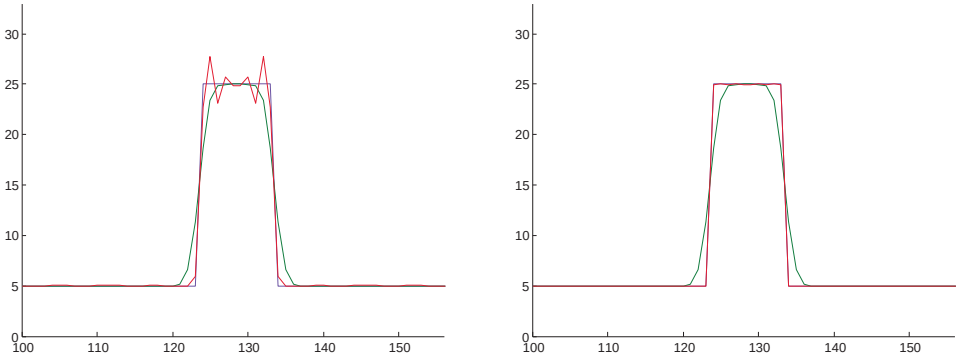
Figure 5.5: One-dimensional simulation without PSF recovery: the blue line represents the real box profile, the green line represents the recorded profile and the red line represents the reconstructed profile (coincident with the recorded one if no PSF recovery is applied)

Using the correct image as the start condition makes the reconstruction useless, so edge artifacts do not appear avoidable by choosing an adequate starting image and in feasible reconstruction conditions (e.g. the number of iterations). The edge amplification seems to be a manifestation of aliasing effects due to *truncation of the PSF kernel* (as suggested in [89]) and so no methods of prevention appear efficacious.

Two-dimensional simulation

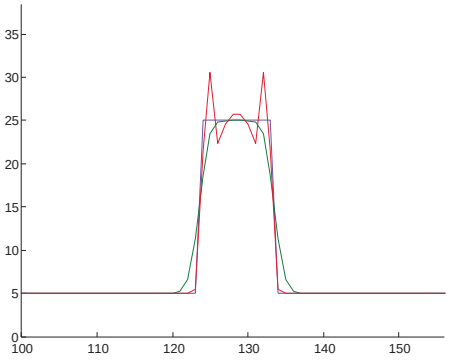
A two-dimensional simulation was performed similarly to the one-dimensional case to compare the simulated results to the images presented in figure 5.4 more directly.

A circle of given diameter superimposed to a fixed background (figure 5.7a) was used as the test image to be reconstructed. The circle was convolved with a symmetric GF having $\sigma = \sigma_r$ and the resulting image simulated the recorded data $\mathbf{y}$ to be "deconvolved" (figure 5.7b). The PSF model was a symmetric GF with $\sigma = \sigma_u$. In the examples here provided, the circle had a diameter of 3 cm and a simulated activity concentration of 5 kBq/cc, while the background activity concentration was 1 kBq/cc. The image simulated a FOV of 70 cm on 256×256 pixels (with a corresponding pixel size of $2.73 \times 2.73 \times 3.27$ mm^3 — where the slice width of 3.27 mm was chosen equal to the one of the considered scanners in chapters 7 and 8, even if its value is irrelevant for the simulation) and the spread GF had $\sigma_r = 3$ mm. Each pixel in the image was filled with the expected counts contained (neglecting attenuation and noise fluctuations). The number of iterations used was — where not differently specified — 280, corresponding to 10 full iterations at 28 subsets. The updating rule of the algorithm was identical to

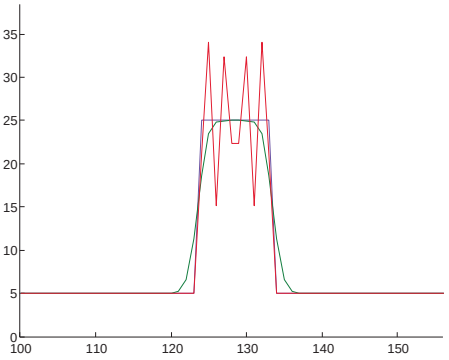


(a) Correct PSF

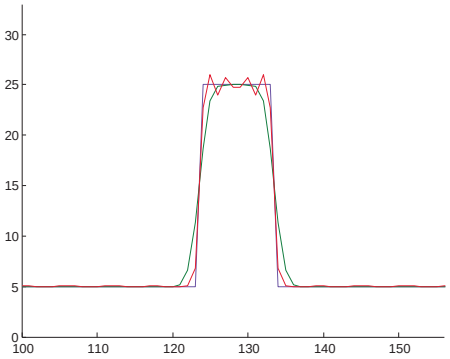
(b) Correct PSF — large iterations



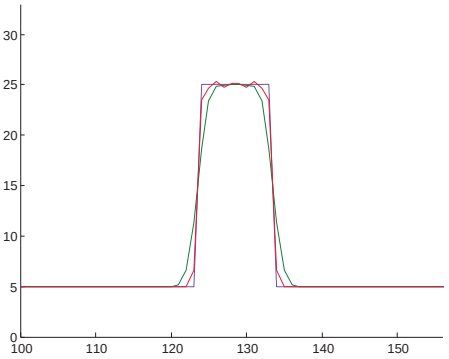
(c) Overestimated PSF



(d) Overestimated PSF — large iterations



(e) Underestimated PSF

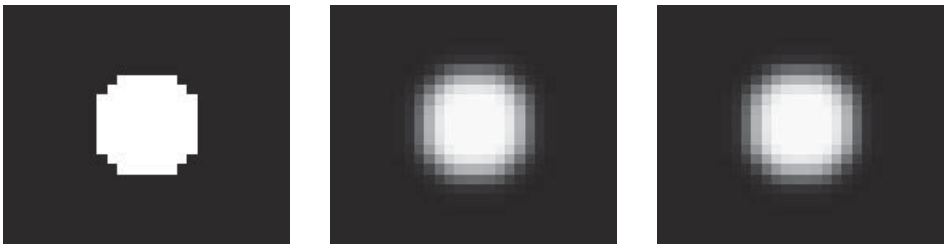


(f) Underestimated PSF — large iterations

Figure 5.6: Main results of data reconstruction one-dimensional simulation with PSF recovery (the blue line represents the real box profile, the green line represents the recorded profile and the red line represents the reconstructed profile): PSF with $\sigma_u = 3$ mm (top), PSF with $\sigma_u = 3.2$ mm (middle), PSF with $\sigma_u = 2.8$ mm (bottom); 280 iterations (left) and 28000 iterations (right)

the one-dimensional case (equation 5.4).

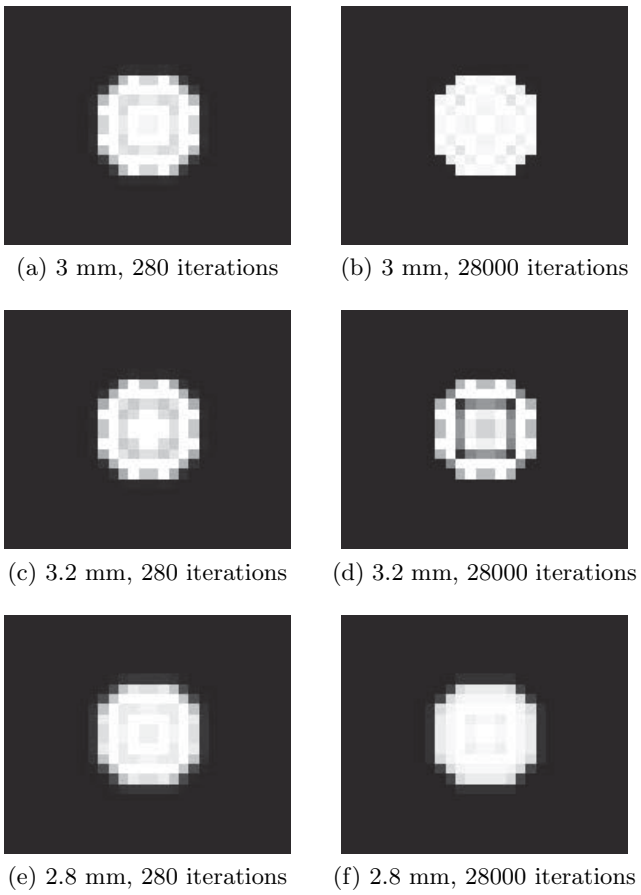
If no PSF recovery is employed, the image remains identical to the recorded data $\mathbf{y}$ (figure 5.7c). If the correct PSF is used, the edge artefact appears (figure 5.8a) very similar to the reality (compare it with figure 5.4). If the PSF is either overestimated ($\sigma_u = 3.2$ mm) or underestimated ($\sigma_u = 2.8$ mm), the artefact is respectively strengthened (figure 5.8c) or lessened (figure 5.8e). This confirms the suggestion found in literature (e.g. see [83]) of underestimating the PSF in order to reduce the edge enhancement. However, as visible in figure 5.8e, in such a case the recovery of spatial resolution is suboptimal, leading to worse definition of the signal (e.g. the smeared edge of the circle). If the number of iterations is strongly increased (28000 iterations), the effect should disappear if the correct PSF were used. In two dimensions the effect appears to be removed more slowly than in one dimension (compare figures 5.8b and 5.6b) and, consequently, even more iterations would be needed to remove it to a sufficient extent for the clinical practice. A comparison between the results with correct PSF after 280, 28000 and 2800000 iterations is presented in figure 5.9. After 28000 iterations with an overestimated PSF (figure 5.8d) the effect amplifies, generating an image very different to the real one; if the PSF is underestimated, instead, (figure 5.8f), the lower edge artefact leads to a reduced spatial resolution recovery, which becomes apparent when compared to the original image (figure 5.7a) and the image obtained with the correct PSF at the same number of iterations (figure 5.8b).



(a) Original image (b) Recorded data (c) No PSF recovery

Figure 5.7: Data used for two-dimensional simulation of edge effect (left, center) and image resulting from no PSF recovery (right)

Both simulations show that the edge effect is avoidable since the beginning only if no iterative updates are necessary, i.e. if the correct image is used as the starting condition of the algorithm. This obviously makes the reconstruction process useless and is a non-feasible approach in the real world. If, instead, a different image is used as the starting condition of the algorithm (in these examples, a uniform one), the effect is present since the first iterations and tends to disappear only if the PSF used for the recovery is correct. Apart from the



(a) 3 mm, 280 iterations (b) 3 mm, 28000 iterations

(c) 3.2 mm, 280 iterations (d) 3.2 mm, 28000 iterations

(e) 2.8 mm, 280 iterations (f) 2.8 mm, 28000 iterations

Figure 5.8: Main results of data reconstruction two-dimensional simulation with PSF recovery: PSF with $\sigma_u = 3$ mm (top), PSF with $\sigma_u = 3.2$ mm (middle), PSF with $\sigma_u = 2.8$ mm (bottom); 280 iterations (left) and 28000 iterations (right)

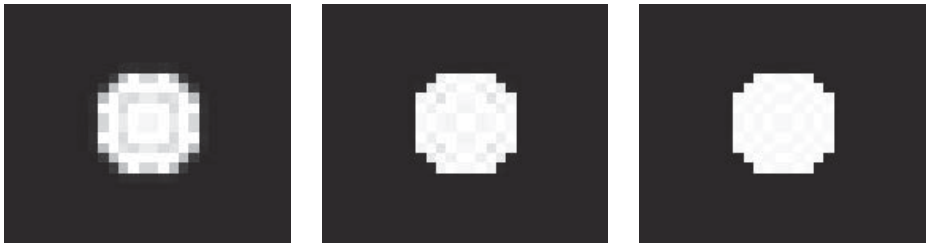


Figure 5.9: Removal of edge artefact with increasing iterations using the correct PSF (σ_u=3 mm): 2.8×10^2 iterations (left), 2.8×10^4 iterations (middle), 2.8×10^6 iterations (right)

impossibility of using a perfectly matched PSF kernel, also the number of iterations required for the removal of the artifact is completely impractical for the real reconstructions. Moreover, the simulations show also that the necessary number of iterations is dramatically dependent on the complexity of the reconstruction: e.g. the change from one-dimensional to two-dimensional simulation makes the required number of iterations increase by about two orders of magnitude. This strengthens the impossibility of using such a method for real reconstructions, which are two-dimensional or, much more often, three-dimensional.

Consequently, no methods of *prevention* appear available and only strategies to *reduce* this effect are usable, one of which is presented in chapter 6.

Chapter 6

Regularization

6.1 Theoretical background

As stated in section 3.5.1, the ML algorithms seek an image making the measured data most likely to occur, i.e. the reconstructed image is the one maximizing the *conditional* probability

$$P(Data|Img) \tag{6.1}$$

of obtaining the projection data $Data$ from the image Img. Being an example of an inverse problem [103], this process generally converges to a quantitatively unbiased but noisy image: consequently, the final result may appear visually unpleasant and thus difficult to be interpreted. To contrast this, a limited number of iterations is usually chosen, reducing the quantitative accuracy due to an incomplete convergence. The incomplete convergence also introduce a strong dependence of the results on the reconstruction parameters (and, specifically, on the number of iterations), causing difficulties in comparing quantitative information from different clinical protocols.

Another approach is to introduce the concept of *evaluation* of the results, allowing to give privileges to a class of images defined in advance, while penalising images not contained in this class. This scheme is based on the Bayes' theorem

$$P(Img|Data) = \frac{P(Data|Img)\,P(Img)}{P(Data)} \tag{6.2}$$

While the MLEM methods are based on the maximization of expression 6.1, the Bayesian methods maximize expression 6.2, where $P(Img|Data)$ is the probability of obtaining image Img having recorded the projections $Data$, $P(Data)$ is the total probability of obtaining the recorded data $Data$ (but, since it does not involve the image Img, its contribution to the image estimation is null) and $P(Img)$ is the *a–priori* probability of obtaining the image Img (i.e. certain kinds

of images are assumed to be most probable and/or to be favoured), thus conditioning the usual MLEM reconstruction.

The expectation step in expression 3.8 is not affected, while the maximization step changes from

$$\left\{\lambda_b^{[k+1]}\right\}_{MLEM} = \arg\max_{\lambda} \left[\log P(Data|Img)\right] = \arg\max_{\lambda} \left[L(\lambda)\right]$$

to

$$\begin{aligned}\left\{\lambda_b^{[k+1]}\right\}_{Bayes} &= \arg\max_{\lambda} \left[\log P(Data|Img) + \log P(Img)\right] = \\ &= \arg\max_{\lambda} \left[L(\lambda) + P(\lambda)\right]\end{aligned}$$

The resulting algorithm is usually referred to as *maximum a posteriori* (MAP), since the reconstruction follows an a-priori constraint. This assumption, contained in $P(Img)$, is crucial for the final result: since the scope of emission tomography reconstruction is to reveal the unknown activity concentration, too strict conditions can mask possible deviations from the normal uptake of the tracer. Consequently, the prior should be ideally capable of leaving the correct image unpenalised, while removing noise and all other spurious effects associated with the reconstruction.

The Bayesian prior is generally formulated according to the Gibbs distribution, whose general form is [36, 40]

$$P(Img) = Ce^{-\beta U(\lambda)}$$

where C is a normalization constant and $U(\lambda)$ is a non–negative energy function having its minimum — corresponding to the maximum of $P(Img)$ — when the image λ meets the prior assumption. The parameter β expresses the relative weight of the prior. If β is close to 0, $P(Img)$ is close to its maximum over a wide range of different λ images and, consequently, the strength of penalisation is modest; with a large β, instead, the prior is more peaked and some images are significantly more favoured than others.

Since the image is discrete, the function $U(\lambda)$ is usually calculated as the sum of the values $U(\lambda, b)$ assumed by $U()$ in each voxel b:

$$U(\lambda) = \sum_b U(\lambda, b) \tag{6.3}$$

A common choice for U is an energy function computed using a potential function $v()$ of the differences between pixels in the neighbourhood N_b of pixel b

$$U(\lambda, b) = \sum_{i \in N_b} w_{bi}\, v\left(\lambda_b - \lambda_i\right) \tag{6.4}$$

where w_{bi} is the weight of pixel i contained inside N_b. Typically, the w_{bi} coefficients are chosen depending on the distance from the central voxel b, following a spatial organization scheme like

$$\begin{matrix} e & d & c & d & e \\ d & b & a & b & d \\ c & a & \star & a & c \\ d & b & a & b & d \\ e & d & c & d & e \end{matrix}$$

where $\star$ indicates the central voxel and is usually set to zero to avoid any regularization effect of each voxel on itself. Different potential functions $v()$ have been proposed in literature [1, 15, 55, 75–77].

As stated before, the M step including the regularization prior becomes

$$\frac{\partial L(\lambda)}{\partial \lambda_b} + \frac{\partial P(\lambda)}{\partial \lambda_b} = 0$$

The same mathematical steps presented in section 3.5.1 and the Gibbs formulation of the prior lead to

$$\sum_{d=1}^{D} \left[-p_{bd} + \frac{x_{bd}^{[k+1]}}{\lambda_b^{[k+1]}} \right] - \beta \frac{\partial U(\lambda)}{\partial \lambda_b} = 0$$

The *one step late* (OSL) algorithm uses the image at the iteration k to calculate the derivative of the energy function $U()$ and penalise the image at iteration $k+1$ [38]. This allows decoupling $\lambda_b^{[k+1]}$ from the prior term: therefore, the updating rule of OSL-MLEM accounting for attenuation, random and scatter coincidences and with the use of Gibbs regularization is

$$\lambda_b^{[k+1]} = \frac{\lambda_b^{[k]}}{\sum_{d=1}^{D} A_d p_{bd} + \beta \left. \frac{\partial U(\lambda)}{\partial \lambda_b} \right|_{\lambda=\lambda^{[k]}}} \sum_{d=1}^{D} \frac{A_d y_d p_{bd}}{A_d \sum_{b'=1}^{B} \lambda_{b'}^{[k]} p_{b'd} + R_d + S_d} \tag{6.5}$$

6.1.1 Variational regularization

One of the most relevant features in an image is represented by the *gradients* contained. The gradients are directly involved in the recognition of particular patterns and structures and in the discrimination of signal and background.

The previously described approach is not directly linked with the gradient analysis, since it takes into account the *difference* between neighbouring voxels (see equation 6.4) regardless of the orientation of this variation. As an example, the two images

0	1	0
0	2	1
0	0	0

0	0	0
1	2	1
0	0	0

would give the same value for $U(\lambda)$ and, consequently, the same penalization term, while they clearly exhibit different patterns of gradients. Moreover, the weight coefficients in equation 6.4 are, in principle, arbitrary and not directly linked with the potential functions $v()$ used: consequently, different choices of w_{bi} for the same regularization function may lead to different results.

By changing the definition of the energy function (equations 6.3 and 6.4), the gradient information can be directly used, while avoiding the arbitrariness of the weight coefficients: this is possible, for example, by using a *variational* approach* [7, 50, 84]:

$$U(\lambda) = \int_{\Omega} \phi(|\nabla \lambda(\omega)|)\, d\omega \tag{6.6}$$

where $\phi(\cdot) \geq 0$ is the *prior* function (or, in the following, simply the *prior*) and ω spans the image domain Ω. The symbol ∇ indicates the *gradient operator* expressed using the ℓ^2 norm.

With the energy function chosen as in equation 6.6, the updating rule of OSL-MLEM (equation 6.5) becomes

$$\lambda_b^{[k+1]} = \frac{\lambda_b^{[k]}}{\sum_{d=1}^{D} A_d p_{bd} - \beta \mathrm{D}_b\left(\lambda^{[k]}\right)} \sum_{d=1}^{D} \frac{A_d y_d p_{bd}}{A_d \sum_{b'=1}^{B} \lambda_{b'}^{[k]} p_{b'd} + R_d + S_d} \tag{6.7}$$

where the regularization term

$$\mathrm{D}(\lambda) = -\frac{\partial U(\lambda)}{\partial \lambda} = \nabla \cdot \left[\phi'(|\nabla\lambda|)\frac{\nabla\lambda}{|\nabla\lambda|}\right] = \nabla \cdot [g(|\nabla\lambda|)\nabla\lambda] \tag{6.8}$$

(with $g(t) = \phi'(t)/t$) results from the variational maximization of equation 6.2, using the energy function specified in equation 6.6 (see section M.5 for the explicit derivation).

The discrete form† for the regularization term $\nabla \cdot (g(|\nabla\lambda|)\nabla\lambda)$ is

$$\begin{aligned}[\nabla \cdot (g(|\nabla\lambda|)\nabla\lambda)]_{i,j,k} = \quad & \frac{1}{h_x}\Delta_-^x \left[g\left(|\nabla\lambda|_{i+\frac{1}{2},j,k}\right)\Delta_+^x \lambda_{i,j,k}\right] + \\ & + \frac{1}{h_y}\Delta_-^y \left[g\left(|\nabla\lambda|_{i,j+\frac{1}{2},k}\right)\Delta_+^y \lambda_{i,j,k}\right] + \\ & + \frac{1}{h_z}\Delta_-^z \left[g\left(|\nabla\lambda|_{i,j,k+\frac{1}{2}}\right)\Delta_+^z \lambda_{i,j,k}\right]\end{aligned} \tag{6.9}$$

*In order to avoid the introduction of new and redundant symbols, λ will indicate either a continuous functions describing the activity distribution or its discretization $\lambda = \{\lambda_b\}$. The meaning of the symbol in each specific case, however, will be clear from the context.

†The indexes used (i, j and k) refer respectively to the x, y and z directions and their domains of definition are

$$i = 1 \ldots N_x \qquad j = 1 \ldots N_y \qquad k = 1 \ldots N_z$$

in which the discrete derivatives are

$$\begin{array}{ll} \Delta_+^x s_{i,j,k} = \frac{1}{h_x}\left(s_{i+1,j,k} - s_{i,j,k}\right) & \Delta_-^x s_{i,j,k} = \frac{1}{h_x}\left(s_{i,j,k} - s_{i-1,j,k}\right) \\ \Delta_+^y s_{i,j,k} = \frac{1}{h_y}\left(s_{i,j+1,k} - s_{i,j,k}\right) & \Delta_-^y s_{i,j,k} = \frac{1}{h_y}\left(s_{i,j,k} - s_{i,j-1,k}\right) \\ \Delta_+^z s_{i,j,k} = \frac{1}{h_z}\left(s_{i,j,k+1} - s_{i,j,k}\right) & \Delta_-^z s_{i,j,k} = \frac{1}{h_z}\left(s_{i,j,k} - s_{i,j,k-1}\right) \end{array} \tag{6.10}$$

and the magnitudes of the (discrete) gradients are given by

$$\begin{aligned} |\nabla\lambda|_{i+\frac{1}{2},j,k} &= \sqrt{\left(\Delta_+^x\lambda\right)^2 + \left[m\left(\Delta_+^y\lambda, \Delta_-^y\lambda\right)\right]^2 + \left[m\left(\Delta_+^z\lambda, \Delta_-^z\lambda\right)\right]^2} \\ |\nabla\lambda|_{i,j+\frac{1}{2},k} &= \sqrt{\left(\Delta_+^y\lambda\right)^2 + \left[m\left(\Delta_+^x\lambda, \Delta_-^x\lambda\right)\right]^2 + \left[m\left(\Delta_+^z\lambda, \Delta_-^z\lambda\right)\right]^2} \\ |\nabla\lambda|_{i,j,k+\frac{1}{2}} &= \sqrt{\left(\Delta_+^z\lambda\right)^2 + \left[m\left(\Delta_+^x\lambda, \Delta_-^x\lambda\right)\right]^2 + \left[m\left(\Delta_+^y\lambda, \Delta_-^y\lambda\right)\right]^2} \end{aligned} \tag{6.11}$$

where λ is a short form for $\lambda_{i,j,k}$ and

$$m(a,b) = \frac{\operatorname{sign} a + \operatorname{sign} b}{2} \min\left(|a|, |b|\right)$$

with the signum function defined as

$$\operatorname{sign} x = \begin{cases} \dfrac{|x|}{x} & x \neq 0 \\ 0 & x = 0 \end{cases} = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

The Neumann boundary conditions (see above and section M.5) are discretized by extending the domain of definition of the indexes i, j, k and assuming

$$\begin{array}{lll} \lambda_{0,j,k} = \lambda_{1,j,k} & \lambda_{i,0,k} = \lambda_{i,1,k} & \lambda_{i,j,0} = \lambda_{i,j,1} \\ \lambda_{N_x,j,k} = \lambda_{N_x+1,j,k} & \lambda_{i,N_y,k} = \lambda_{i,N_y+1,k} & \lambda_{i,j,N_z} = \lambda_{i,j,N_z+1} \end{array}$$

The effects of a variational regularization strategy on the final image depend on the mathematical characteristics of the prior $\phi(x)$. In particular, *convex* functions ($\phi''(x) > 0$) smooth the image (with increasing effect for larger values of $\phi''(x)$), lowering the noise level but reducing the sharpness of the edges in the image, while *concave* functions ($\phi''(x) < 0$) enhance the edges (with increasing effect for larger values of $|\phi''(x)|$), even if leading to possible amplification of noisy textures and creation of "patchy" artefacts. Both these behaviours are, in principle, contemporarily needed for PET, since a reconstruction algorithm for PET should be able to smooth uniform areas (in order to take noise under control, all the more – if the algorithm is iterative – while iterations proceed) and, contemporarily, to preserve the spatial resolution, to obtain high definition of the activity distribution.

For this reason, the case with $\phi''(x) = 0$ has been subject of interest, since it corresponds to the so-called "total variation" (TV) function, which neither smooths nor enhances edges: this specific penalization provides an excellent edge

preservation and smoothing of flat regions. Unfortunately, in regions with gradual variations (as often encountered in PET) it may lead to unnatural staircasing effects.

Many different penalty functions have been proposed [50]. The most used Bayesian regularization strategy is based on the Gaussian prior [36], which corresponds to the variational penalty function $\phi_G(x) = x^2/2$. Since $\phi''_G(x) = 1$ and considering that the function is used with argument $x = |\nabla\lambda|$, the prior smooths all the gradient intensities with equal relative strength, resulting in a strong smoothing effect.

A modification of the Gaussian prior is the Huber prior [44] or, equivalently in the variational framework, the Gauss-Total Variation (GTV) prior [50]:

$$\phi_{GTV}(x) = \begin{cases} x^2/2 & x < \delta \\ \delta x - \delta^2/2 & x \geq \delta \end{cases} = \begin{cases} \phi_G(x) & x < \delta \\ \phi_{TV}(x) & x \geq \delta \end{cases}$$

where the parameter δ is a threshold to discriminate between different regularization behaviours to be applied respectively to background regions – Gaussian (GR) component $\phi''_G(x) = 1$ for low x – for noise suppression and signal regions – Total Variation (TV) component $\phi''_{TV}(x) = 0$ for high x – for edge preservation.

Another promising strategy is represented by the generalized Gaussian or p-Gaussian prior (PR), as proposed in [15]: the corresponding variational penalty function is

$$\phi_P(x, p) = \frac{x^p}{p}$$

with $1 < p < 2$. Since its second derivative is $\phi''_P(x, p) = (p-1)\,x^{p-2}$, this regularization scheme has the capability of strongly penalizing small gradients, while reducing the smoothing effect on larger gradients.

6.2 Proposal of a new variational regularization prior

In this thesis a new prior is proposed as a generalization of $\phi_P(x, 4/3)$:

$$\phi(x) = \begin{cases} 3x^{4/3}/4 & x < \delta \\ x - d\,(d+\delta)\ln|x+d| + c & x \geq \delta \end{cases}$$

with

$$c = d\,[(d+\delta)\ln|d+\delta| - \delta] - \frac{\delta^{4/3}}{4}$$

$$d = \left(1 - \sqrt[3]{\delta}\right)$$

to have $\phi(x)$ and $\phi'(x)$ continuous in $x = \delta$. Consequently,

$$\phi'(x) = \begin{cases} \sqrt[3]{x} & x < \delta \\ \dfrac{x + d\,(1-d-\delta)}{x+d} & x \geq \delta \end{cases}$$

and

$$\phi''(x) = \begin{cases} \dfrac{1}{\sqrt[3]{27x^2}} & x < \delta \\ \dfrac{d\,(d+\delta)}{(x+d)^2} & x \geq \delta \end{cases}$$

The value $p = 4/3$ was chosen by considering a cylindrical phantom divided into six sectors, each with a different activity concentration (expressed in kBq/cm^3, 3.5, 11.0, 13.4, 16.7, 17.4, 22.0), immersed into an elliptical cylindrical tank (dimensions: 28×22×16.8 cm^3) filled with a uniform activity concentration of 1.9 kBq/cm^3. The activity concentrations were chosen to contemporarily simulate gradual (e.g. between the coldest sector and the background or between the intermediate sectors) and steep (e.g. between all sectors but the coldest one and the background) transitions, to test the different characteristics introduced by the p-Gaussian prior (i.e. the starting point of the proposed modification) in presence of small and large transitions, in particular the possible sawtoothing of edges as indicated in the literature (e.g. [6]). The phantom was acquired for 5 min, recording 50.9×10^6 counts, and then reconstructed with an algorithm with PSF recovery and variational regularization (see section 6.4) with p-Gauss prior, using different values of p (corresponding to different ϕ functions):

p	$1.01 = 101/100$	$1.2 = 6/5$	$1.\bar{3} = 4/3$	$1.5 = 3/2$	$1.8 = 9/5$
$\phi(x)$	$\frac{100}{101}\sqrt[100]{x^{101}}$	$\frac{5}{6}\sqrt[5]{x^6}$	$\frac{3}{4}\sqrt[3]{x^4}$	$\frac{2}{3}\sqrt{x^3}$	$\frac{5}{9}\sqrt[5]{x^9}$

In this phase, the value of β is non influential, since it is common to all the examined values of p. Consequently, to select the best p the empirical value of $\beta = 0.002$, capable of providing a good smoothing while retaining a globally natural appearance of the image with the different values of p, was chosen.

The left part of figure 6.1 presents the images of the phantom with the different choices of p, while on the center part of the figure a zoomed region of the phantom is shown. Taking the zoomed image with $p = 4/3$ as the reference image, the right column in figure 6.1 is obtained by subtracting the reference image to the center column: this operation should help showing the differences between the different values of p. As easily recognizable, the lowest value ($p = 1.01$) preserves high spatial resolution, but produces saw-toothed edges and innaturally flat regions, very similarly to the Total Variation behaviour [50]. The largest value considered ($p = 1.80$), instead, does not introduce any saw-tooth effect, but produces insufficient noise control and large spatial resolution loss. The intermediate values ($p = 1.2$, $p = 1.\bar{3}$ and $p = 1.5$) generate very similar images one another, with slightly higher spatial resolution and noise control, but also saw-tooth appearance, for smaller p values. For this reason, the central value $p = 4/3 = 1.\bar{3}$ appears to represent a good compromise.

The p value has been also heuristically studied by clinical evaluation of whole body PET images reconstructed with an algorithm using PSF and the generalized Gaussian prior, in which different values of the parameter p, spanning the range $1 < p < 2$, were used; once again, the value $p = 4/3$ gave the best compromise between preservation of spatial resolution, noise suppression and natural appearance of the clinical image.

This value, which is also in agreement with indications from other studies (e.g. [6, 109]), was thus chosen as a reference point for the modified prior.

The parameter δ, as described above, represents a threshold to discriminate between different regularization behaviours.

Below the threshold ($x < \delta$), since $\phi''(x) > 0$ the proposed prior smoothes with a strength independent of the threshold but depending only on x. Moreover, in the interval $0 < x < \alpha = 1/\left(3\sqrt{3}\right)$, this smoothing effect is greater than the GR one and, all the more, than the TV one (since $\phi''(x) > \phi''_G(x) > \phi''_{TV}(x)$), while in the interval $\alpha < x < \delta$ the smoothing effect is intermediate between GR and TV ($\phi''_{TV}(x) < \phi''(x) < \phi''_G(x)$).

Above the threshold ($x > \delta$), instead, the behaviour depends on the value of δ.

If $\delta < 1$, the prior smooths the image ($\phi''(x) > 0$), but much more lightly than the GR prior and with decreasing strength for larger x. Consequently, the result is a preservation of most edges, while globally maintaining "natural" transitions thanks to the very small smoothing effect.

If $\delta > 1$, the prior emphasises the edges ($\phi''(x) < 0$), contributing to a gain in spatial resolution. Consequently, the result is an enhancement of the contrast between regions of large x and regions with low x. In this case, however, it is important to note that the prior is not convex anymore.

Both for $\delta < 1$ and $\delta > 1$, for $x \to +\infty$, $\phi''(x) \to 0$, with a consequent asymptotic TV behaviour.

In regions above the threshold, for every choice of δ, the proposed prior smoothes less than the corresponding p-Gaussian prior $\phi_P(x, 4/3)$, while below the threshold the behaviour is identical.

Since the penalty function is used with argument $x = |\nabla\lambda|$, the strong smoothing is applied in background regions (in which $|\nabla\lambda|$ is small), while the "natural" edge preservation (or edge enhancement) is used in signal regions (i.e. regions with high $|\nabla\lambda|$).

A summary of the prior characteristics is provided in table 6.1, while in figure 6.2 a graphical comparison of $\phi(x)$ and $\phi''(x)$ for the Gauss-Total Variation, the p-Gaussian and the proposed priors is presented.

Both the GTV and the proposed regularization schemes are thus based on two parameters: β (i.e. the contribution of the regularization to the resulting image) and δ (i.e. the threshold to discriminate between background and signal

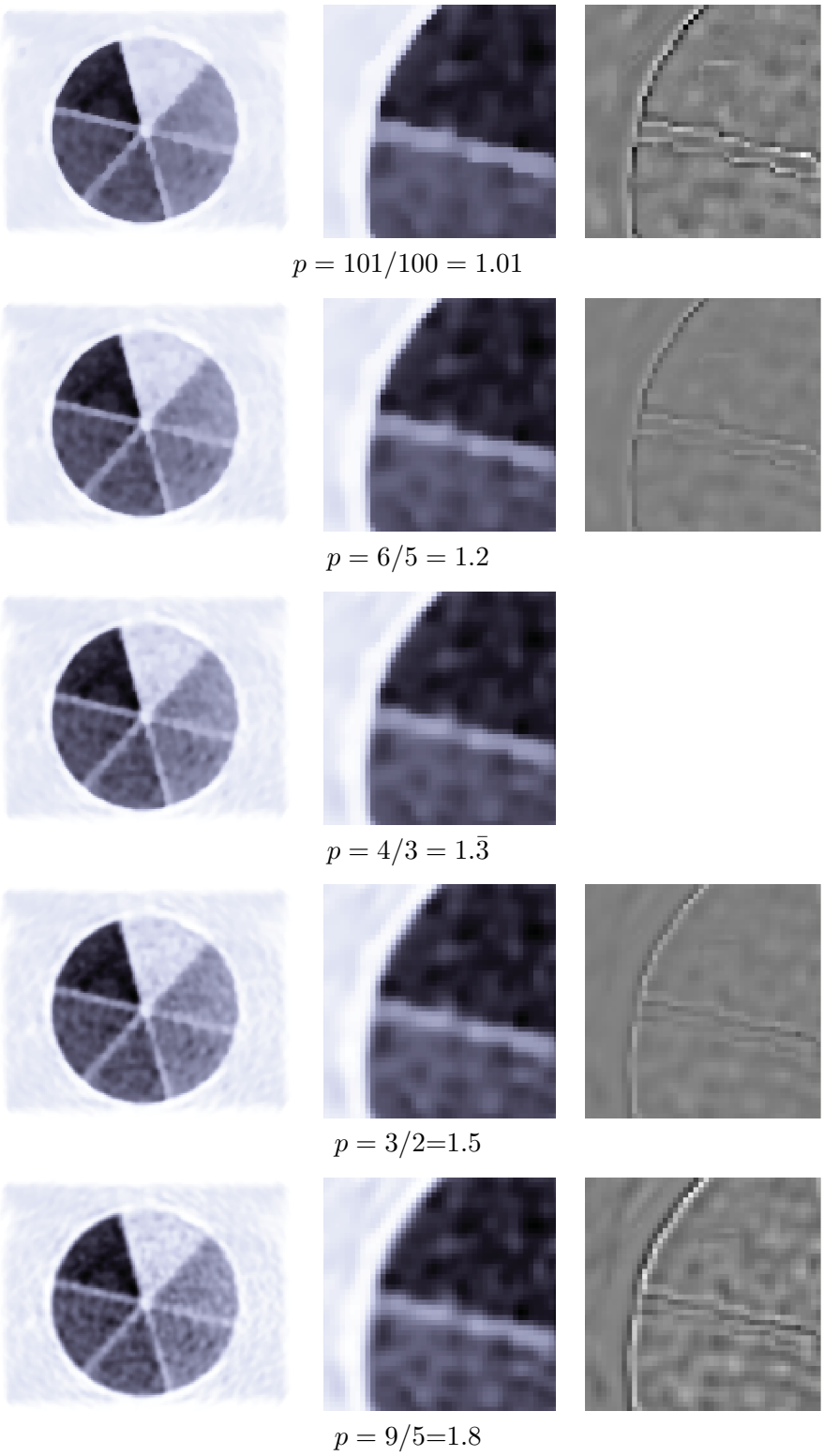


Figure 6.1: Effects of different p values in the p-Gaussian prior: left, sectorial phantom; center, zoom on a portion of the phantom; right: difference between the central column and the zoomed reference image ($p = 4/3$)

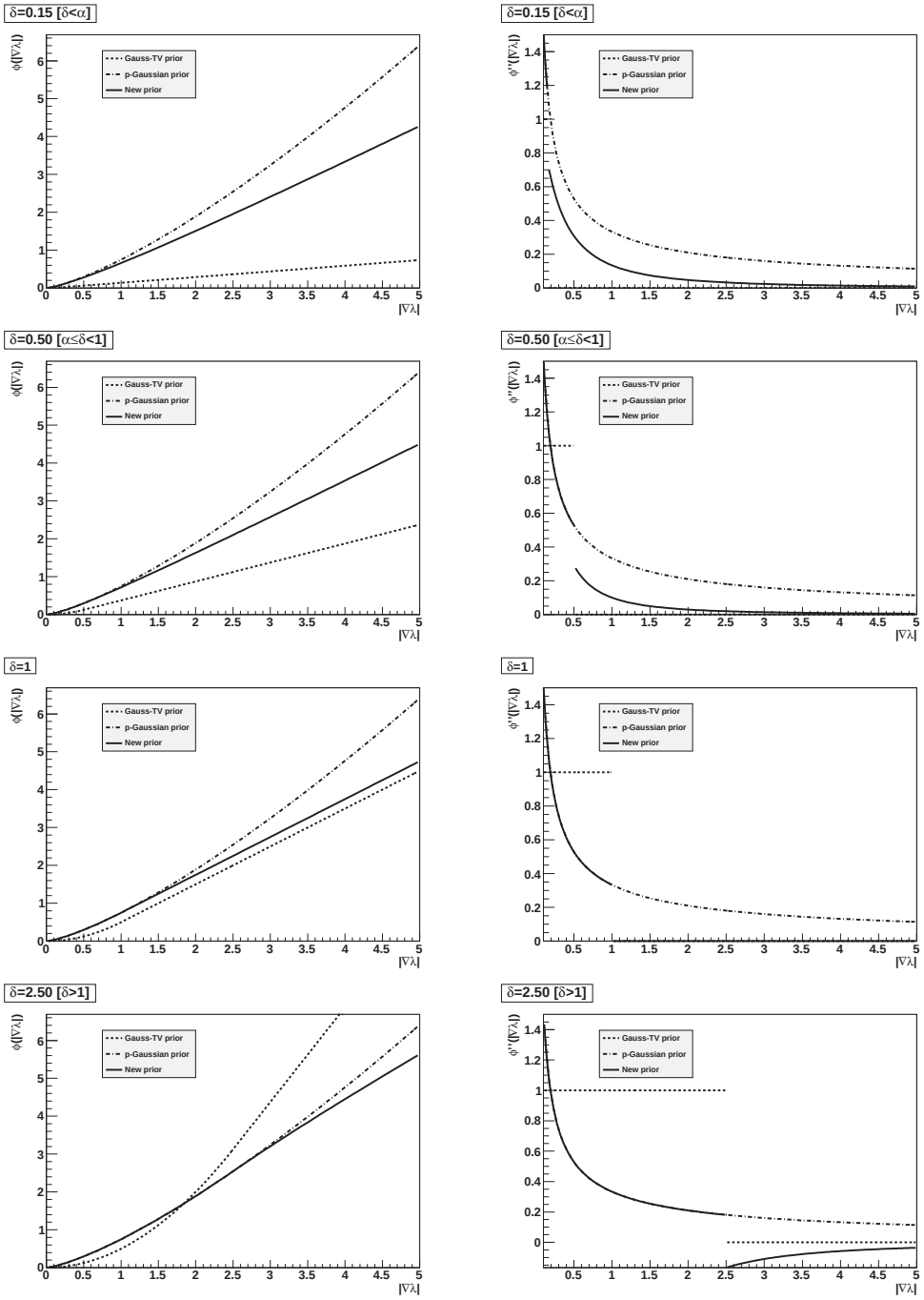


Figure 6.2: $\phi(x)$ (left) and $\phi''(x)$ (right) with different values of the parameter δ for the Gauss-Total Variation, the p-Gaussian and the proposed priors

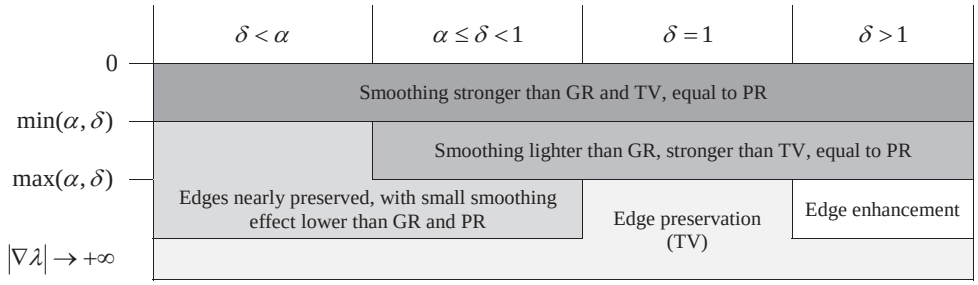


Table 6.1: Characteristics of the proposed prior compared to Gauss-Total Variation and p-Gaussian priors ($\alpha = 1/\left(3\sqrt{3}\right) \approx 0.1925$)

regions). These two parameters have to be set properly to obtain the desired effects on the final image.

6.3 Optimization of the regularization parameters

In this thesis the tuning of the regularization parameters has been obtained by proposing an index of image quality (which could take into account both qualitative and quantitative information), analysing the images obtained by using different sets of parameters and choosing the couple of parameters which maximized this index.

The index proposed, the *detectability* D, was chosen as the product between two factors, C_{qnt} and C_{qlt}, respectively a quantitative and qualitative figure of merit:

$$D = C_{qnt} C_{qlt}$$

The quantitative factor was chosen as the hot contrast recovery coefficient (CR)

$$C_{qnt} = CR = 100\,\frac{\mu_S/\mu_B - 1}{R - 1}$$

where R is the real signal-to-background ratio and μ_S (μ_B) is the mean content of the voxels inside a spherical VOI drawn in the signal (background) region. The qualitative factor, instead, models the discrimination between signal and background regions (with mean values of activity concentration respectively of μ_S and μ_B) characterised by standard deviations of counts (chosen as a representation of "noise" perceived by the human operator) σ_S and σ_B, respectively. It assumes a logarithmic response of the human eye [35, 106]:

$$C_{qlt} = \ln \frac{\mu_S - \mu_B}{\sigma_S + \sigma_B}$$

The complete expression for the detectability is therefore

$$D = 100\,\frac{\mu_S/\mu_B - 1}{R - 1} \ln \frac{\mu_S - \mu_B}{\sigma_S + \sigma_B}$$

Larger values of D should correspond to better definition and visibility of the lesions and, thus, higher probability of detection by a human operator.

A simulation, using synthetic images, has been employed to test the proposed index and evaluate its capability to predict the outcome of image analysis by human operators.

Each synthetic image was initially created by choosing a uniform activity concentration in the background (13 kBq/cm^3) and a simulated acquisition time (2 min). The image was represented by a matrix 256 pixels $\times$ 256 pixels $\times$ 47 slices on a FOV of 600 mm (transaxial) and 154 mm (axial), corresponding to a voxel size of 2.34 mm $\times$ 2.34 mm $\times$ 3.27 mm. Then, a random number (uniformly distributed between 0 and 15) of spheres (simulating oncological lesions) was inserted in the image, at random locations (with the only constraint of reciprocal non-superposition) and each characterised by a random signal-to-background ratio (SBR), uniformly distributed between 2 and 10. All the radii were chosen equal to 5 mm, since this is the dimension of the real sphere used in the following optimization of the parameters (see sections 7.4.1 and 8.4.1).

For each voxel, then, the product of proper (background or signal) activity concentration, acquisition time and voxel volume was calculated, obtaining the total number of counts in the voxel, respectively M_b and M_s.

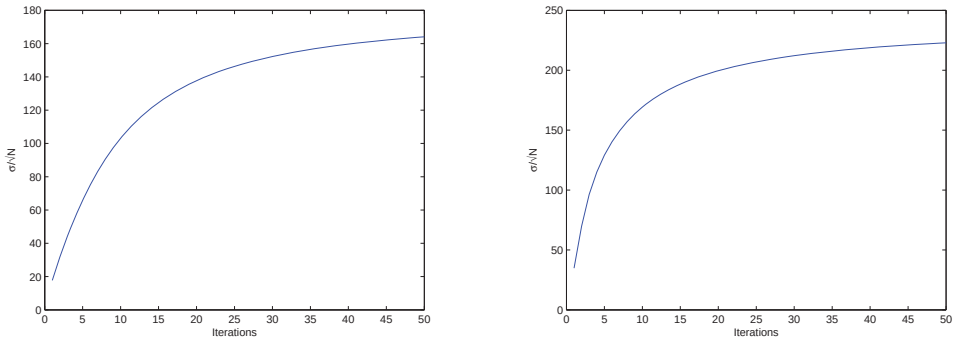
A pseudo-Poisson noise was also added to the entire image. To do so, the standard deviation of counts (for background and signal regions) in a real image of a NEMA IEC Body Phantom[‡] (acquired statistics: 52.4 Mcounts) was analysed, in order to evaluate the deviation from a standard Poisson noise. For each voxel, containing N_c counts, inside a background region (with activity concentration of 13 kBq/cm^3) the standard deviation of counts STD_c was calculated. The *Poisson coefficient*

$$P = \frac{STD_c}{\sqrt{N_c}}$$

was then calculated for each voxel and, finally, the obtained values were averaged to obtain the mean Poisson coefficient $\bar{P}$. This procedure was repeated also for the voxels of a sphere of diameter 10 mm contained in the NEMA Phantom (with SBR equal to 4.4). In figure 6.3 the mean Poisson coefficient for the background and the signal regions is plotted versus the number of iterations.

Choosing 10 iterations as a representative number of iterations for a PET reconstruction with PSF, the $\bar{P}$ coefficient in signal (i.e. a VOI of diameter 10 mm corresponding to the smallest sphere in the NEMA phantom) and background regions was found to be about 175 and 100, respectively. In order to test the detectability index also considering possible additional effects which could worsen the quality of the image, it was decided to introduce a "margin of security" and

[‡]For additional details on this phantom, see section 7.3.1.



(a) Background (b) 10 mm sphere

Figure 6.3: Deviation from pure Poisson noise in background regions and inside a sphere with diameter 10 mm as a function of the iteration number

set the standard deviation S_s (S_b) of counts in signal (background) regions to twice the value reported above:

$$S_s = 350\sqrt{M_s} \qquad S_b = 200\sqrt{M_b}$$

The signal and background regions were filled with respectively $(M_s/S_s)^2$ and $(M_b/S_b)^2$. A pseudo-Poisson random noise was subsequently applied by numerically solving the inverse transform sampling method applied to the cumulative distribution function

$$cdf_{PseudoP} = \frac{\Gamma(x - 0.5, \mu)}{\Gamma(x - 0.5)}$$

which is a generalization of the Poisson cumulative distribution function

$$cdf_{Poiss} = \frac{\Gamma(\lfloor k + 1 \rfloor, \mu)}{\Gamma(\lfloor k + 1 \rfloor)}$$

where μ is the mean of the pseudo-Poisson distribution, $\lfloor \cdot \rfloor$ is the floor function and

$$\Gamma(s) = \int_0^{+\infty} y^{s-1} e^{-y} dy \qquad \Gamma(s,t) = \int_t^{+\infty} y^{s-1} e^{-y} dy$$

are respectively the Gamma function and the incomplete Gamma function. The resulting image, after setting to zero the voxel with negative values, was rescaled by multiplying signal and background regions with factors respectively S_s^2/M_s and S_b^2/M_b: this procedure results in a good approximation of an image characterised by mean counts in a signal (background) region equal to M_s (M_b) and standard deviation S_s (S_b). Each slice of the three-dimensional image was then filtered with a two-dimensional Gaussian of FWHM=5 mm and a 3-point mean was applied

between the different slices: this filtering process was introduced to obtain a final result which could simulate well the images used by the clinicians to report. In figure 6.4 an example of a simulated image is provided, while in figure 6.5 the comparison between the voxel contents in a real image and in a simulated image is performed: both figures show that the proposed method generates a good approximation of the real reconstructed images.

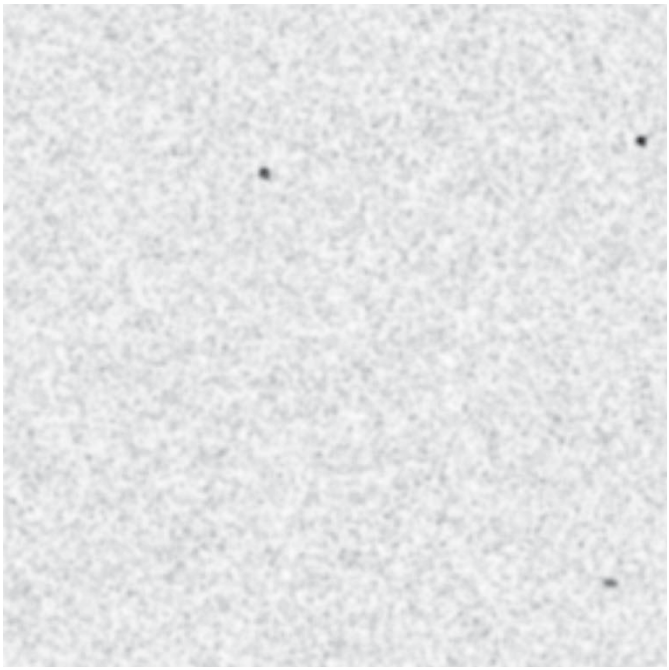


Figure 6.4: Example of simulated PET image

The D value for each sphere in the image was automatically calculated; parallely, a human operator analysed the image without knowing anything about the spheres contained and stated how many lesions, in his/her own opinion, were present. The entire process was repeated for all the 250 images generated.

The images were then divided into two groups: a *training* set and a *test* set. The former set was used to obtain an optimal threshold for D to discriminate between detectable and undetectable lesions, while the latter one was used to evaluate the correlation between the number of lesions detected by the human operator and the number of lesions recognised by using the thresholded D.

Training phase

For each image I in the training set Φ_{train} the spheres $s_i^{(I)}$ are sorted in order of decreasing D value. For a given threshold t, $n_d^{(I)}$ spheres have a D value greater than or equal to t and thus they are considered to be detectable: these spheres

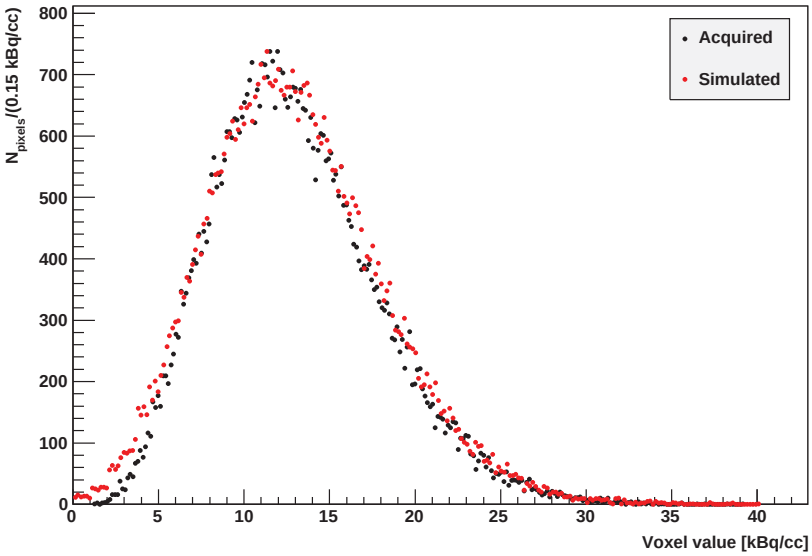


Figure 6.5: Comparison of the voxel content in simulated and real PET image

constitute the set

$$S_t^{(I)} = \left\{ s_i^{(I)} \left| D\left(s_i^{(I)}\right) \geqslant t \right. \right\}$$

$$\left| S_t^{(I)} \right| = n_d$$

Calling n_o the number of spheres recognised in the image by the human operator during the image analysis, a second set $S_o^{(I)}$ is populated with the first n_o (if available) spheres in decreasing order of D value:

$$S_o^{(I)} = \left\{ s_k^{(I)} \left| D\left(s_k^{(I)}\right) \geqslant D\left(s_{k+1}^{(I)}\right) \right. \;,\; k = 1 \ldots \min\left(n_o, n_d\right) \right\}$$

The optimal threshold $\hat{t}$ is then chosen as

$$\hat{t} = \arg\min_t \sum_{I \in \Phi_{train}} \sum_{s_j^{(I)} \in \Psi} \left[D\left(s_j^{(I)}\right) - t \right]^2$$

$$\Psi = S_o^{(I)} \,\Delta\, S_t^{(I)}$$

where Δ indicates the symmetric difference between sets (i.e. Ψ contains the $|n_d - n_o|$ remaining spheres from the larger set between $S_t^{(I)}$ and $S_o^{(I)}$). From the computational point of view, the search for $\hat{t}$ is performed by changing discretely t (in steps of 0.1) and selecting the value minimizing the sum presented above.

Test phase

Each image I in the test set Φ_{test} is used to evaluate the correlation between the number of lesions detected by the human operator ($n_o^{(I)}$) and the number of lesions detected with the thresholded detectability index ($n_d^{(I)}$), i.e. the number of lesions with $D \geqslant \hat{t}$.

By introducing the vectors

$$X = \left\{ n_o^{(I)} | \, I \in \Phi_{test} \right\} \qquad Y = \left\{ n_d^{(I)} | \, I \in \Phi_{test} \right\}$$

the correlation index between X and Y is calculated as

$$\rho = \frac{\mathrm{cov}\,(X, Y)}{\sqrt{\mathrm{var}\,(X)}\sqrt{\mathrm{var}\,(Y)}}$$

where cov and var indicate covariance and variance, respectively.

Whether each image should belong to Φ_{train} or Φ_{test} is an arbitrary decision. To avoid this arbitrariness, 200000 different assignments were performed; for each assignment, 100 images (out of the entire set of 250 images) were randomly selected for Φ_{train}, while the remaining 150 images populated Φ_{test}.

The results of the whole test, presented in figure 6.6a, show that — in conditions worse than the clinical ones — the correlation coefficient between the human and the automatic detection is

$$\rho = (97.6 \pm 0.3)\ \%$$

As an additional validation test, the same procedure described above was performed also by assigning to each image a random number (uniformly distributed between 0 and 15) of spheres detected by the operator and:

- by using, for the detectability of each sphere, the expression presented above;
- by randomly generating, for each sphere, a real number from a uniform distribution between 0 and 60 and using it as the detectability of the sphere.

The results of these additional tests are presented in figures 6.6b and 6.6c. In particular, the correlation coefficients between the human and the automatic detection are, respectively,

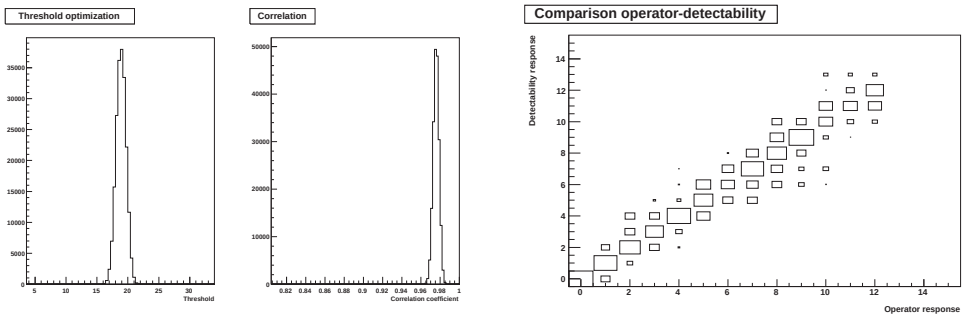
$$\rho = (8.8 \pm 4.5)\ \%$$

and

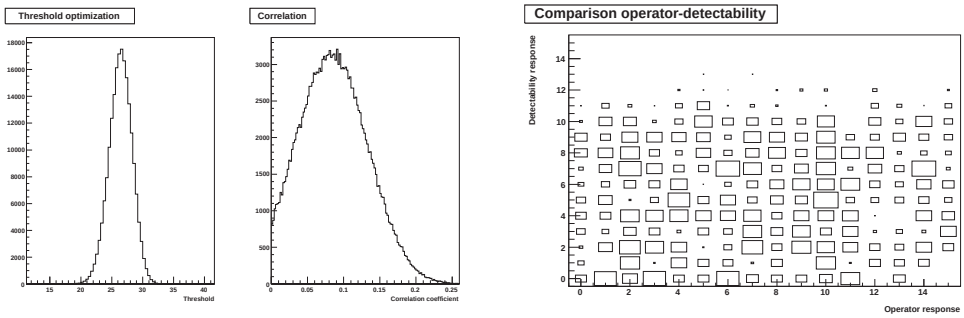
$$\rho = (12.6 \pm 0.5)\ \%$$

Consequently, the proposed index D can mimic quite well the response of the human operator and it may thus be used to optimize the parameters β and

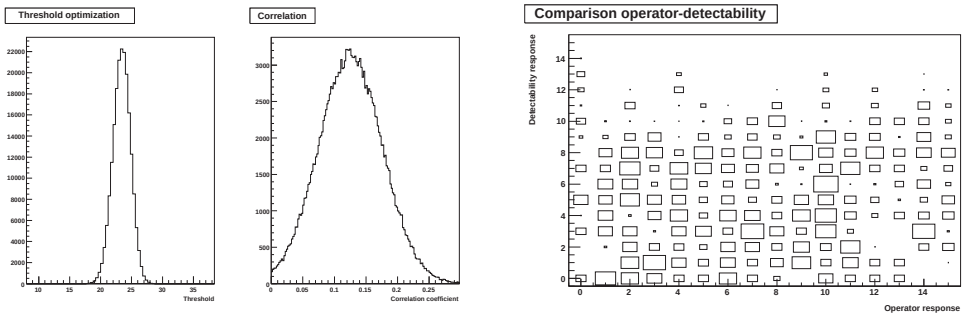
δ. As described in sections 7.4.1 and 8.4.1, the optimization is performed by reconstructing the same phantom using different sets of regularization parameters (β, δ), among which the one maximizing the detectability index D is considered to be optimal.



(a) Detectability index D, real number of lesions detected by the human observer



(b) Detectability index D, random number of lesions detected by the human observer



(c) Random detectability, random number of lesions detected by the human observer

Figure 6.6: Validation procedure for the detectability index: left, histograms of the optimal thresholds (left part of the plot) and of the correlation coefficients (right part of the plot) for the different assignments; right, comparison between the number of lesions recognized by the detectability index and by the human operator (larger boxes represent higher frequencies)

6.4 Implementation of the variational regularization

To implement the variational regularization inside the OSEM algorithm accounting for the PSF, the starting point is equation 6.7:

$$\lambda_b^{[k+1]} = \frac{\lambda_b^{[k]}}{\mathrm{B}_b A_d - \beta \mathrm{D}_b\left(\lambda^{[k]}\right)} \mathrm{B}_b \frac{A_d y_d}{A_d \mathrm{P}_d \lambda_{b'}^{[k]} + R_d + S_d}$$

In principle, it is possible to implement the PSF in the above equation by redefining the projector and backprojector as described in equation 5.3. In this case, however, the β parameter does not directly control the strength of the prior, since the regularization term $\beta \mathrm{D}_b$ is compared to the (modified) backprojection of the attenuation sinogram, which obviously depends on the specific data to be reconstructed. Consequently, it would be very difficult to set a *global* reference scale for β.

For this reason, a different approach has been followed: the (modified) backprojection of the attenuation was "factorised" by implementing the regularization using the multiplicative approach described in [68]:

$$\lambda_b^{[k+1]} = \frac{\lambda_b^{[k]}}{(\mathrm{B}_b A_d) \cdot \left[1 - \beta \mathrm{D}\left(\lambda^{[k]} \cdot \mathrm{B} A_d\right)\right]_b} \mathrm{B}_b \frac{A_d y_d}{A_d \mathrm{P}_d \lambda_{b'}^{[k]} + R_d + S_d} \tag{6.12}$$

With this strategy the β parameter has a natural reference scale — since the regularization term is always compared with the unitary vector 1, regardless of the attenuation factor —, which justifies a generalization of the optimized parameters to the patient studies and a comparison of the results obtained with phantom and clinical data.

Part III

Results

Chapter 7

Results — Non TOF imaging

7.1 Description of the GE Discovery STE scanner

The Discovery STE (DSTE) is an integrated system which combines a multi–slice CT scanner and a BGO PET tomograph [95].

The CT system is the "LightSpeed" model with technology at 16 multi–slices. The detector matrix is characterized by 24 parallel rows of small solid state detectors (16 rows with thickness of 0.625 mm each and 8 rows with thickness of 0.25 mm each) for a total of 21888 elements. The "LightSpeed" detector configuration allows full 360 degree rotational scans with a variable period from 0.5 s to 1.0 s, with steps of 0.1 s (allowing a maximum of 16 slices per rotation and 32 images per second) with the possibility of retrospectively reconstructing the acquired data with different slice thickness.

The PET scanner is a multi–ring design system. The detector system consists of 13440 BGO detectors with dimensions of $4.7 \times 6.3 \times 30$ mm^3. The basic PET detection unit is a block of 48 (8×6) individual BGO crystals read–out by a single squared photomultiplier tube (PMT) with 4 anodes for each block. The connection of the PMTs to the crystal detectors is optical and directed. The DSTE uses a new electronic system that bidirectionally communicates with the PMTs, correcting single temporal events for each crystal rather than for the whole block. Considering the axial direction, four blocks of six crystals are used: the 24 crystal ring detectors originate 47 slices, which correspond — on the axial FOV of 15.7 cm — to an axial sampling of 3.27 mm. Each ring (diameter is 88.6 cm) is composed of 70 blocks; the transaxial FOV diameter is 70 cm. The DSTE has the capability of both 2D (using tungsten septa with length of 5.4 cm and thickness of 0.8 mm) and 3D (without septa) configuration. In 2D, the axial combination is 5 and 6 planes for direct and cross planes respectively, while in 3D the axial combination for coincidence acceptance is up to ± 23 planes. Moreover, in the

2D configuration the DSTE uses a coincidence time window of 11.7 ns and a low energy threshold of 375 keV, while in the 3D configuration the coincidence time window is reduced to 9.7 ns and the low energy threshold is raised to 425 keV. Finally, the PET system has a ^{68}Ge rod source with activity of 55.5 MBq, used for the system calibration and daily quality controls.

7.2 PSF measurements

The PSF of the DSTE was experimentally determined as described in chapter 4. The transaxial FOV was sampled at distances (from the FOV centre) ranging from 2.5 cm to 25 cm in steps of about 2.5 cm, both to the right and to the left of the centre. For each radial distance, the entire axial FOV was sampled in steps of about 1.5 cm.

An example of a fit on an acquired image is reported in figure 7.1.

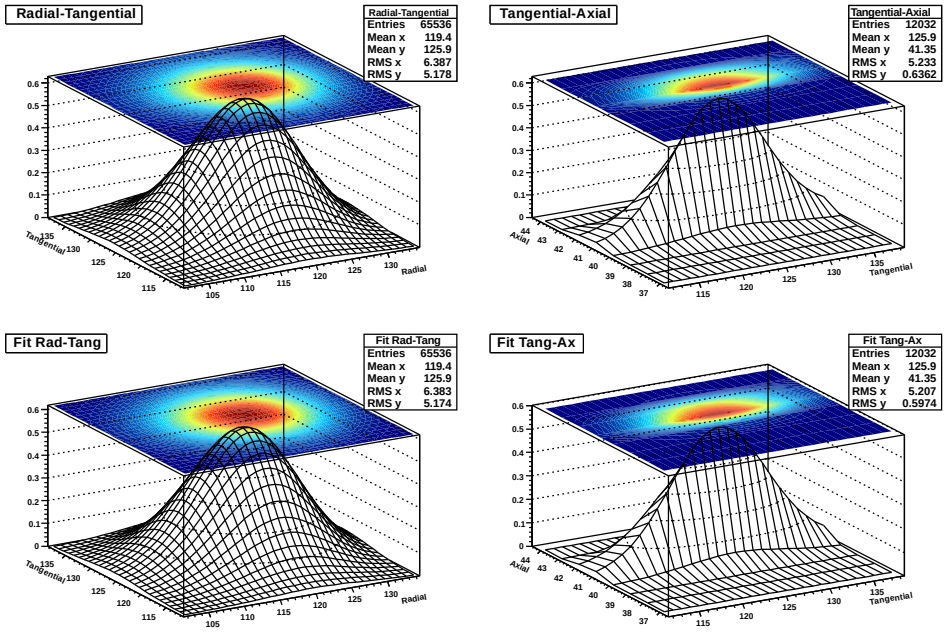


Figure 7.1: Example of PSF fits on DSTE. Top: radial–tangential and tangential–axial images; bottom: corresponding fits

All the obtained values for each spread parameter were then graphed as a function of both radial and axial distances from the scanner FOV centre (figure 7.2): it was then possible to fit each set of data with a polynomial quadratic in both distances, symmetric with respect to the centre of axial FOV, as defined in equation 4.8.

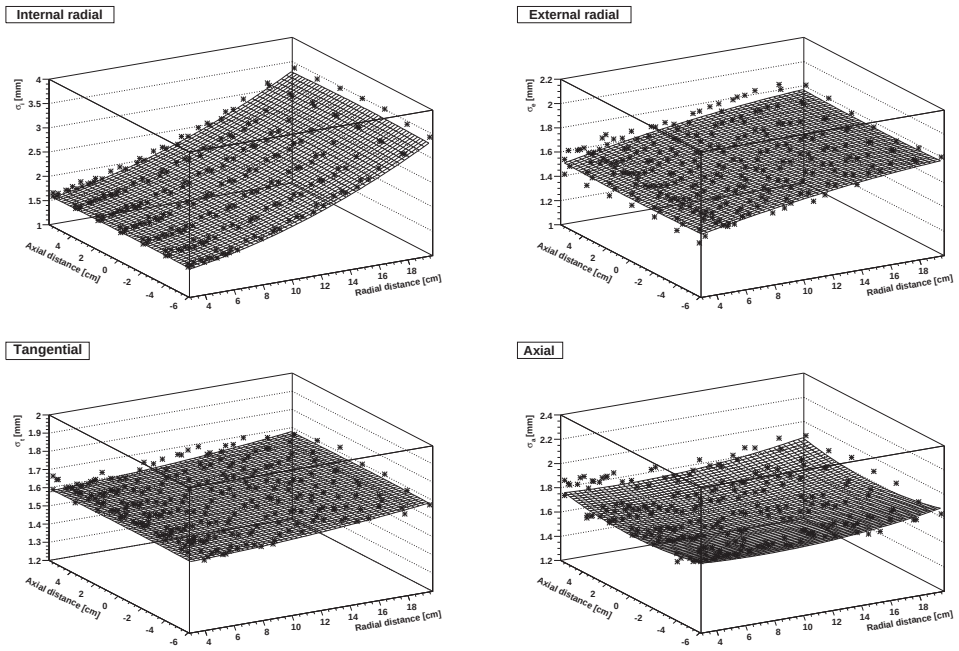


Figure 7.2: Determination of the spread parameters dependences on the radial and axial distances in Discovery STE: internal radial (top left), external radial (top right), tangential (bottom left) and axial (bottom right)

The relative weight of the square term in the axial dependence of the transaxial spread parameters was found to be about one order of magnitude lower than the square term in the axial dependence of the axial spread parameter: therefore, the axial dependence of the transaxial spread parameters was assumed to be negligible, in order to reduce the complexity of the model.

The dependences used in the work were then

$$\begin{aligned} \sigma_i &= 1.443 + 0.05572\, r + 2.437 \times 10^{-3}\, r^2 \\ \sigma_e &= 1.473 + 0.01114\, r + 1.318 \times 10^{-4}\, r^2 \\ \sigma_t &= 1.499 - 1.233 \times 10^{-3}\, r + 2.908 \times 10^{-4}\, r^2 \\ \sigma_a &= \left(1.521 - 0.03099\, r + 2.011 \times 10^{-3}\, r^2\right)\left(1 + 0.01664\, a^2\right) \end{aligned} \tag{7.1}$$

where the radial (r) and axial (a) distances are measured in centimeters and the spread parameters are measured in millimeters.

In figure 7.3 the radial, tangential and axial profiles drawn on an image are compared with the corresponding ones drawn on the resulting fit function and on the final kernel function* used in the reconstruction algorithm. The figure reveals

*To compare the three profiles, the values of σ_X^f (the spread parameters evaluated from the

the very good resemblance between the different profiles and, consequently, the reliability of the fitting results.

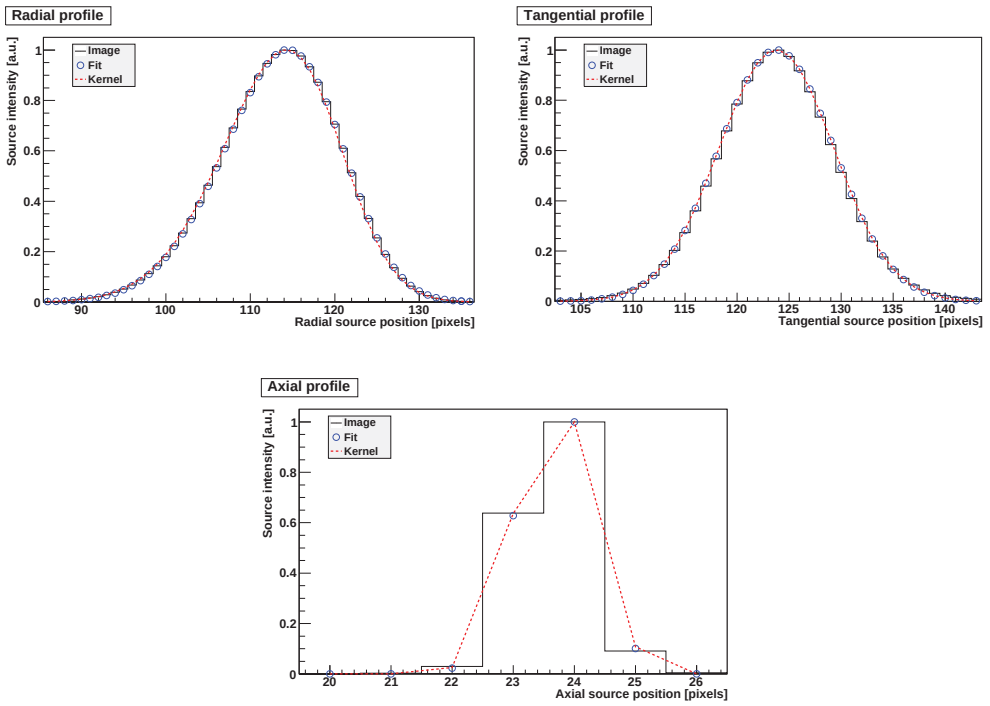


Figure 7.3: Comparison between the radial (top left), tangential (top right) and axial (bottom) profiles drawn on an image, on the resulting fit and on the corresponding kernel used in the reconstruction

7.3 Improvements thanks to PSF

The new algorithm accounting for the PSF of the DSTE scanner was validated on phantom and clinical studies. For both types of data, two sets of reconstructions were performed using the 3D OSEM algorithm without (RnoPSF) and with the PSF (RwPSF). A further reconstruction was performed without the use of the PSF but with a post-filter applied on the final reconstructed images (RnoPSF-Filt). The validation consisted in comparing the images obtained with the different algorithms. All the validations were qualitative; with phantom data, also quantitative comparison were performed.

fit on the specific image) and σ_X^k (the spread parameters evaluated from the global fit on the entire set of data) were substituted in equations 4.3, 4.4 and 4.6 and the resulting functions (respectively, the fitted functions on the specific image and the fitting functions corresponding to the final model) were used to extract the required profiles.

Moreover, to evaluate the efficacy of the variant PSF model, a second implementation of the reconstruction algorithm was obtained by using a transaxially-invariant PSF. Two single invariant PSFs were considered to be representative of the response of the PET system, evaluated at 0 cm (PSF_0) and 15 cm (PSF_{15}) from the transaxial FOV centre, in the center of the axial FOV. The corresponding algorithms will be referred to as $RwPSF_0$ and $RwPSF_{15}$, respectively.

7.3.1 Quantitative validation

^{22}Na sources

Six coplanar ^{22}Na non–collimated point sources (whose activity at the time of acquisition was 5.22 μCi) were positioned, on a polystyrene support, at different radial distances (ranging from 3 cm to 18 cm) from the centre of the scanner transaxial FOV to test the spatial resolution recovery introduced by PSF compensation. A CT scan was performed to be used for attenuation correction followed by a 3D emission scan recording 339.6 Mcounts.

The reconstruction parameters common to all the algorithms were: image matrix 256 pixels × 256 pixels, 28 subsets, 5 iterations, reconstruction FOV 25.6 cm, pixel size 1 mm, reconstruction x offset 10 cm. The post filter applied in RnoPSF-Filt was a symmetric two-dimensional Gaussian filter with FWHM=2 mm in the transaxial planes.

A one-dimensional profile of the activity distribution was generated from the reconstructed images by averaging along the tangential direction.

In figure 7.4 the RnoPSF, RnoPSF-Filt and RwPSF images at 5 iterations are compared, while in figue 7.5 the profiles obtained from these images are shown. The RnoPSF images show asymmetric profiles, in particular when the distance from the scanner centre increases, with a consequent decrease of the peak heights and very irregular peaks for the different point sources. The application of a 2 mm FHWM Gaussian filter (RnoPSF-Filt) mitigates this noisy pattern. The PSF in the reconstruction clearly contributes to recovering the symmetry of the profiles, to obtaining a more uniform resolution across the reconstruction FOV and to recovering similar peak heights across the FOV. To support this latter statement, the coefficient of variation (defined as the ratio between the standard deviation and the mean value) of the peak heights was calculated. The obtained values for this figure of merit were 0.180, 0.133 and 0.061 for RnoPSF, RnoPSF-Filt and RwPSF, respectively. It is important to note that, even if with a higher number of iterations the results here presented remained stable, 5 iterations were considered a limit for acceptability of noise in RnoPSF and thus chosen for this validation.

In figure 7.6 the results obtained by using the invariant PSFs are shown and compared with those obtained by using RwPSF. When the invariant PSF (e.g.

PSF_0) underestimates the real PSF in a specific point of the scanner FOV (e.g. in the sources farthest from the centre), the recovery of the corresponding point sources is incomplete both in terms of shape and peak heights, while if the invariant PSF overestimates the real PSF (e.g. PSF_{15} used at 3-6 cm from the scanner centre), also the recovery of spatial resolution is overstressed, leading to spatially variant peak heights, decreasing with increasing distance from the scanner centre. It is also important to note that the profiles obtained with the invariant and the variant kernels are very similar near locations where the invariant PSF was derived (respectively, 0 cm for PSF_0 and 15 cm for PSF_{15}), as expected.

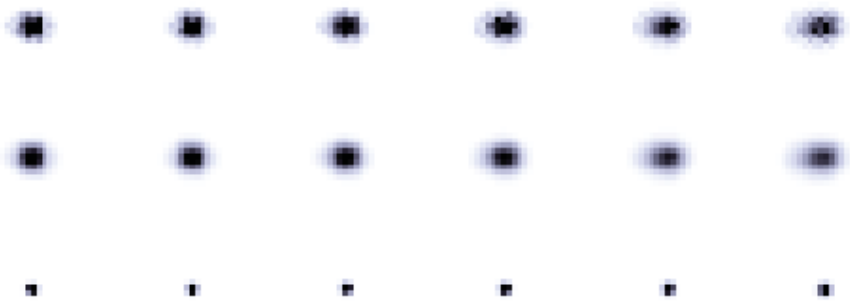


Figure 7.4: ^{22}Na sources at 5 iterations: RnoPSF (top), RnoPSF filtered with a FWHM=2 mm Gaussian (middle), RwPSF (bottom) [the images are shown using the same display parameters]

NEMA IEC Body Phantom

The NEMA IEC Body Phantom Set^{TM}[26] consists of a body phantom (180 mm long), a cylindrical insert (whose outside diameter is 51 mm) simulating lung and an insert with six fillable spheres (whose inner diameters are 10 mm, 13 mm, 17 mm, 22 mm, 28 mm and 37 mm). The spheres have their centres coplanar and all at a distance of 5 cm from the phantom centre. Figure 7.7 presents a photograph of this phantom.

This phantom was used to simulate an oncologic whole body ^{18}F-FDG study. Both the body and the spheres were filled with a ^{18}F-FDG solution in order to have a LBR equal to 4.4. After centring the phantom in the scanner FOV, a CT scan was performed to be used for attenuation correction followed by a 3D emission scan of 2 min recording 52.4 Mcounts.

The reconstruction parameters common to all the algorithms were: image matrix 256×256, 28 subsets, up to 50 iterations, reconstruction FOV 60 cm, pixel size 2.34 mm. The post filter applied in RnoPSF-Filt was a symmetric

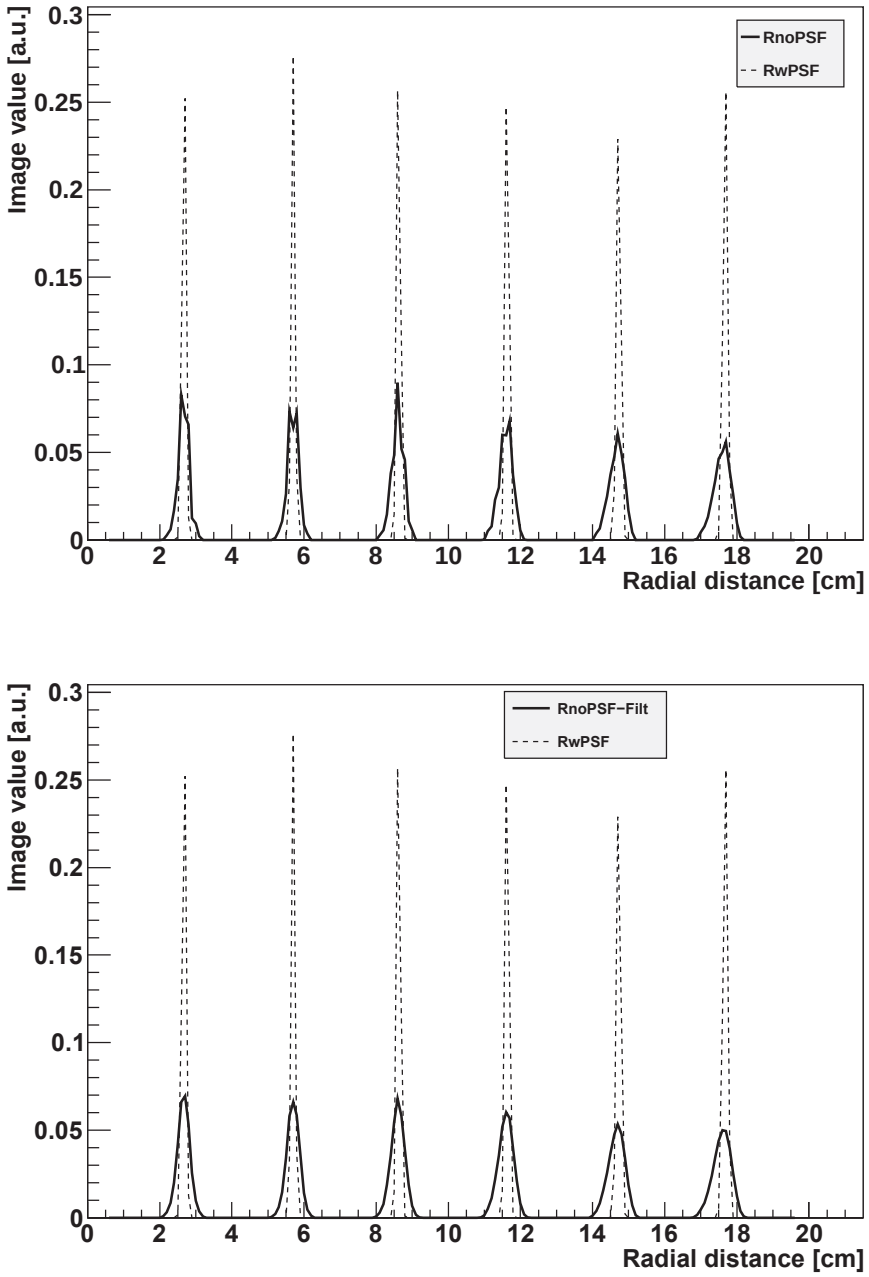


Figure 7.5: Comparison between RnoPSF and RwPSF (top) and RnoPSF-Filt (FWHM 2 mm) and RwPSF (bottom) with ^{22}Na sources

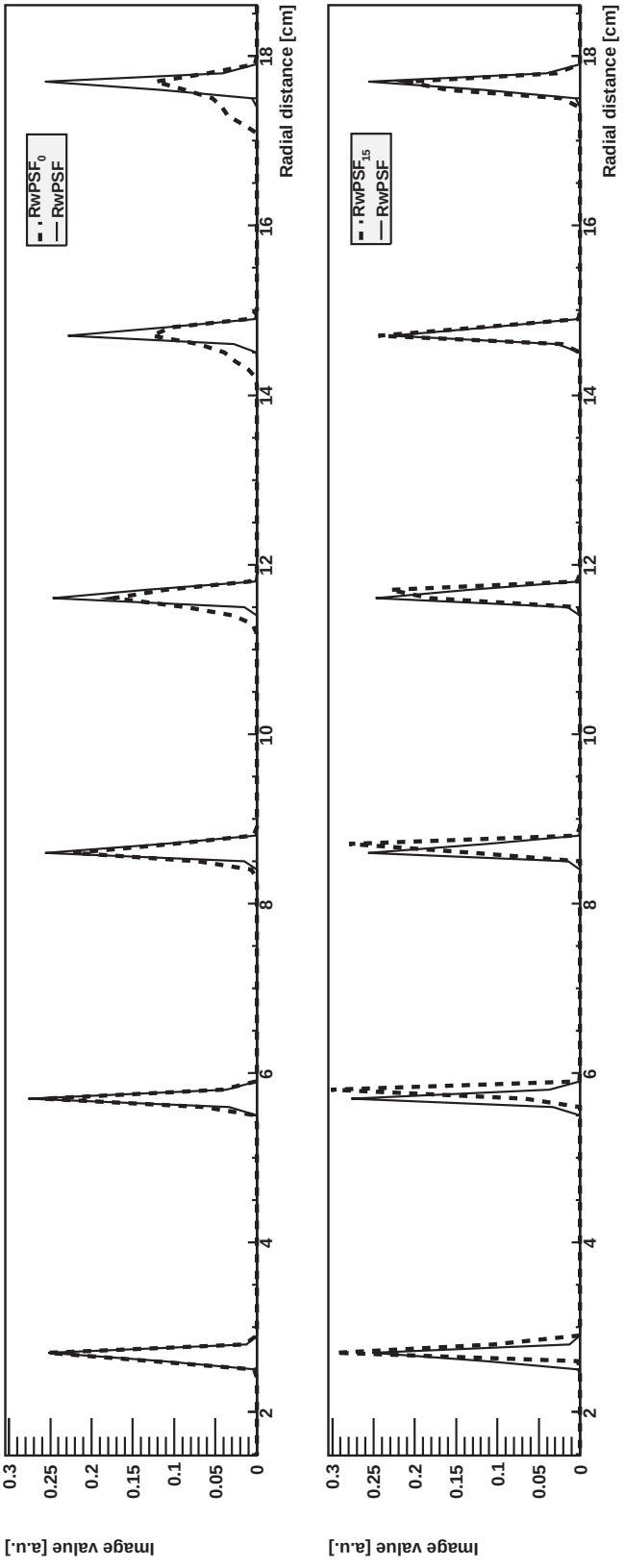


Figure 7.6: Comparison between RwPSF and RwPSF_0 (top) and RwPSF and RwPSF_{15} (bottom) with ^{22}Na sources

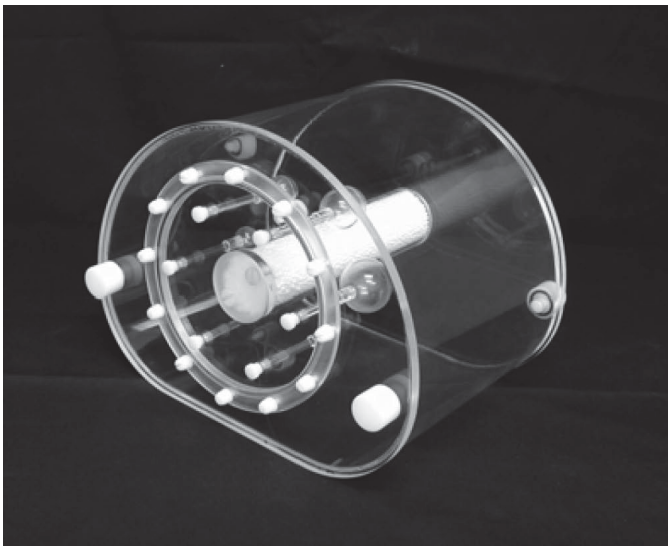


Figure 7.7: Photograph of the NEMA IEC Body Phantom Set™[26]

two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a weighted 3-point mean along the axial direction.

The quantitative analysis was performed by using two parameters.

Background coefficient of variation

A uniform region in the object under study should be reconstructed as uniform: the MLEM algorithm is able to overcome the statistical fluctuations, since the Poissonian nature of emission is incorporated inside the algorithm itself. In the real case, though, a uniform region contains more than pure Poissonian events (e.g. an imperfect correction of physical effects or round–off errors in the computational representation) and, as a consequence, some residual noise will be present. Unfortunately, iterative algorithms amplify this noise iteration by iteration. This amplification is measurable for example with the *standard deviation* of noise: if n_i is the content of voxel i inside the region R (composed of N voxels) and μ is the mean value of contents in region R, the standard deviation is defined by

$$STD = \sqrt{\frac{1}{N-1}\sum_{\substack{i=1 \\ i\in R}}^{N}(n_i-\mu)^2} \tag{7.2}$$

where the Bessel's correction has been applied to the ideal definition. The coef-

ficient of variation (COV) is then defined as

$$COV = \frac{STD}{\mu} = \frac{1}{\mu}\sqrt{\frac{1}{N-1}\sum_{\substack{i=1 \\ i\in R}}^{N}(n_i - \mu)^2} \tag{7.3}$$

Hot contrast recovery

If the reconstruction of a PET scan were quantitatively correct, the ratio of the activities in a *signal* region and in a *background* region — the *contrast* of the signal with respect to the background — calculated on the reconstructed image would resemble the real one. Due to PVE and PSF, instead, the measured contrast is always lower. The *hot contrast recovery coefficient* CR_{hot} is the percentage of contrast, with respect to the original contrast, measured on the reconstructed image:

$$CR_{hot} = 100\,\frac{\mu_S/\mu_B - 1}{R - 1} \tag{7.4}$$

where R is the real signal-to-background ratio and μ_S (μ_B) is the mean content of the voxels inside a spherical VOI drawn in the signal (background) region.

Ten spherical VOIs were drawn, four in background regions and six for the signal (i.e. the six spheres). The signal spheres were centred on CT and the radii were chosen according to the specifications from the manufacturer of the spheres.

The background mean value μ_B and the standard deviation STD were calculated over all voxels in the first set of VOIs, while the second set of VOIs was used to calculate the contrast recovery coefficients for the spheres.

As shown in figure 7.8, the PSF contributes to an increase of the CR_{hot} — after 50 complete iterations — ranging from 8.4% (for the largest sphere, diameter 37 mm) to 59.7% (for the smallest sphere, diameter 10 mm). At a more "clinical" number of iterations (e.g. 5-10 iterations), the CR_{hot} coefficients become respectively 4.8%–6.0% (largest sphere) and 15.1%–35.3% (smallest sphere) higher than the corresponding ones from RnoPSF. Furthermore, the presence of PSF in the reconstruction algorithm reduces the background COV by 53.2% at 5 iterations, 50.9% at 10 iterations and 12.5% at 50 iterations. In figure 7.9 the reconstructed images of the NEMA IEC Body Phantom at 5 iterations are shown. The images visually confirm the quantitative results for higher spatial resolution, higher contrast and lower noise in RwPSF compared to RnoPSF.

If RnoPSF-Filt is compared with RwPSF the spatial resolution and contrast in the images are higher for the latter, while the background COVs in RnoPSF-Filt become increasingly lower (-39.9% at 5 iterations, -53.4% at 10 iterations and -80.6% at 50 iterations) with respect to the corresponding results obtained using PSF; thus, RwPSF showed increased spatial resolution with respect to RnoPSF images while maintaining an acceptable level of background variations if compared with the poor resolution images obtained with RnoPSF-Filt.

Also for the NEMA IEC Body Phantom a comparison between the results obtained with invariant and variant kernels was performed. To this purpose, up to 200 iterations were used to further investigate the asymptotic behaviour of the considered algorithms. The results obtained are presented in figure 7.10. The use of an invariant PSF confirms the results obtained with the point sources. In fact, since the centres of the spheres in the NEMA IEC phantom are at a radial distance of 5 cm from the centre of the phantom (which was coincident with the centre of the PET FOV), the RwPSF_0 underestimated the recovery over all the spheres in the phantom when compared with the results obtained with the RwPSF, leading to lower values of the CR_{hot}. Instead, the RwPSF_{15} (which overestimates the response of the system in the region of the spheres) provided the highest contrast recovery, but also generated unphysical results: as can be seen in figure 7.10, the CR coefficient curve for the largest sphere crosses 100 after about 200 iterations, suggesting that the images have been overcorrected.

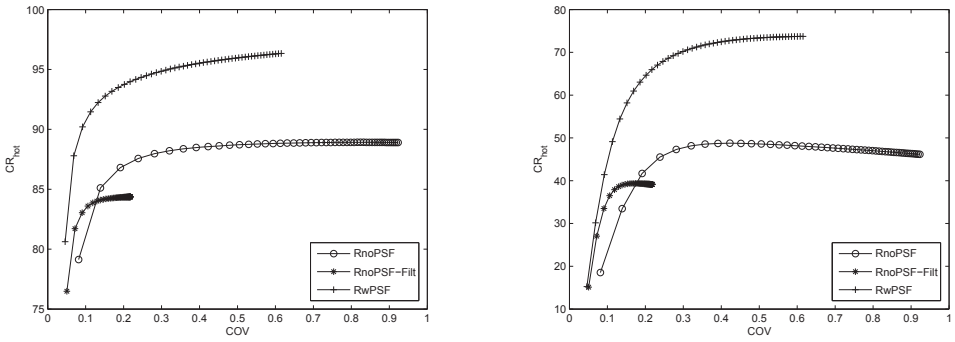


Figure 7.8: NEMA IEC phantom: comparison between the contrast recovery coefficients (CR_{hot}) for the 37-mm diameter sphere (left) and for the 10-mm diameter sphere (right) versus the background coefficient of variation [each point represents one iteration] for RnoPSF, RnoPSF-Filt and RwPSF

Figure 7.9: NEMA IEC phantom at 5 iterations: RnoPSF (top left), RnoPSF filtered with a FWHM=5 mm Gaussian (top right), RwPSF (bottom) [the images are shown using the same display parameters]

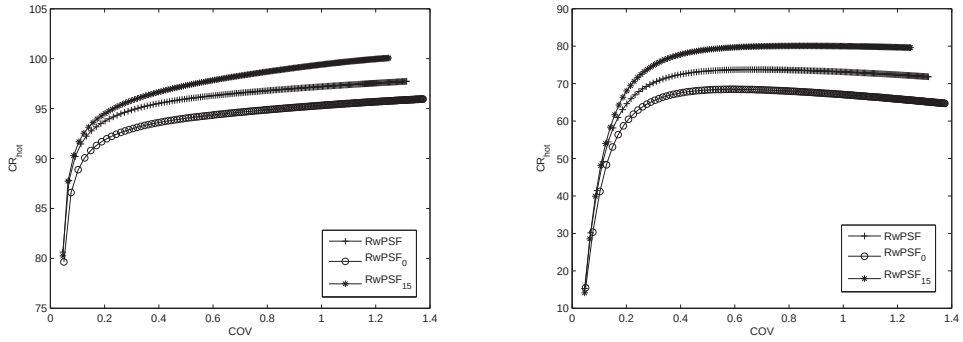


Figure 7.10: NEMA IEC phantom: comparison between the contrast recovery coefficients (CR_{hot}) for the 37-mm diameter sphere (left) and for the 10-mm diameter sphere (right) versus the background coefficient of variation [each point represents one iteration] for RwPSF, RwPSF_0 and RwPSF_{15}

Cylindrical phantom

A cylindrical phantom (with diameter and height 20 cm) containing a sphere (inner diameter: 8 mm) was used to assess the dependence of the convergence rate and the recovery of spatial resolution on the lesion-to-background ratio (LBR). For these reasons, neither RnoPSF-Filt nor the invariant kernels were considered for this phantom.

Both the cylinder and the sphere were filled with a ^{18}F-FDG solution in water to obtain a known ratio between the activity concentrations in the sphere and in the background: the considered ratios were 3.8 and 20, to simulate lesions respectively in the liver and in the lungs.

After positioning the phantom in the scanner FOV to have the sphere axially centred at a radial distance of 15 cm, a CT scan was performed to be used for attenuation correction followed by a 3D emission scan of 15 min recording respectively 348.7 and 457.6 Mcounts.

The reconstruction parameters common to all the algorithms were: image matrix 256×256, 28 subsets, up to 100 iterations, reconstruction FOV 30 cm, pixel size 1.17 mm, reconstruction offset along the vertical (y) direction: 10 cm.

As shown in figure 7.11, the PSF increases the CR_{hot} coefficients for both the considered LBRs, confirming the results presented in figure 7.8. Moreover, the asymptotic distance between the two curves is reduced by the use of PSF (at 100 iterations, 5.1% for RnoPSF and 2.27% for RwPSF), showing the increase in quantitative accuracy and the reduction of its dependence on the LBR. On the other hand, the higher LBR (representative of the lung district) leads to a faster convergence, while for the lower LBR (representative of the liver district) more

iterations with PSF are required to obtain the same recovery. It is important to underline, though, that the sphere was located about 15 cm far from the scanner centre and, consequently, the real convergence speed in the clinical practice may be higher, since most lesions are located within about 10 cm from the scanner centre (e.g. brain and lung).

In figure 7.12 the reconstructed images (with RnoPSF and RwPSF) of the cylindrical phantom at 10 iterations are shown. The images visually confirm the quantitative results in terms of higher contrast and lower noise in the RwPSF compared to the RnoPSF.

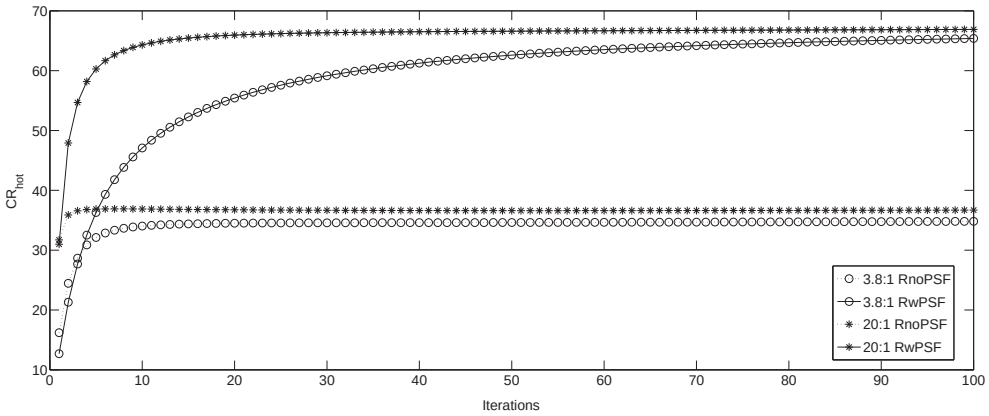


Figure 7.11: Cylindrical phantom: comparison between the contrast recovery coefficients (CR_{hot}) obtained with RnoPSF and RwPSF as a function of the number of iterations

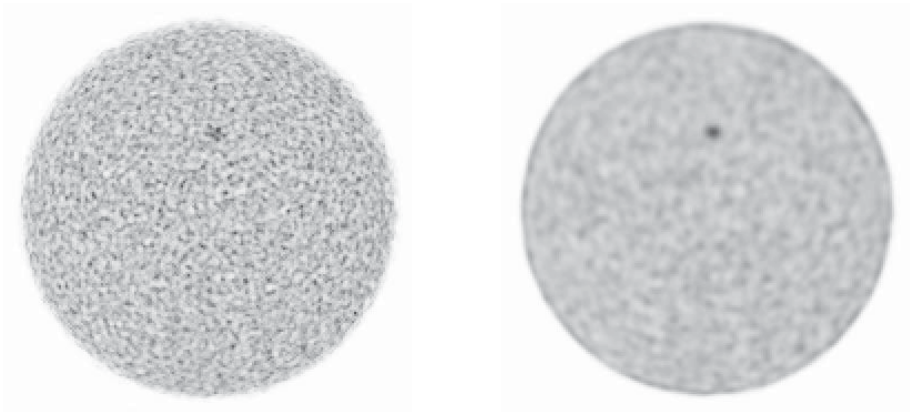


Figure 7.12: Cylindrical phantom: comparison between the RnoPSF (left) and RwPSF (right) images at 10 iterations for the 3.8:1 LBR [the images are shown using the same display parameters]

7.3.2 Qualitative validation

Oncological patient

An oncological patient received an injection of ^{18}F-FDG (370 MBq); the tracer uptake time was 60 min. The acquisition protocol consisted in a whole body CT scan to be used for anatomical localization and attenuation correction and a 3D whole body emission scan (2.5 min per bed position, 6 bed positions, 7-slice overlap between bed positions) recording 289.3 Mcounts.

The reconstruction parameters common to all the algorithms were: image matrix 256×256, 28 subsets, up to 5 iterations, reconstruction FOV 60 cm, pixel size 2.34 mm. The applied post–filter was characterized by the same parameters as in the clinical practice, i.e. symmetric two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a weighted 3-point mean along the axial direction.

To assess the improvements introduced by PSF, a qualitative comparison of the three different sets of images (RnoPSF, RnoPSF-Filt and RwPSF) was performed by an expert oncologist on the images obtained after 2, 3 and 5 iterations.

In figure 7.13 the resulting coronal images are conveyed as a Maximum Intensity Projection (MIP) representation, while in figure 7.14 an example of transaxial images is provided. The "pure" RnoPSF images show an unacceptable level of noise for clinical practice, even after few iterations; consequently, the post–filter is mandatory to allow a clinical evaluation. The RwPSF images, compared to both the RnoPSF and RnoPSF-Filt images, are characterized by higher contrast of the lesions. As far as resolution is concerned, at 2 iterations the RnoPSF and RwPSF images appear similar, probably with a slightly lower resolution for RwPSF. After 3 iterations the RwPSF starts to show higher spatial resolution with respect to both the RnoPSF reconstructions: this is a confirmation of the results obtained from the NEMA IEC phantom (e.g. see figures 7.8 and 7.9), in which 3 iterations appeared to be a "crossing point" for resolution recovery, particularly in the smallest spheres (e.g. for the 10-mm sphere, the CR_{hot} from RwPSF becomes higher than the one from RnoPSF starting from the third iteration). The "visual" noise appears progressively lowest in the RnoPSF-Filt, even if this affects the contrast and the definition of the lesions.

The dark bands visible in the whole-body MIP RnoPSF images originate in the spatial regions of overlap between adjacent bed positions when more than one PET axial FOV is acquired (like in whole-body scans). In fact, as the profile of the axial sensitivity of a scanner operating in 3D mode is triangularly-shaped and peaked in the middle point of the axial FOV, at its axial borders the efficiency of the scanner is lower than in the centre: consequently, the corresponding images are noisier and hence the MIP — since it substitutes each pixel with the maximum value along the projection line — tends to emphasise these noisier re-

gions. Usually 9- or 11-slice overlap is used for whole body applications; in the case presented in figure 7.13 only a 7-slice overlap was used. The bands are less visible in the RnoPSF-Filt and in the RwPSF images, thanks particularly to the effects along z given by the weighted mean and the PSF, respectively.

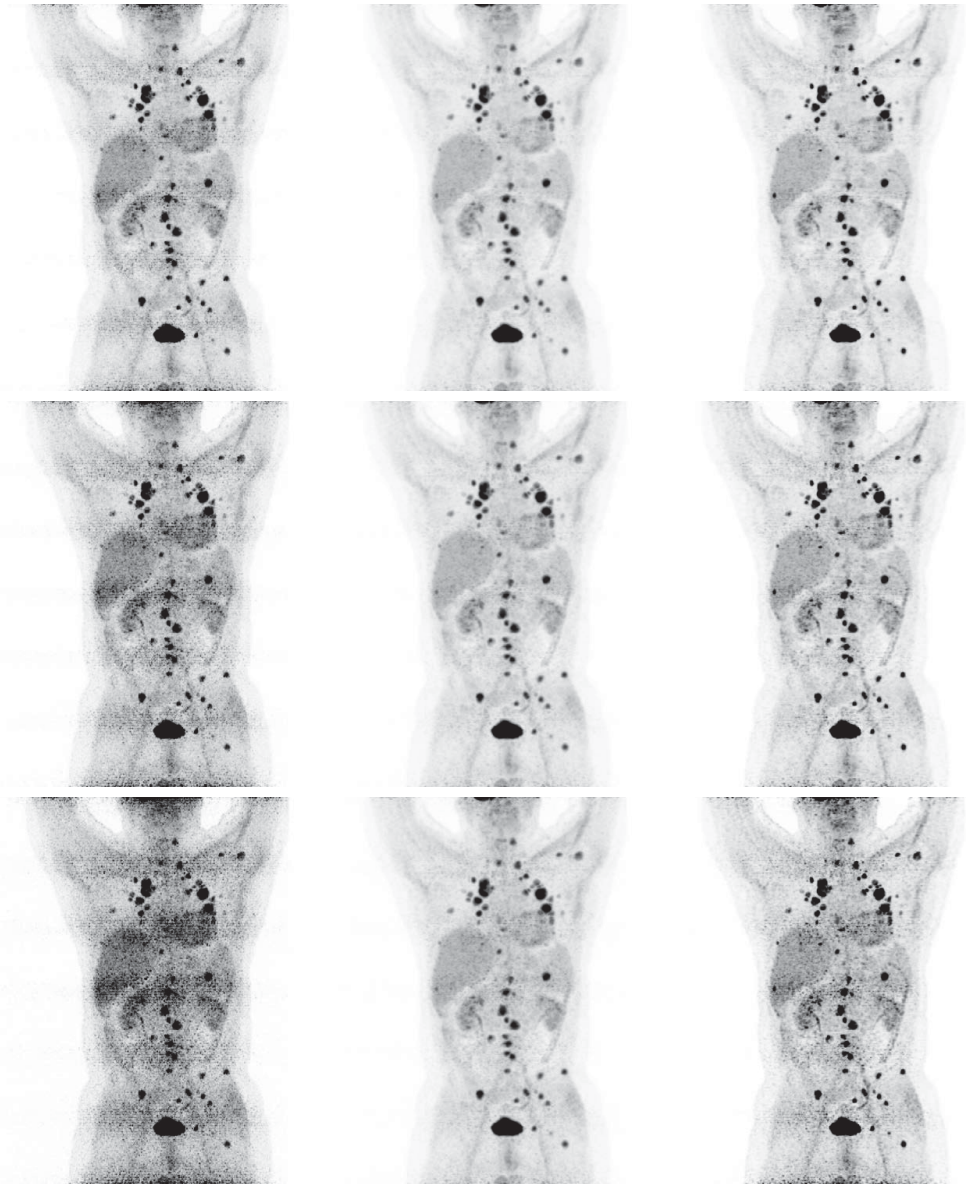


Figure 7.13: Coronal MIP images of the oncology patient reconstructed using 2 (top), 3 (middle) and 5 (bottom) iterations: RnoPSF (left), RnoPSF filtered with a FWHM=5 mm Gaussian (centre), RwPSF (right) [the images are shown using the same display parameters]

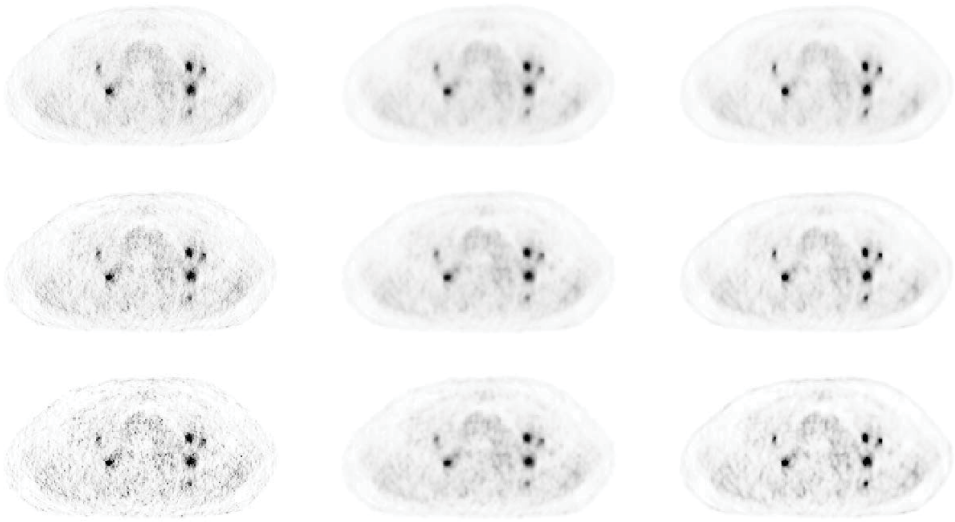


Figure 7.14: Transaxial images of the oncology patient reconstructed using 2 (top), 3 (middle) and 5 (bottom) iterations: RnoPSF (left), RnoPSF filtered with a FWHM=5 mm Gaussian (centre), RwPSF (right) [the images are shown using the same display parameters]

Neurological patient

A neurological patient received an injection of ^{18}F-FDG (193 MBq); the tracer uptake time was 60 min. The acquisition protocol consisted in a CT scan to be used for anatomical localization and attenuation correction and a 3D emission scan of 15 min (single PET FOV) recording 311.6 Mcounts.

The reconstruction parameters common to all the algorithms were: image matrix 256×256, 28 subsets, up to 5 iterations, reconstruction FOV 50 cm, pixel size: 1.95 mm. The applied post–filter was characterized by the same parameters as in the clinical practice, i.e. symmetric two-dimensional Gaussian filter with FWHM=3 mm in the transaxial planes and a weighted 3-point mean along the axial direction.

To assess the improvements introduced by PSF, a qualitative comparison of the three different sets of images (RnoPSF, RnoPSF-Filt and RwPSF) was performed by an expert neurologist on the images obtained after 2, 3 and 5 iterations.

In figures 7.15 and 7.16 some representative transaxial and sagittal images of the neurological patient are shown. As stated for the oncology patient, the RnoPSF reconstruction is characterized by a high level of noise and needs the clinical filtering. The introduction of a filter, even if very light (FWHM=3 mm), drastically reduces the visual noise but also the spatial resolution in the image.

RwPSF maintains an acceptable level of visual noise, while enhancing the different structures and their "functional" morphology, thanks to the recovery of the spatial resolution and the increase of the contrast: this improvement in image quality is particularly noticeable in the sagittal images.

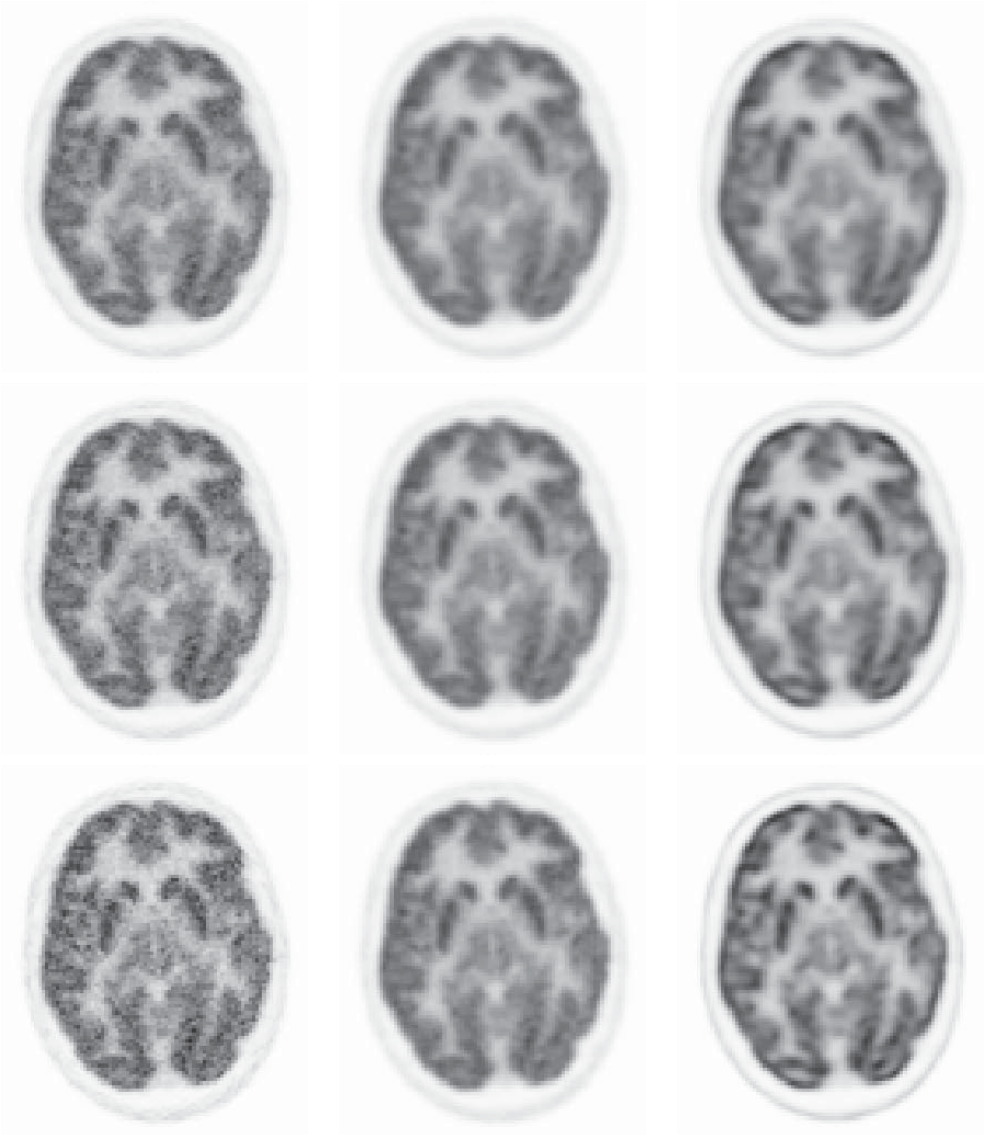


Figure 7.15: Transaxial images of the neurological patient reconstructed using 2 (top), 3 (middle) and 5 iterations (bottom): RnoPSF (left), RnoPSF filtered with a FWHM=3 mm Gaussian (centre), RwPSF (right) [the images are shown using the same display parameters]

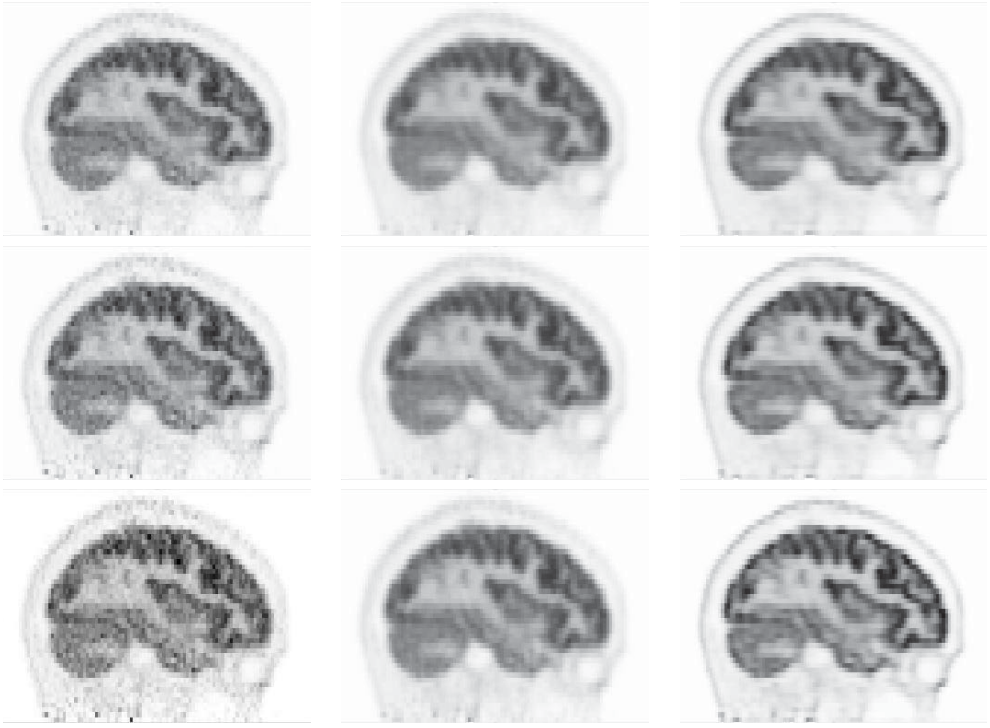


Figure 7.16: Sagittal images of the neurological patient reconstructed using 2 (top), 3 (middle) and 5 iterations (bottom): RnoPSF (left), RnoPSF filtered with a FWHM=3 mm Gaussian (centre), RwPSF (right) [the images are shown using the same display parameters]

7.4 Effects of the regularization

7.4.1 Optimization of the regularization parameters

The optimization was performed by reconstructing an acquisition of a NEMA IEC Body Phantom (acquired statistics: 52.4 Mcounts, see section 7.3.1 for further details) using different sets of parameters (β, δ) and choosing the set which maximized the detectability index D (see section 6.3) relative to the smallest sphere (diameter 10 mm), i.e. a representative sphere of many oncological lesions. Larger lesions would suffer from less spatial resolution loss due to the regularization process, with consequent better results in terms of definition and contrast of the activity distribution.

The optimization was performed using 10 iterations, considered as a compromise between clinical requirements of time and recovery of spatial resolution thanks to the presence of PSF in the reconstruction algorithm.

In figure 7.17a the detectability for the smallest sphere is plotted as a function of the parameters β and δ for the proposed prior. The maximum detectability is

obtained for

$$\boxed{\beta = 0.002, \quad \delta = 0.3} \tag{7.5}$$

With a similar strategy, it is possible to obtain the optimal parameters for the GTV prior. In figure 7.17b the detectability for the smallest sphere using the GTV prior is plotted as a function of the parameters β and δ: the maximization is obtained for

$$\boxed{\beta = 0.015, \quad \delta = 0.2} \tag{7.6}$$

The plots in figure 7.17 also reveal a basic difference between the GTV prior and the proposed one.

Both priors need two regularization parameters: the regularization strength β (which controls the effect of the regularization strategy on the final image) and the signal-background threshold δ (which distinguishes the different behaviours to be applied in the different regions of the image). Among them, the latter parameter is, in principle, more intricate and delicate to set, since different regions in the same study might require different thresholds to maximize the general spatial resolution preservation.

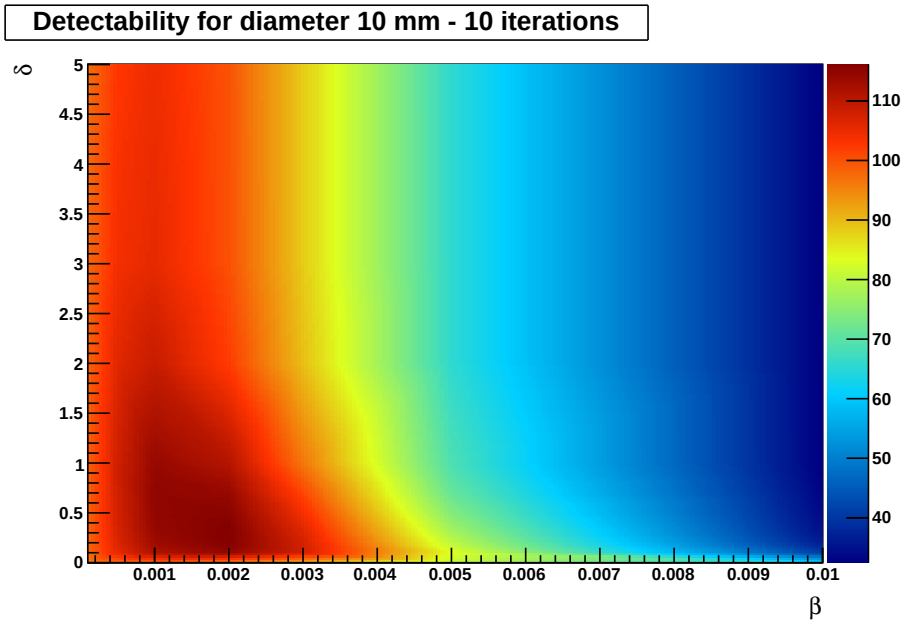
As stated above, the maximization results hint a different behaviour of the two priors: the GTV prior shows a very narrow peak along the direction of the δ parameter, suggesting very different behaviours for small changes of this parameter and, consequently, indicating that setting this parameter could be a delicate process. For the proposed prior, small changes of the δ parameter around the maximum lead to very little differences in its behaviour, guaranteeing higher stability and "margins of security".

7.4.2 The edge effect

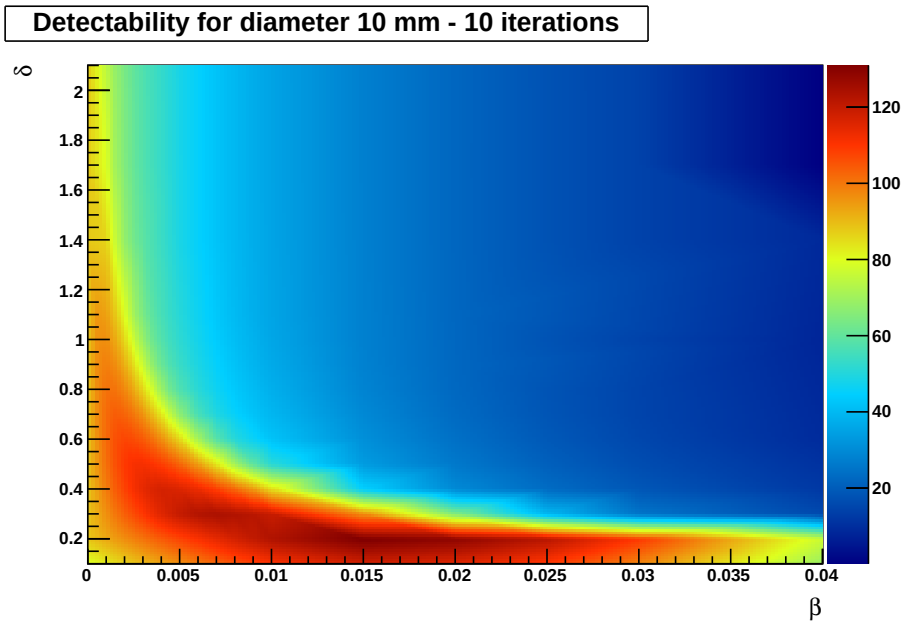
As demonstrated in section 5.3, the use of PSF introduces an *unavoidable* enhancement of sharp transitions, originating artefacts and spurious structures. As partially shown in figure 5.4, an example of this effect may be shown by filling three spheres (in the case here presented, with diameters 30 mm, 23 mm and 15 mm) with a *homogeneous* FDG solution (activity concentration of 90 kBq/ml), acquiring them in air (in the case here presented, for 90 s, recording 11.1 Mcounts) and reconstructing them in absence and in presence of PSF recovery. In this section the previous example will be expanded, in particular to test the capability of the different regularization priors considered to reduce the strength of the edge effect.

In figure 7.18 a comparison of the different reconstruction algorithms considered is presented.

The chosen reconstruction parameters were: image matrix 256×256, 28 subsets, reconstruction FOV 60 cm, pixel size 2.34 mm. The clinical post-filter used in RnoPSF was coincident with the one used in the clinical practice, i.e. it was



(a) Proposed prior



(b) Gauss-Total Variation

Figure 7.17: Optimization of the regularization parameters for the proposed (top) and the Gauss-Total Variation (bottom) priors by maximizing the D index

composed of a symmetric two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a three-point mean along the axial direction. The post-filter used in RwPSF-Filt was composed of a symmetric two-dimensional Gaussian filter with FWHM=4.28 mm in the transaxial planes and a three-point mean along the axial direction. In the case of RwPSF-GTVR and RwPSF-R, the regularization parameters were set equal to the optimal values found in section 7.4.1, i.e. respectively $(0.015, 0.2)$ and $(0.002, 0.3)$. For RwPSF-PR, the parameter β was set equal to the RwPSF-R one, i.e. $\beta = 0.002$.

The convergence for spheres in air is expected to be faster than in the other phantom studies and so the number of iterations was set to 5, to limit the noise content in RnoPSF (and, to a lesser extent, in RwPSF) to an acceptable level (in particular for the drawn activity profiles). For higher number of iterations the regularized images did not change appreciably, confirming the faster convergence and the stability of the algorithm.

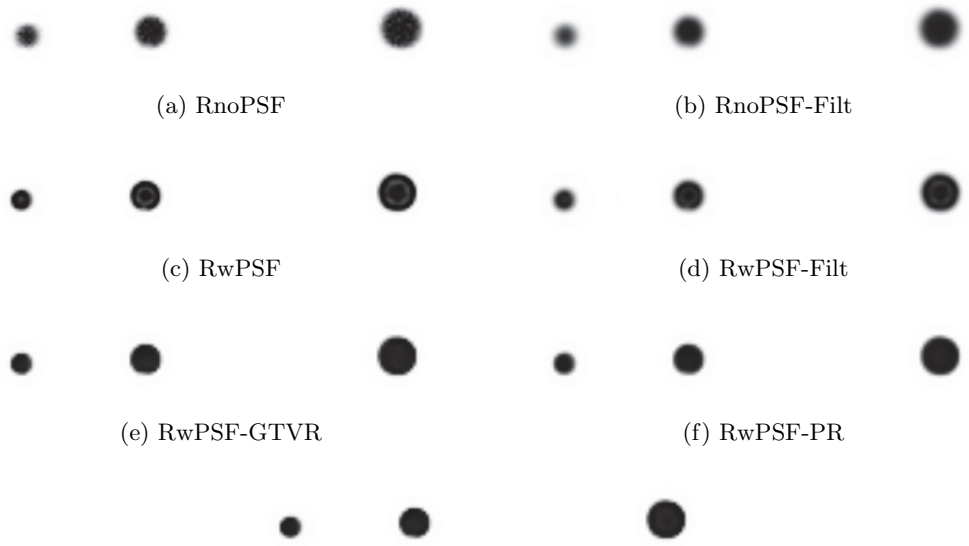
The edge effect is absent from RnoPSF (figure 7.18a) and RnoPSF-Filt (figure 7.18b) images, with the only difference being the expected poorer spatial resolution and lower noise of the latter ones. As anticipated and demonstrated in section 5.3, the reconstructed images obtained with RwPSF (figure 7.18c) show very pronounced edges in each sphere. Furthermore, inside the sphere the activity distribution oscillates along each radial direction similarly to a typical Gibbs effect [42, 108]. These effects are partially mitigated (but still easily recognizable) in RwPSF-Filt images (figure 7.18d), but with an evident loss of spatial resolution.

The three regularization priors — RwPSF-GTVR (figure 7.18e), RwPSF-PR (figure 7.18f) and RwPSF-R (figure 7.18g) — appear to control the edge over-enhancement and the ringing effects with limited loss of spatial resolution (with slightly better results for RwPSF-R and RwPSF-GTVR); moreover, no unnatural image flattening or "sawtoothing" effects (e.g. in the outer contour of the spheres) are visible, as produced instead by some regularization strategies [50].

These qualitative evaluations are confirmed also by drawing two line profiles (along the radial and tangential directions) passing through the centre of the largest sphere, in order to analyze the oscillations in the two directions. The profiles obtained for RwPSF, RwPSF-Filt, RwPSF-GTVR, RwPSF-PR and RwPSF-R are presented in figure 7.19. The three regularization priors are comparable in terms of reduction in the edge over-enhancement — with slightly higher uniformity for RwPSF-PR — and loss of spatial resolution — with slightly better results for RwPSF-R and RwPSF-GTVR — confirming the qualitative results.

7.4.3 Quantitative accuracy

The quantitative validation was performed by using the NEMA IEC Body Phantom presented in section 7.3.1, with identical acquisition protocols. The chosen



(a) RnoPSF (b) RnoPSF-Filt

(c) RwPSF (d) RwPSF-Filt

(e) RwPSF-GTVR (f) RwPSF-PR

(g) RwPSF-R

Figure 7.18: Comparison of the different reconstruction algorithms concerning the edge effect

reconstruction parameters were: image matrix 256×256, 28 subsets, reconstruction FOV 60 cm, pixel size 2.34 mm, up to 50 iterations. The clinical post-filter used in RnoPSF was coincident with the one used in the clinical practice, i.e. it was composed of a symmetric two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a three-point mean along the axial direction. The post-filter used in RwPSF-Filt was composed of a symmetric two-dimensional Gaussian filter with FWHM=4.28 mm in the transaxial planes and a three-point mean along the axial direction. In the case of RwPSF-GTVR and RwPSF-R, the regularization parameters were set equal to the optimal values found in section 7.4.1, i.e. respectively $(0.015, 0.2)$ and $(0.002, 0.3)$. For RwPSF-PR, the parameter β was set equal to the RwPSF-R one, i.e. $\beta = 0.002$.

In table 7.1 the percent differences (with respect to RnoPSF reconstructions) of CR_{hot} coefficients, obtained with the different algorithms, for the largest sphere (37-mm diameter) and the smallest sphere (10-mm diameter) and of the COV at 5, 10 and 50 iterations are reported. Similarly, in table 7.2 the percent differences, with respect to the clinical RwPSF-Filt reconstructions, obtained from the different regularized algorithms and from RwPSF are reported. In figure 7.21 the CR_{hot} coefficients for the largest and smallest spheres are plotted versus the background coefficient of variation, while in figure 7.20 an example of the reconstructed images using the different algorithms is presented.

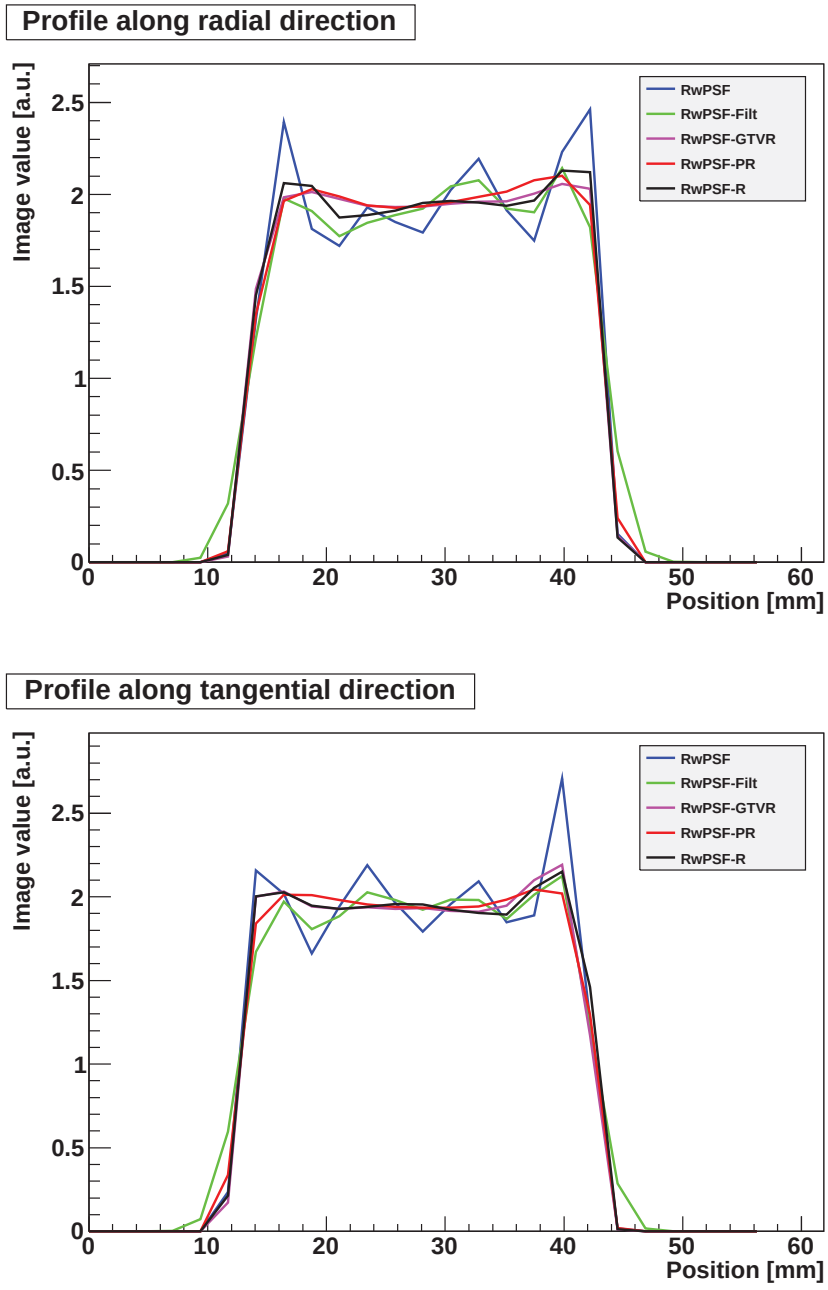


Figure 7.19: Activity line profiles in the uniform spheres obtained with the different regularization priors: top, profile drawn along the scanner radial direction; bottom, profile drawn along the scanner tangential direction

As it can be seen from tables 7.1 and 7.2 and figure 7.21, RwPSF-R results in the most accurate quantification among the other regularization priors (RwPSF-PR and RwPSF-GTVR). It also provides better quantification than RnoPSF-Filt and — for the smallest sphere in the first iterations (up to about 15 iterations, i.e. also farther than the optimization point of the regularization parameters) — than RwPSF-Filt. For the largest sphere, instead, RwPSF-R gives always better results than RwPSF-Filt. The contrast recovery coefficients obtained with RwPSF-R are lower, as expected, than in the RwPSF case.

Moreover, RwPSF-R also contributes to noise reduction, with background uniformity higher than RnoPSF-Filt, RwPSF and RwPSF-Filt and slightly lower, but comparable, to RwPSF-PR. RwPSF-GTVR provides the lowest background COV, but at the expense of a moderate reduction in the quantification accuracy.

From the graphs in figure 7.21 the regularized reconstruction algorithms show also their higher capability to get closer to the convergence of the algorithm with respect to RnoPSF, RnoPSF-Filt, RwPSF and RwPSF-Filt. In fact, each of RwPSF-GTVR, RwPSF-PR and RwPSF-R graphs converges towards a specific point $\left(\overline{CR_{hot}}, \overline{COV}\right)$, meaning that the image remains nearly equal to itself and, consequently, gives stable quantitative results. This stability is reached after about 25 iterations, with lower number of iterations for larger dimensions of the sphere.

The images[†] of the phantom (figure 7.20) confirm the numerical results. RwPSF-GTVR drastically reduces the noise content, but with an important loss of spatial resolution and an unnatural flattening of background (as confirmed in literature, e.g. [50]). RwPSF-PR and RwPSF-R provide high background uniformity (with natural appearance), while retaining good definition of the spheres, with slightly better spatial resolution for RwPSF-R.

[†]All the images presented hereafter are compared at 5 iterations, to propose a direct comparison with the clinical practice.

Recon algorithm	**ΔCR_{hot} for the largest sphere**		
	5 it.	10 it.	50 it.
RnoPSF-Filt	-6.6%	-7.0%	-7.6%
RwPSF	+4.9%	+6.2%	+7.8%
RwPSF-Filt	-0.6%	-0.1%	0.0%
RwPSF-GTVR	-1.9%	-1.7%	-2.4%
RwPSF-PR	+2.4%	+2.6%	+2.1%
RwPSF-R	+3.4%	+4.0%	+3.7%

Recon algorithm	**ΔCR_{hot} for the smallest sphere**		
	5 it.	10 it.	50 it.
RnoPSF-Filt	-29.4%	-29.0%	-30.1%
RwPSF	+12.3%	+31.6%	+43.6%
RwPSF-Filt	-15.9%	-2.9%	+6.1%
RwPSF-GTVR	-37.0%	-25.6%	-23.5%
RwPSF-PR	-17.7%	-10.3%	-9.0%
RwPSF-R	-11.7%	+0.6%	+3.8%

Recon algorithm	**ΔCOV**		
	5 it.	10 it.	50 it.
RnoPSF-Filt	-71.9%	-77.1%	-83.0%
RwPSF	-53.2%	-50.9%	-12.5%
RwPSF-Filt	-73.3%	-75.0%	-70.3%
RwPSF-GTVR	-87.6%	-92.1%	-95.1%
RwPSF-PR	-79.6%	-85.6%	-90.6%
RwPSF-R	-79.3%	-85.2%	-90.2%

Table 7.1: Percent differences of CR_{hot} coefficients for the largest sphere (37-mm diameter) and the smallest sphere (10-mm diameter) and of the background coefficient of variation COV with respect to RnoPSF reconstructions

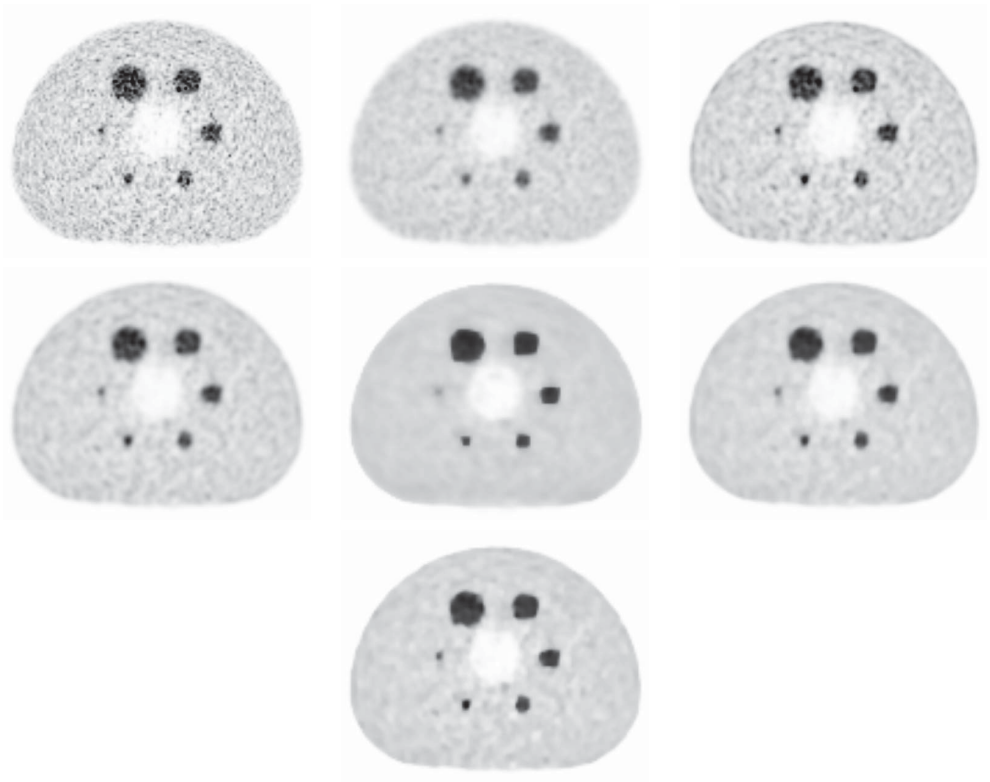


Figure 7.20: NEMA IEC phantom at 5 iterations: RnoPSF (top left), RnoPSF filtered with a FWHM=5 mm Gaussian (top center), RwPSF (top right), RwPSF filtered with a FWHM=4 mm Gaussian (middle left), RwPSF-GTVR (middle center), RwPSF-PR (middle right) and RwPSF-R (bottom) [the images are shown using the same display parameters]

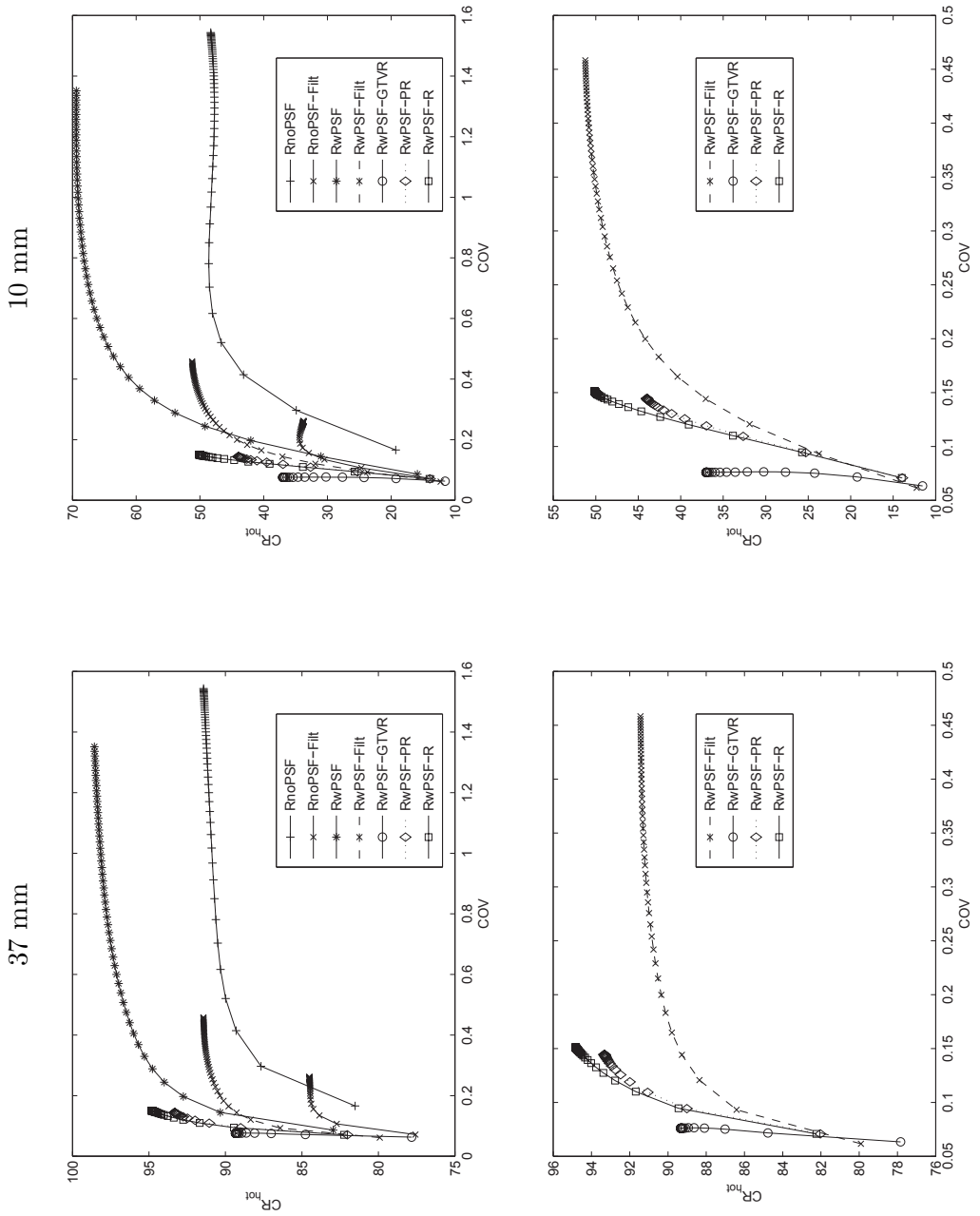


Figure 7.21: NEMA IEC phantom: comparison between the contrast recovery coefficients (CR_{hot}) for the 37-mm diameter sphere (left) and for the 10-mm diameter sphere (right) versus the background coefficient of variation for the reconstruction algorithms considered (top) and RwPSF-Filt and the different regularization priors used (bottom) [each point represents one iteration]

Recon algorithm	**ΔCR_{hot} for the largest sphere**		
	5 it.	10 it.	50 it.
RwPSF	+5.5%	+6.3%	+7.8%
RwPSF-GTVR	-1.3%	-1.6%	-2.4%
RwPSF-PR	+3.0%	+2.7%	+2.1%
RwPSF-R	+4.0%	+4.1%	+3.7%

Recon algorithm	**ΔCR_{hot} for the smallest sphere**		
	5 it.	10 it.	50 it.
RwPSF	+33.6%	+35.4%	+35.4%
RwPSF-GTVR	-25.1%	-23.4%	-27.9%
RwPSF-PR	-2.1%	-7.7%	-14.2%
RwPSF-R	+5.1%	+3.6%	-2.1%

Recon algorithm	**ΔCOV**		
	5 it.	10 it.	50 it.
RwPSF	+75.1%	+96.3%	+194.8%
RwPSF-GTVR	-53.6%	-68.6%	-83.4%
RwPSF-PR	-23.7%	-42.3%	-68.4%
RwPSF-R	-22.7%	-40.7%	-67.0%

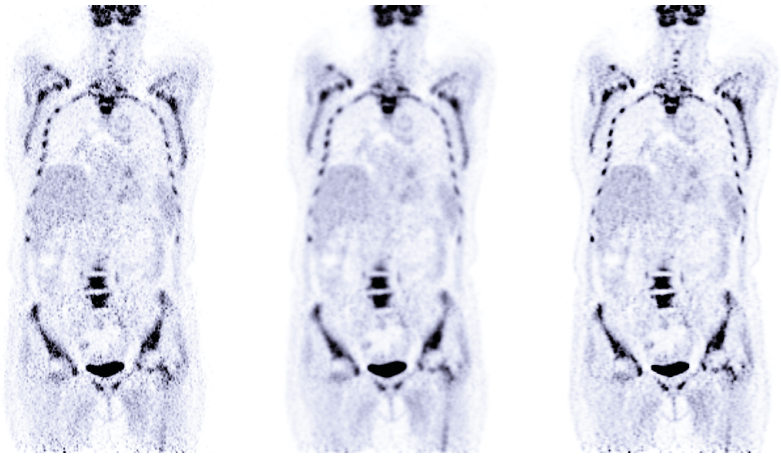
Table 7.2: Percent differences of CR_{hot} coefficients for the largest sphere (37-mm diameter) and the smallest sphere (10-mm diameter) and of the background coefficient of variation COV with respect to RwPSF-Filt reconstructions

7.4.4 Qualitative improvements

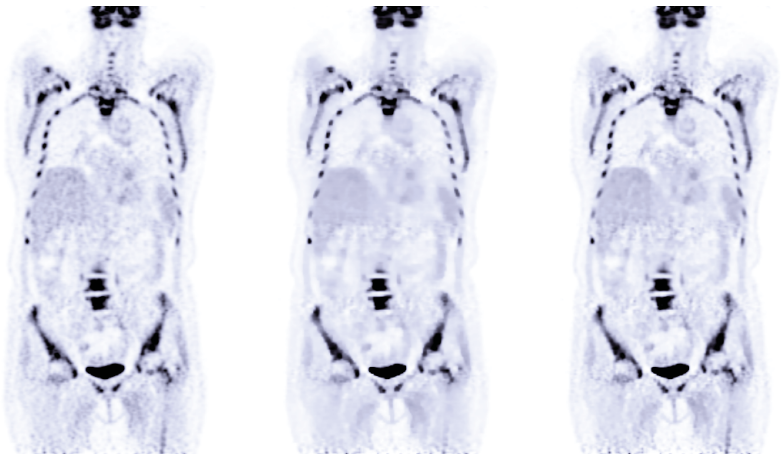
Two oncological patients A and B (56 kg and 53 kg, respectively) received an injection of ^{18}F-FDG (296 MBq and 277.5 MBq, respectively); the tracer uptake time was 60 min. The acquisition protocol consisted in a whole body CT scan to be used for anatomical localization and attenuation correction and a 3D whole body emission scan (2.5 min per bed position, 6 bed positions, 7-slice overlap between bed positions) recording 390.7 and 267.5 Mcounts, respectively.

The reconstruction parameters common to all the algorithms were: image matrix 256×256, 28 subsets, 5 iterations, reconstruction FOV 60 cm, pixel size 2.34 mm. The post–filter applied in RnoPSF-Filt was characterized by the same parameters as in the clinical practice, i.e. symmetric two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a weighted 3-point mean along the axial direction. The post-filter used in RwPSF-Filt was composed of a symmetric two-dimensional Gaussian filter with FWHM=4.28 mm in the transaxial planes and a three-point mean along the axial direction. In the case of RwPSF-GTVR and RwPSF-R, the regularization parameters were set equal to the optimal values found in section 7.4.1, i.e. respectively $(0.015, 0.2)$ and $(0.002, 0.3)$. For RwPSF-PR, the parameter β was set equal to the RwPSF-R one, i.e. $\beta = 0.002$.

In figures 7.22 and 7.23 some representative examples of the images obtained from clinical studies at 5 iterations are provided. The figures qualitatively confirm the (qualitative and quantitative) results obtained with the NEMA IEC phantom. RnoPSF yields very high level of noise, which is drastically reduced by the postfilter in RnoPSF-Filt, but at the expense of an additional spatial resolution loss. RwPSF improves the level of details recognizable in the images, but needs some post treatment to be used for clinical report due to the excessive level of noise. The use of a postfilter (RwPSF-Filt) dramatically spoils the gain in spatial resolution thanks to PSF modelling. RwPSF-GTVR shows very smooth regions, confirming (and, in these cases, also enhancing) the unnatural appearance noticed with NEMA IEC Body Phantom. RwPSF-PR provides a good control of the noise with some loss of spatial resolution, while RwPSF-R appears to yield very positive results in terms of uniformity of regions (e.g. the liver) and definition of the activity distribution (e.g. ribs and vertebrae in the first patient, hepatic lesions in the second one).



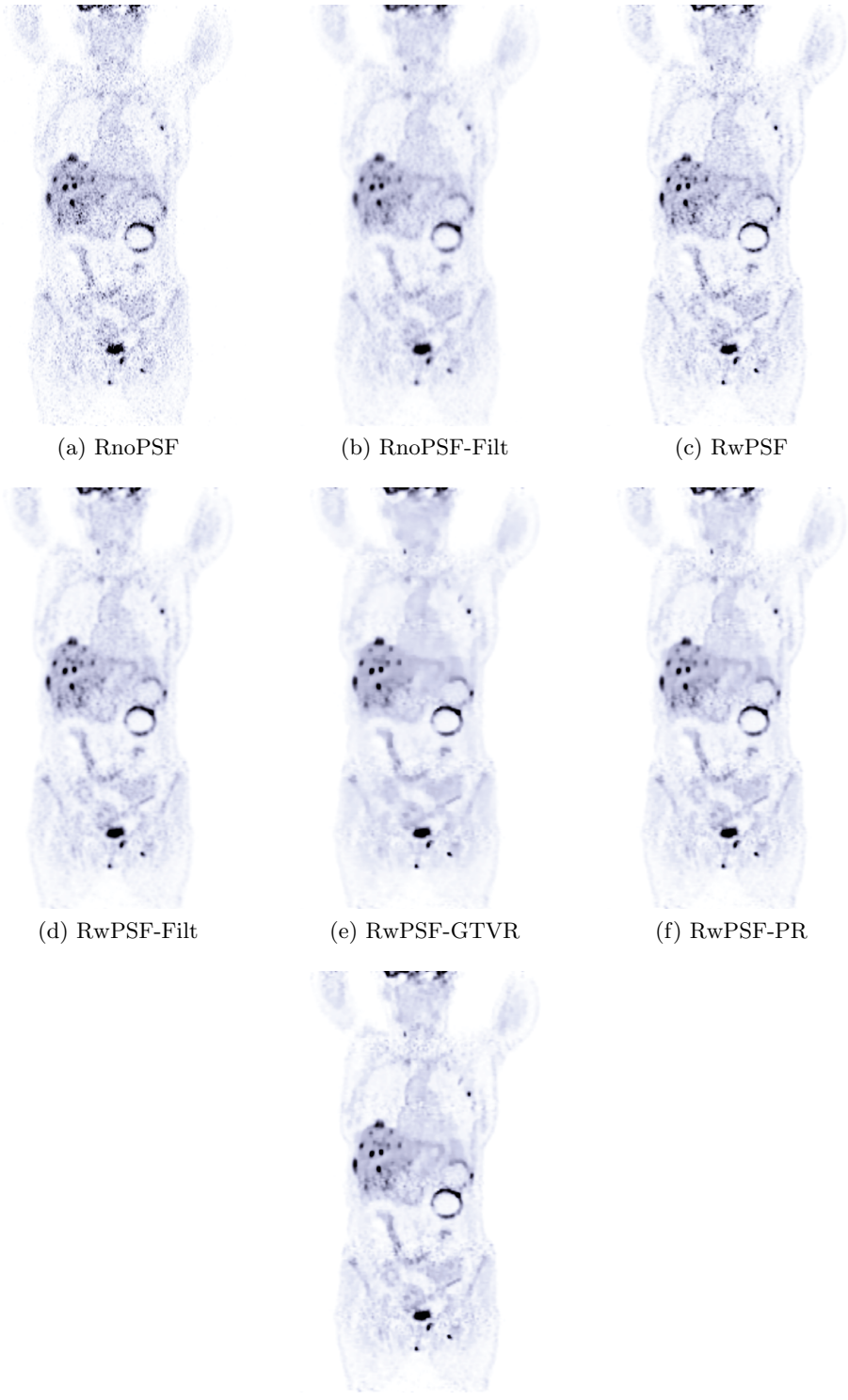
(a) RnoPSF (b) RnoPSF-Filt (c) RwPSF



(d) RwPSF-Filt (e) RwPSF-GTVR (f) RwPSF-PR

(g) RwPSF-R

Figure 7.22: Oncological patient A: comparison of the different reconstruction algorithms [the images are shown using the same display parameters]



(a) RnoPSF (b) RnoPSF-Filt (c) RwPSF

(d) RwPSF-Filt (e) RwPSF-GTVR (f) RwPSF-PR

(g) RwPSF-R

Figure 7.23: Oncological patient B: comparison of the different reconstruction algorithms [the images are shown using the same display parameters]

Chapter 8

Results — TOF imaging

8.1 Description of the GE Discovery 690 scanner

The Discovery 690 (D690) is an integrated system which combines a LYSO block detector PET tomograph with a 64-slice CT scanner.

The PET scanner has a multi-ring system design. The PET tomograph consists of 13824 LYSO crystals with dimensions of $4.2 \times 6.3 \times 25$ mm^3. The PET detection unit is a block of 54 (9×6) individual LYSO crystals coupled to a single squared photomultiplier tube with 4 anodes. The D690 uses a low energy threshold of 425 keV and a coincidence time window of 4.9 ns. The D690 consists of 24 rings of detectors for an axial FOV of 157 mm. The transaxial FOV is 70 cm.

The D690 operates only in three-dimensional mode with an axial coincidence acceptance of ± 23 planes. A radioactive pin source of ^{68}Ge (18 MBq) is used for system calibration and daily quality control. The D690 is equipped with a powerful modular array of cell processors (IBM Blade Centre) specifically designed to speed up image reconstruction and data processing. The D690 system uses, as the standard PET reconstruction, a fully 3D OSEM algorithm with all corrections (scatter, random, dead time, attenuation, and normalization) incorporated into the iterative reconstruction scheme (official name: VUE-point HD). Furthermore, new reconstruction algorithms are available on the D690, which add to the standard VUE-point HD configuration the time of flight information (VUE-point FX) and/or a 3D model of the D690 PET point spread function (Sharp-IR). The PSF model implemented in the D690 system is based on experimental measurements of a point source acquired in different positions within the 3D-PET FOV. The resulting spatially variant PSF was modelled and then coded in a system matrix, in the projection space, to be used in the reconstruction scheme of a 3D-OSEM algorithm, as described in [4].

The CT system is the LightSpeed VCT with 64 slices. The detector matrix is characterized by 912 channels × 64 rows for a total of 58368 solid state detector elements. The LightSpeed VCT allows full 360-degree rotational scans with a variable rotational time ranging from 0.35 to 2 seconds and slice thickness of 64×0.625 mm, 32×1.25 mm, 16×2.5 mm, 8×5 mm and 4×10 mm.

8.2 PSF measurements

The PSF of the D690 was experimentally determined as described in chapter 4. The transaxial FOV was sampled at distances (from the FOV centre) ranging from 3 cm to 25 cm in steps of about 3 cm, both to the right and to the left of the centre. For each radial distance, the entire axial FOV was sampled in steps of about 1.5 cm.

An example of a fit on an acquired image is reported in figure 8.1.

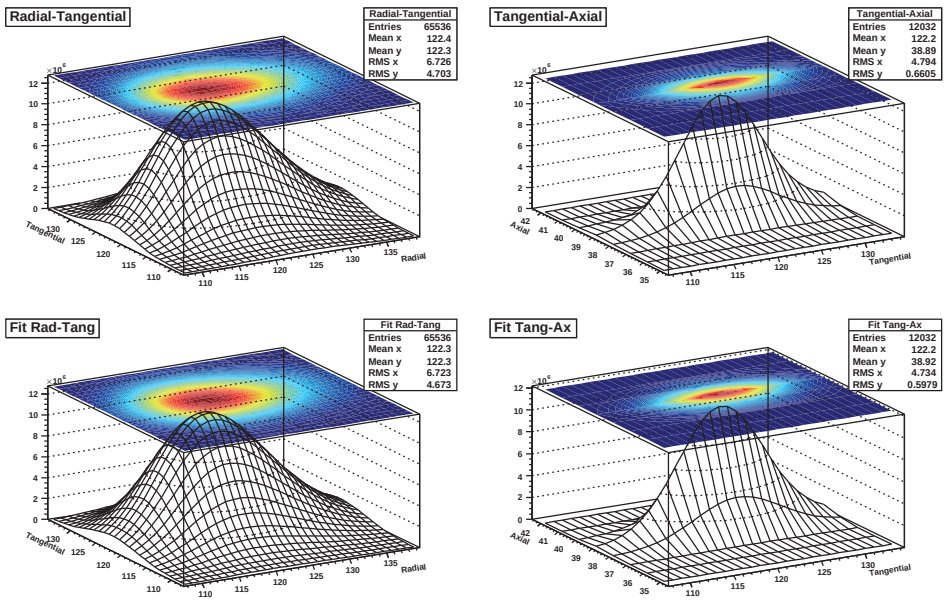


Figure 8.1: Example of PSF fits on D690. Top: radial–tangential and tangential–axial images; bottom: corresponding fits

All the obtained values for each spread parameter were then graphed as a function of both radial and axial distances from the scanner FOV centre (figure 8.2): it was then possible to fit each set of data with a polynomial quadratic in both distances, symmetric with respect to the centre of axial FOV, as defined in equation 4.8.

The relative weight of the square term in the axial dependence of the transaxial spread parameters was found to be more than two orders of magnitude lower

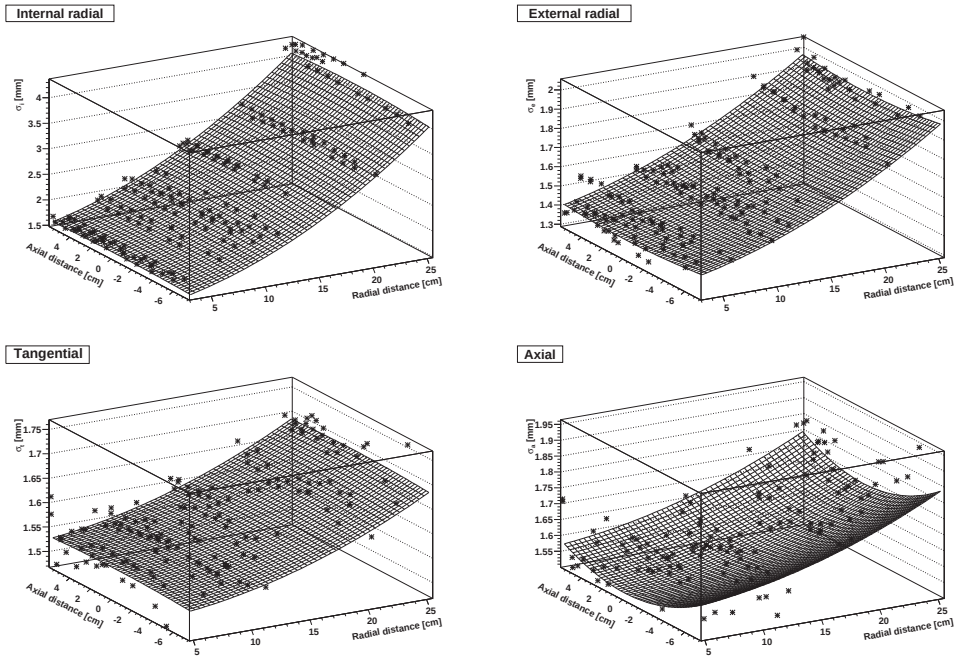


Figure 8.2: Determination of the spread parameters dependences on the radial and axial distances in Discovery 690: internal radial (top left), external radial (top right), tangential (bottom left) and axial (bottom right)

than the square term in the axial dependence of the axial spread parameter: therefore, the axial dependence of the transaxial spread parameters was assumed to be negligible, in order to reduce the complexity of the model.

The dependences used in the work were then

$$\begin{aligned}
\sigma_i &= 1.340 + 0.05816\, r + 2.297 \times 10^{-3}\, r^2 \\
\sigma_e &= 1.373 + 0.01105\, r + 5.539 \times 10^{-4}\, r^2 \\
\sigma_t &= 1.427 + 3.461 \times 10^{-4}\, r + 2.582 \times 10^{-4}\, r^2 \\
\sigma_a &= \left(1.511 + 9.8 \times 10^{-10}\, r + 3.640 \times 10^{-4}\, r^2\right)\left(1 + 1.110 \times 10^{-3}\, a^2\right)
\end{aligned} \tag{8.1}$$

where the radial (r) and axial (a) distances are measured in centimeters and the spread parameters are measured in millimeters.

In figure 8.3 the radial, tangential and axial profiles drawn on an image are compared with the corresponding ones drawn on the resulting fit function and on the final kernel function used in the reconstruction algorithm. The figure reveals the good resemblance between the different profiles and, consequently, the reliability of the fitting results.

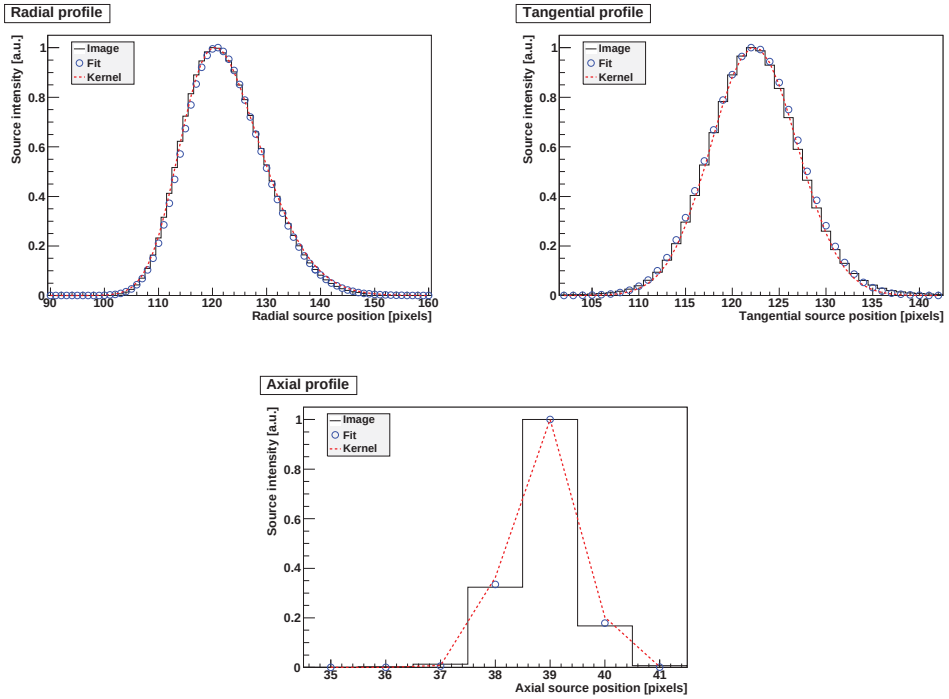


Figure 8.3: Comparison between the radial (top left), tangential (top right) and axial (bottom) profiles drawn on an image, on the resulting fit and on the corresponding kernel used in the reconstruction

8.3 Improvements thanks to PSF

The new algorithm accounting for the PSF of the D690 scanner was validated on phantom and clinical studies. All datasets were reconstructed using the 3D OSEM algorithm without (RnoPSF) and with the PSF (RwPSF). Further reconstructions were performed by applying a post-filter on the final reconstructed images (respectively, RnoPSF-Filt and RwPSF-Filt). The four reconstruction algorithms were replicated by adding also the TOF information (and obtaining as a consequence the TOF RnoPSF, TOF RnoPSF-Filt, TOF RwPSF and TOF RwPSF-Filt algorithms). The validation consisted in comparing the images obtained with the different algorithms. All the validations were qualitative; with phantom data, also quantitative comparison were performed.

8.3.1 Quantitative validation

NEMA IEC Body Phantom

The NEMA IEC Body Phantom (as described in section 7.3.1 and pictured in figure 7.7) was used to simulate an oncologic whole body ^{18}F-FDG study. Both

the body and the spheres (except the two biggest ones) were filled with a ^{18}F-FDG solution in order to have a LBR equal to 4.4. The two largest spheres (with diameters 28 mm and 37 mm) were filled with pure water, in order to simulate cold regions. After centring the phantom in the scanner FOV, a CT scan was performed to be used for attenuation correction followed by a 3D emission scan of 4 min recording 50.2 Mcounts.

The reconstruction parameters common to all the algorithms were: image matrix 256×256, 18 subsets, up to 50 iterations, reconstruction FOV 50 cm, pixel size 1.95 mm. The post filter applied in RnoPSF-Filt was a symmetric two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a weighted 3-point mean along the axial direction. The post filter applied in RwPSF-Filt was a symmetric two-dimensional Gaussian filter with FWHM=4 mm in the transaxial planes and a weighted 3-point mean along the axial direction.

The images were analysed by using the same figures of merit described in section 7.3.1:

- the *background coefficient of variation COV* was used to assess the noise content in the images;
- the *hot contrast recovery* CR_{hot} was used to measure the quantitative accuracy of the different reconstruction algorithms when dealing with hot lesions on a warm background.

Moreover, a new figure of merit was introduced, to assess also how the reconstruction algorithm represents cold regions. The *cold contrast recovery*

$$CR_{cold} = 100 \left(1 - \frac{\mu_C}{\mu_B}\right)$$

(where μ_C and μ_B are the mean content of the voxels inside a spherical VOI drawn in cold and background regions, respectively) would ideally be equal to 100 in a cold region ($\mu_C = 0$) if the scanner had infinite spatial resolution. Due to PSF (spill-in) effects, instead, the mean activity concentration in a cold region is always non-zero and, consequently, the CR_{cold} is always lower than 100.

Ten spherical VOIs were drawn, four in background regions and six for the signal (i.e. the six spheres). The signal spheres were centred on CT and the radii were chosen according to the specifications from the manufacturer of the spheres.

The background mean value μ_B and the standard deviation STD were calculated over all voxels in the first set of VOIs, while the second set of VOIs was used to calculate the contrast recovery coefficients for the spheres: μ_S was measured on each of the four smallest spheres, while μ_C on the two largest ones.

In figure 8.5 the contrast recovery coefficients for the largest (CR_{cold}) and smallest (CR_{hot}) spheres obtained with the different reconstruction algorithms

are plotted as a function of the background coefficient of variation COV, while in figure 8.6 they are plotted as a function of the number of iterations. In table 8.1 the CR_{cold} on the largest sphere, the CR_{hot} on the smallest sphere and the background COV are compared between the different reconstruction algorithms by calculating their percent variations (with respect to RnoPSF) at 5, 10 and 50 iterations. In table 8.2, instead, the reference reconstruction algorithm is TOF RnoPSF.

As expected, the effects of the post filter on the largest, cold sphere are much less pronounced than on the smallest, hot one. Also the effect of PSF in non TOF reconstruction appears to be much more important on the small hot sphere: this is probably due to the increased importance of resolution effects on small structures and, parallely, to the reduced effect of PSF on spill-in with respect to spill-out.

TOF information contributes to a significant improvement in the quantitative accuracy both on cold and hot coefficients; moreover, the reduction of the percent difference with respect to the noPSF reconstruction as iterations proceed hints an increased rate of convergence thanks to TOF (as apparent also by focusing on the convergence point of RnoPSF and TOF RnoPSF graphs in figures 8.5 and 8.6). For the hot smallest sphere, approximately 15 iterations are needed to obtain with PSF the same CR_{hot} as with TOF; for the cold largest sphere, the CR_{cold} from TOF is always higher than the one from PSF: this confirms the strong contribution given by TOF in presence of cold regions (TOF is outperformed only by the combination of TOF and PSF, see below) and, at the same time, the reduced effect introduced by PSF on regions affected by spill-in. The synergy of TOF and PSF leads to huge improvements in the contrast recoveries (at 10 iterations, +45.4% for the hot smallest sphere, +17.9% for the cold largest sphere). It is also important to note how the post-filters — identical to the one used in standard clinical practice — dramatically reduce the contrast recovery coefficients, in particular for the hot smallest sphere.

As far as noise is concerned, the introduction of TOF increases the COV for the first (approximately 15) iterations with respect to RnoPSF, while for a higher number of iterations the COV from TOF RnoPSF results lower than in RnoPSF. Moreover, as particularly apparent in figure 8.6, the convergence point of the algorithm has a lower COV than in the RnoPSF plot, all the more given the lower convergence speed of the latter algorithm. The introduction of PSF reduces the COV, demonstrating once again how the PSF introduces also a filtering effect in the background regions. For every reconstruction algorithm (RnoPSF, RwPSF, TOF RnoPSF and TOF RwPSF) the introduction of a post-filter dramatically reduces the noise content with respect to the corresponding non-filtered algorithm, at the expense of the quantitative accuracy.

The results here presented appear in agreement with those of other authors [28, 49, 92, 110].

In figure 8.4 a comparison of the images obtained after 10 iterations with the different algorithms is provided. The number of iterations was chosen by searching a compromise between the readability of the resulting images, the improvement in convergence and quantitative content and the clinical requests in terms of reconstruction time, also considering the improved computational power available on the new scanners. The images visually confirm the numerical results. The use of the TOF information (right-hand side of the figure) increases the definition of the smallest spheres and, in particular, dramatically improves the content of the cold regions, which appear with clearly lower activity concentration and much more uniform. A post-filter strongly reduces, in every case, both the noise content and the definition of the signal regions. The PSF alone improves the spatial resolution and the intensity of the signal and has a positive — but moderate — effect on the cold regions, while if used in conjunction with TOF it provides the best definition of details and coldness of the largest spheres.

Recon algorithm	**ΔCR_{cold} for the largest sphere**		
	5 it.	10 it.	50 it.
RnoPSF-Filt	-3.9%	-4.3%	-4.3%
RwPSF	+0.9%	+3.0%	+9.3%
RwPSF-Filt	-2.9%	-1.3%	+3.6%
TOF RnoPSF	+18.3%	+13.7%	+12.5%
TOF RnoPSF-Filt	+13.6%	+8.9%	+7.7%
TOF RwPSF	+20.2%	+17.9%	+20.8%
TOF RwPSF-Filt	+15.7%	+13.0%	+15.0%

Recon algorithm	**ΔCR_{hot} for the smallest sphere**		
	5 it.	10 it.	50 it.
RnoPSF-Filt	-32.5%	-35.2%	-38.4%
RwPSF	+1.0%	+15.2%	+28.8%
RwPSF-Filt	-29.0%	-22.4%	-18.7%
TOF RnoPSF	+30.9%	+22.6%	+15.4%
TOF RnoPSF-Filt	-17.6%	-25.9%	-31.1%
TOF RwPSF	+44.3%	+45.4%	+57.4%
TOF RwPSF-Filt	-3.3%	-6.1%	-7.2%

Recon algorithm	**ΔCOV**		
	5 it.	10 it.	50 it.
RnoPSF-Filt	-70.1%	-76.6%	-85.2%
RwPSF	-54.3%	-54.8%	-31.4%
RwPSF-Filt	-74.3%	-77.6%	-78.5%
TOF RnoPSF	+17.9%	+8.1%	-10.6%
TOF RnoPSF-Filt	-71.2%	-79.9%	-88.7%
TOF RwPSF	-47.3%	-50.0%	-29.6%
TOF RwPSF-Filt	-73.1%	-78.2%	-81.6%

Table 8.1: Percent differences of CR_{cold} coefficient for the largest sphere (37-mm diameter), of CR_{hot} for the smallest sphere (10-mm diameter) and of the background coefficient of variation COV with respect to RnoPSF reconstructions

Recon algorithm	**ΔCR_{cold} for the largest sphere**		
	5 it.	10 it.	50 it.
RnoPSF	-15.5%	-13.7%	-11.1%
RnoPSF-Filt	-18.8%	-15.9%	-15.0%
RwPSF	-14.7%	-9.4%	-2.9%
RwPSF-Filt	-17.9%	-13.2%	-7.9%
TOF RnoPSF-Filt	-3.9%	-4.2%	-4.3%
TOF RwPSF	+1.6%	+3.7%	+7.3%
TOF RwPSF-Filt	-2.2%	-0.7%	+2.2%

Recon algorithm	**ΔCR_{hot} for the smallest sphere**		
	5 it.	10 it.	50 it.
RnoPSF	-23.6%	-22.6%	-13.4%
RnoPSF-Filt	-48.4%	-47.1%	-46.6%
RwPSF	-22.9%	-6.0%	+11.6%
RwPSF-Filt	-45.8%	-36.7%	-29.6%
TOF RnoPSF-Filt	-37.1%	-39.6%	-40.3%
TOF RwPSF	+10.2%	+18.7%	+36.4%
TOF RwPSF-Filt	-26.2%	-23.4%	-19.6%

Recon algorithm	**ΔCOV**		
	5 it.	10 it.	50 it.
RnoPSF	-15.2%	-8.1%	+11.8%
RnoPSF-Filt	-74.6%	-78.3%	-83.5%
RwPSF	-61.2%	-58.2%	-23.3%
RwPSF-Filt	-78.2%	-79.2%	-76.0%
TOF RnoPSF-Filt	-75.6%	-81.4%	-87.3%
TOF RwPSF	-55.3%	-53.8%	-21.3%
TOF RwPSF-Filt	-77.2%	-79.9%	-79.5%

Table 8.2: Percent differences of CR_{cold} coefficient for the largest sphere (37-mm diameter), of CR_{hot} for the smallest sphere (10-mm diameter) and of the background coefficient of variation COV with respect to TOF RnoPSF reconstructions

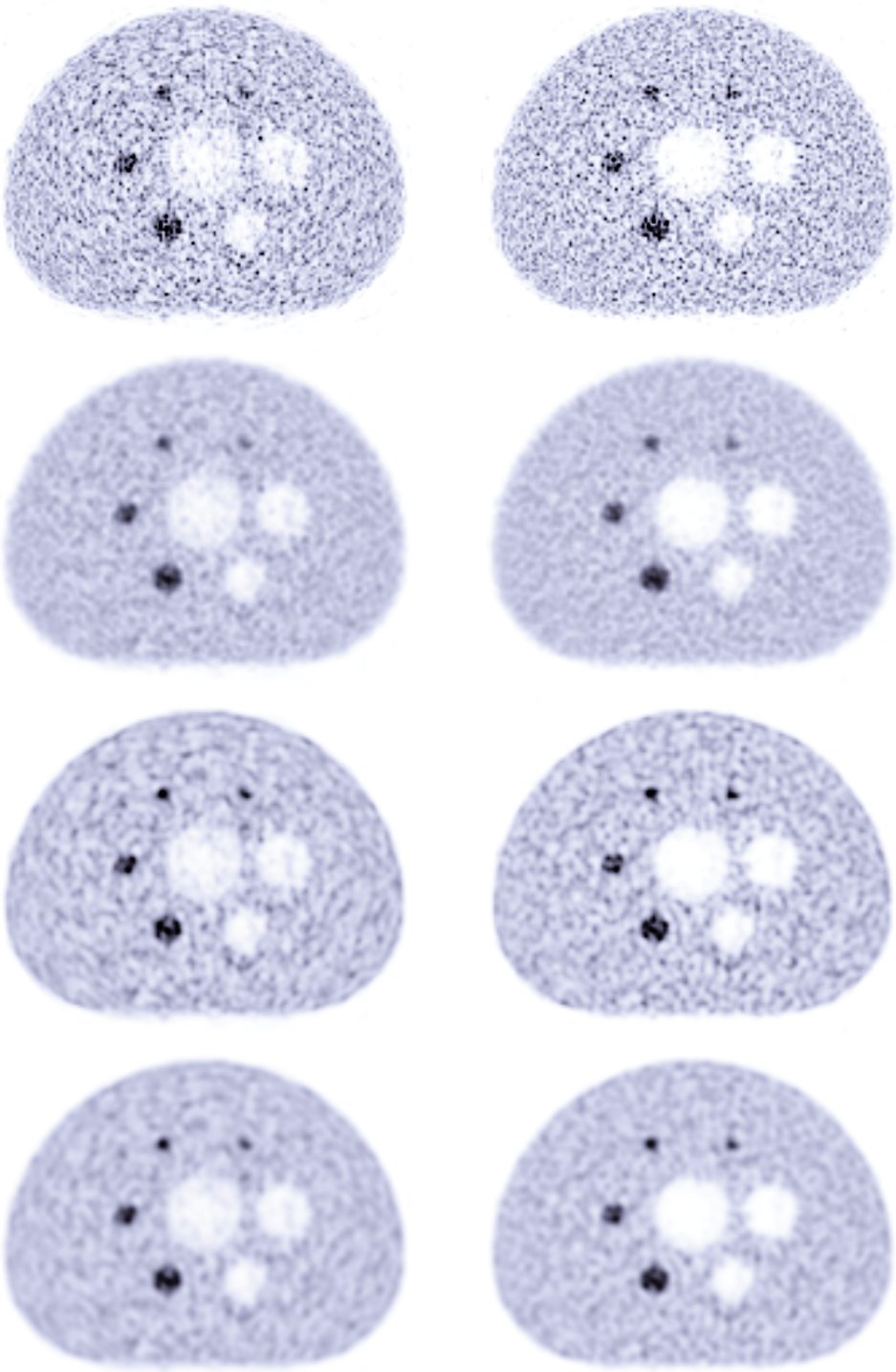


Figure 8.4: NEMA IEC Body Phantom at 10 iterations: comparison of RnoPSF, RnoPSF-Filt, RwPSF and RwPSF-Filt (from top to bottom) without (left) and with (right) TOF information [the images are shown using the same display parameters]

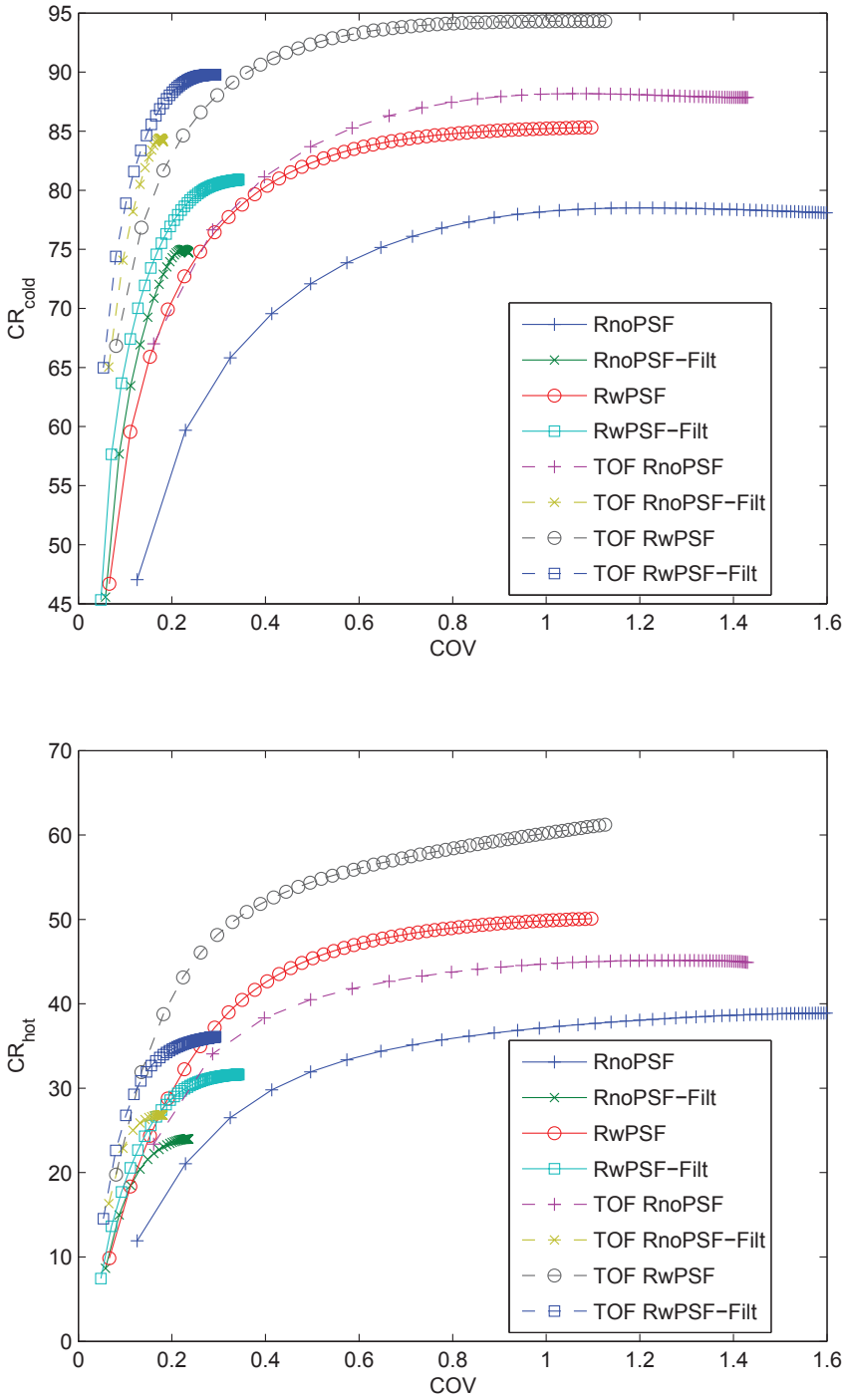


Figure 8.5: CR_{cold} coefficient for the largest sphere (top) and CR_{hot} coefficient for the smallest sphere (bottom) versus the background COV for non TOF and TOF reconstruction algorithms [each point represents one iteration]

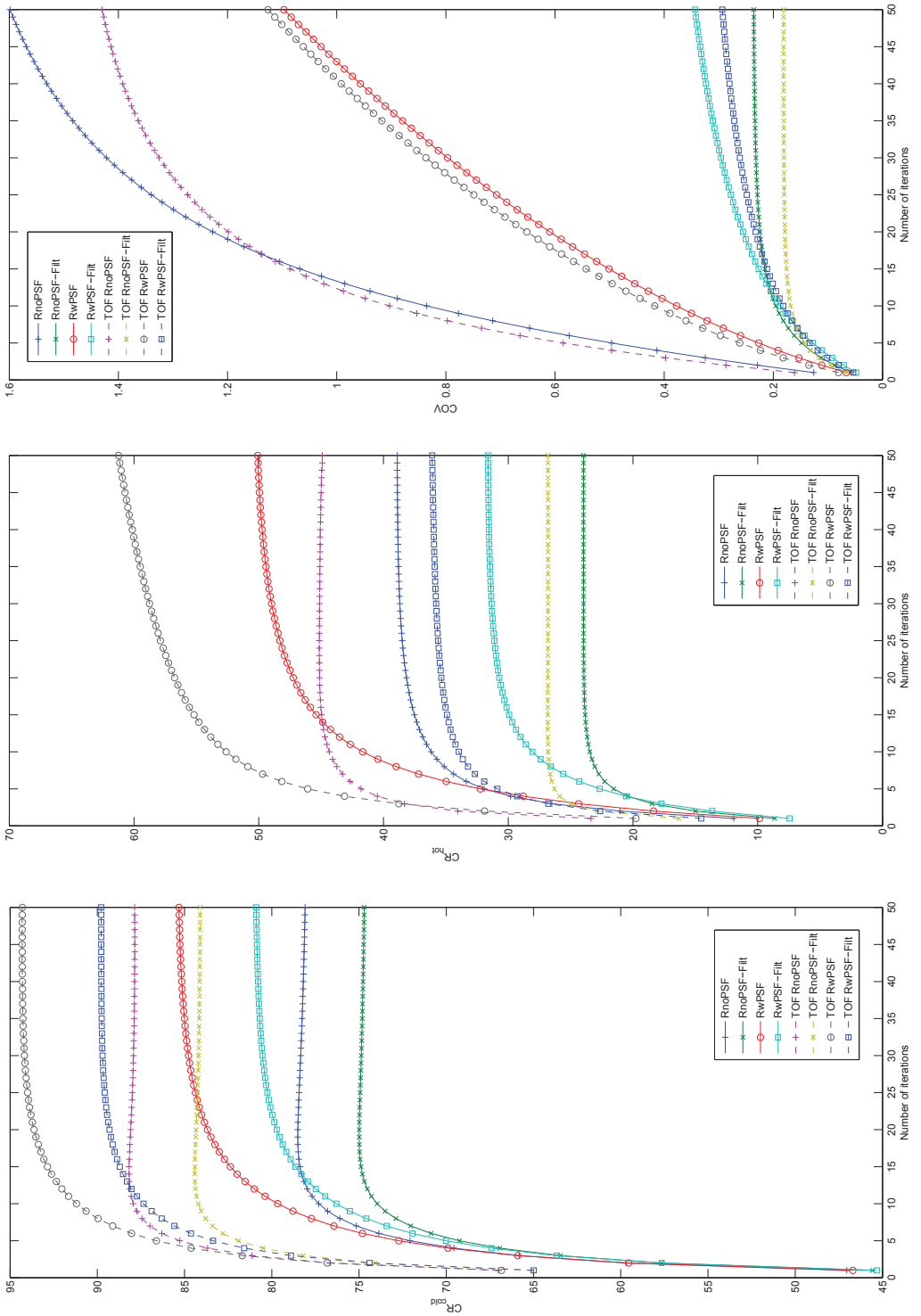


Figure 8.6: CR_{cold} coefficient for the largest sphere (left), CR_{hot} coefficient for the smallest sphere (center) and background coefficient of variation COV (right) versus the number of iterations for non TOF and TOF reconstruction algorithms

8.3.2 Qualitative validation

Clinical data

The acquisitions of two oncological patients (C and D) were used to assess the contribution of TOF and/or PSF on the final results.

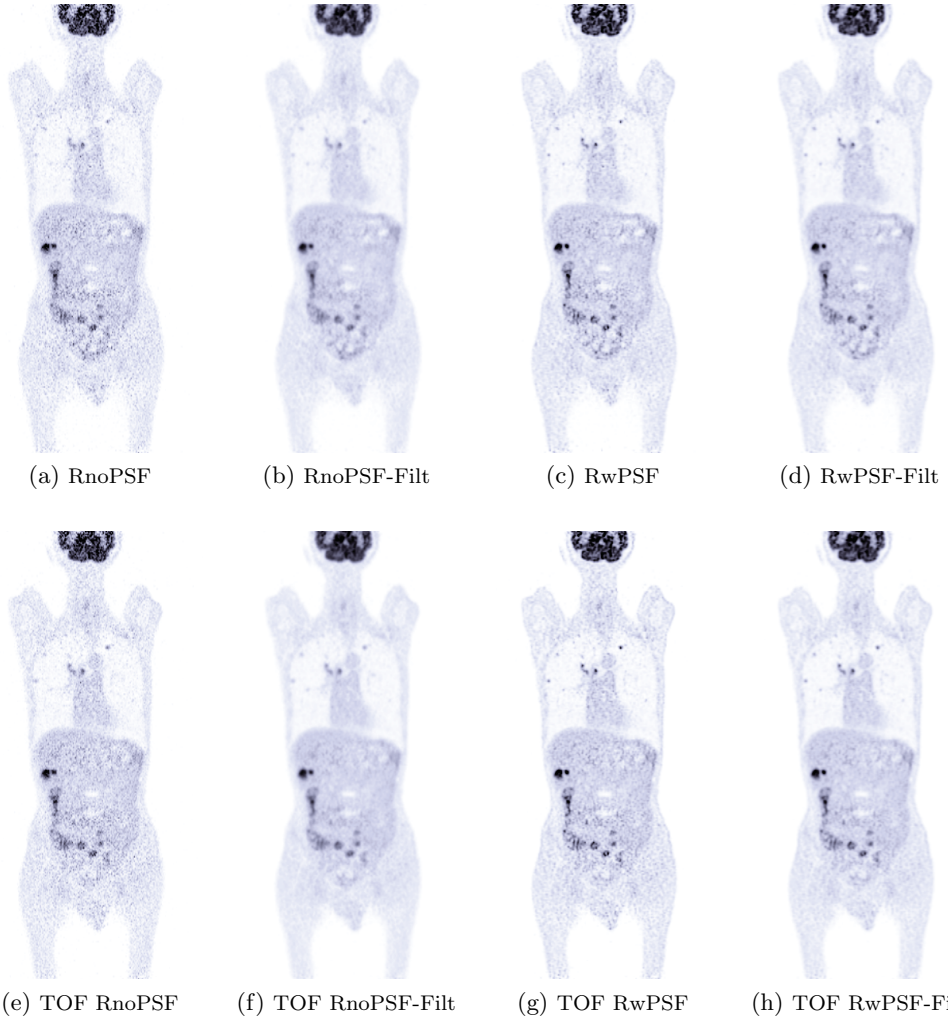
Patient C (height 175 cm, weight 53 kg) received an injection of ^{18}F-FDG (351.5 MBq); the tracer uptake time was 60 min. The acquisition protocol consisted of a whole body CT scan to be used for anatomical localization and attenuation correction and a 3D whole body emission scan (4 min for each bed containing the liver, 2 min per bed position otherwise, 8 bed positions, 9-slice overlap between adjacent beds) recording 511.6 Mcounts.

Patient D (height 175 cm, weight 75 kg) received an injection of ^{18}F-FDG (370 MBq); the tracer uptake time was 60 min. The acquisition protocol consisted of a whole body CT scan to be used for anatomical localization and attenuation correction and a 3D whole body emission scan (4 min for each bed containing the liver, 2 min per bed position otherwise, 7 bed positions, 9-slice overlap between adjacent beds) recording 341.7 Mcounts.

The chosen reconstruction parameters were: image matrix 256×256, 18 subsets, reconstruction FOV 60 cm, pixel size 2.34 mm, 10 iterations. The clinical post filter used in RnoPSF-Filt was composed of symmetric two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a three-point weighted mean along the axial direction. The clinical post filter used in RwPSF-Filt was composed of symmetric two-dimensional Gaussian filter with FWHM=4 mm in the transaxial planes and three-point weighted mean along the axial direction.

As far as patient C is concerned, in figures 8.7 and 8.8 the comparison between the different reconstruction algorithms is performed on some coronal images, while one set of transaxial images is provided in figure 8.9. Patient C is slim and, consequently, the spatial extension of the activity distribution is expected to be small, with consequent reduced effect of the TOF: in effect, this is confirmed by the images, in which — except for few small lesions — the contribution of TOF appears to be moderate, even if some improvements in the contrast of the lesions are present (e.g. see figure 8.9). The introduction of PSF, both in non TOF and TOF images, allows increasing the definition and the contrast of the different structures (e.g. the pulmonary nodules, the lesions in the liver and the thighbone). The combination of both TOF and PSF yields the best results in terms of spatial resolution, definition of the activity distribution and contrast of the different regions.

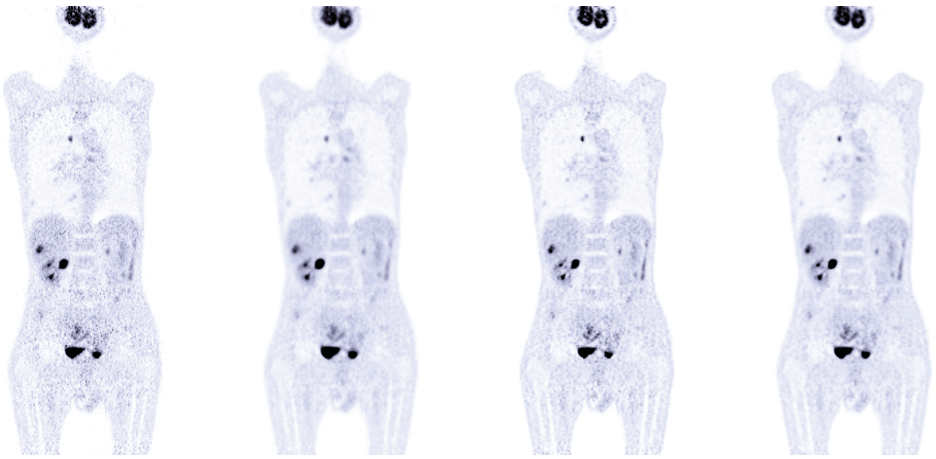
For what concerns patient D, in figures 8.10 and 8.11 the comparison between the different reconstruction algorithms is performed on a set of coronal images, while some transaxial images are provided in figure 8.12. Patient D is



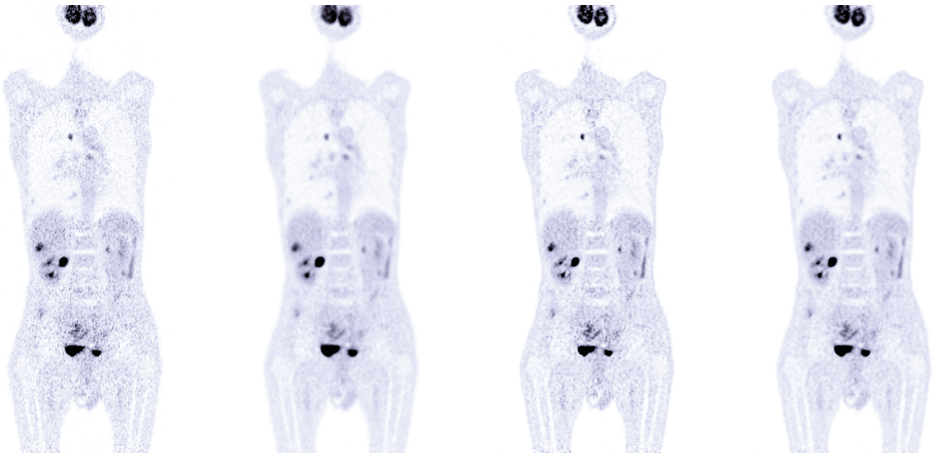
(a) RnoPSF (b) RnoPSF-Filt (c) RwPSF (d) RwPSF-Filt

(e) TOF RnoPSF (f) TOF RnoPSF-Filt (g) TOF RwPSF (h) TOF RwPSF-Filt

Figure 8.7: Coronal images of oncological patient C, comparison of the different reconstruction algorithms without (top) and with (bottom) TOF information: from left to right, RnoPSF, RnoPSF-Filt, RwPSF, RwPSF-Filt [the images are shown using the same display parameters]

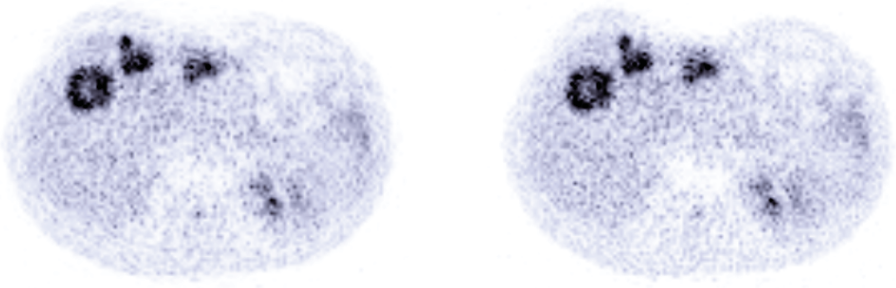


(a) RnoPSF (b) RnoPSF-Filt (c) RwPSF (d) RwPSF-Filt

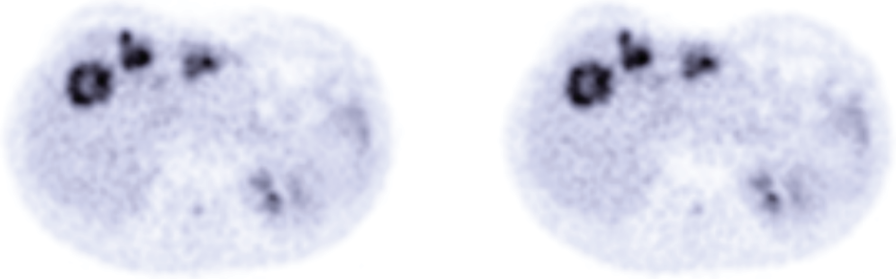


(e) TOF RnoPSF (f) TOF RnoPSF-Filt (g) TOF RwPSF (h) TOF RwPSF-Filt

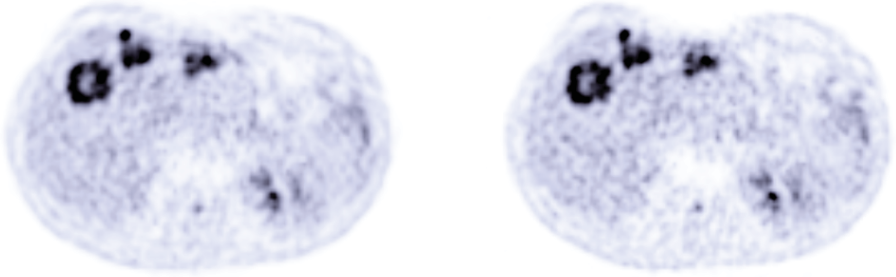
Figure 8.8: Coronal images of oncological patient C, comparison of the different reconstruction algorithms without (top) and with (bottom) TOF information: from left to right, RnoPSF, RnoPSF-Filt, RwPSF, RwPSF-Filt [the images are shown using the same display parameters]



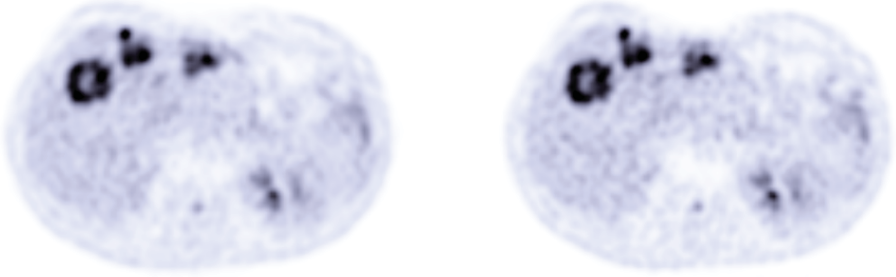
(a) RnoPSF (b) TOF RnoPSF



(c) RnoPSF-Filt (d) TOF RnoPSF-Filt



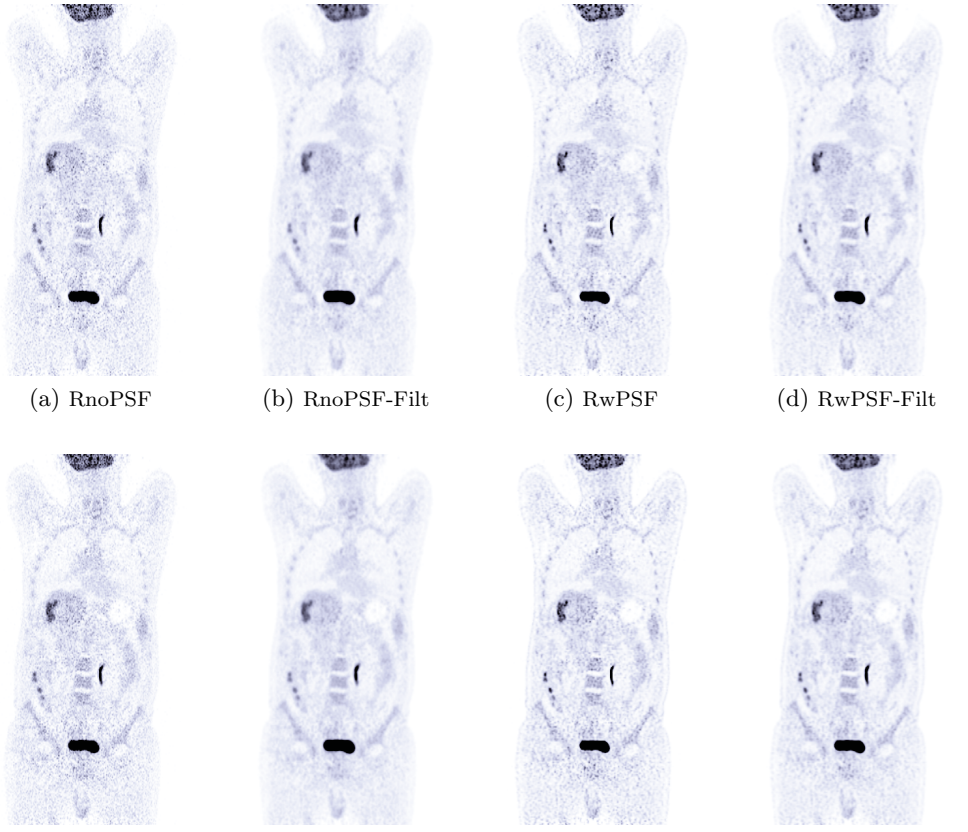
(e) RwPSF (f) TOF RwPSF



(g) RwPSF-Filt (h) TOF RwPSF-Filt

Figure 8.9: Transaxial images of oncological patient C, comparison of the different reconstruction algorithms without (left) and with (right) TOF information: from top to bottom, RnoPSF, RnoPSF-Filt, RwPSF, RwPSF-Filt [the images are shown using the same display parameters]

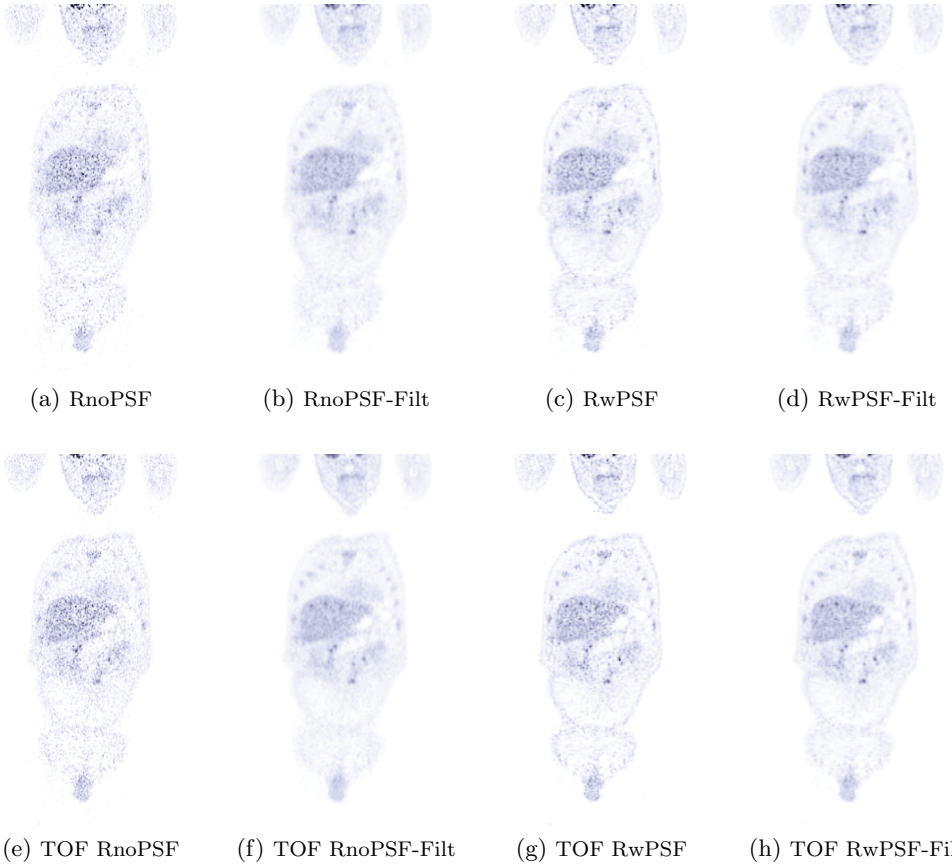
fatter than patient C: consequently, the activity distribution is expected to have larger dimensions, increasing the effect of TOF. The figures confirm a slight improvement in the definition and contrast of the hot regions thanks to the TOF information, in particular for the lesions and structures farthest from the scanner center. The introduction of PSF dramatically increases the spatial resolution and contrast of the lesions, especially the smallest ones. Finally, as already noted, the synergy between TOF and PSF appears to yield the best results, allowing to better discriminate — when compared to RwPSF — the lesions from the surrounding background. Moreover, their combined use reduces the activity in the cold regions, leading to more contrasted images.



(a) RnoPSF (b) RnoPSF-Filt (c) RwPSF (d) RwPSF-Filt

(e) TOF RnoPSF (f) TOF RnoPSF-Filt (g) TOF RwPSF (h) TOF RwPSF-Filt

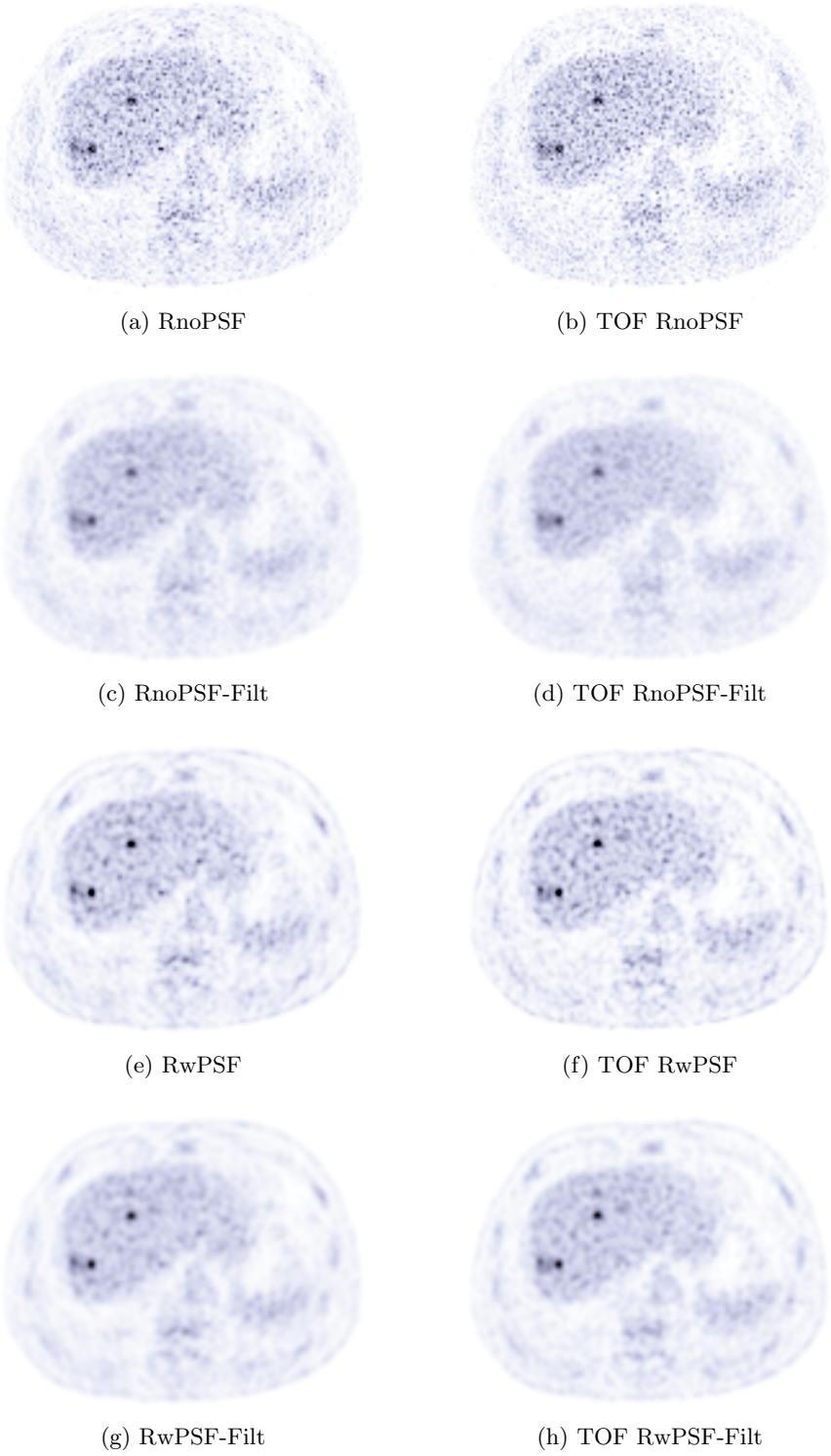
Figure 8.10: Coronal images of oncological patient D, comparison of the different reconstruction algorithms without (top) and with (bottom) TOF information: from left to right, RnoPSF, RnoPSF-Filt, RwPSF, RwPSF-Filt [the images are shown using the same display parameters]



(a) RnoPSF (b) RnoPSF-Filt (c) RwPSF (d) RwPSF-Filt

(e) TOF RnoPSF (f) TOF RnoPSF-Filt (g) TOF RwPSF (h) TOF RwPSF-Filt

Figure 8.11: Coronal images of oncological patient D, comparison of the different reconstruction algorithms without (top) and with (bottom) TOF information: from left to right, RnoPSF, RnoPSF-Filt, RwPSF, RwPSF-Filt [the images are shown using the same display parameters]



(a) RnoPSF (b) TOF RnoPSF

(c) RnoPSF-Filt (d) TOF RnoPSF-Filt

(e) RwPSF (f) TOF RwPSF

(g) RwPSF-Filt (h) TOF RwPSF-Filt

Figure 8.12: Transaxial images of oncological patient D, comparison of the different reconstruction algorithms without (left) and with (right) TOF information: from top to bottom, RnoPSF, RnoPSF-Filt, RwPSF, RwPSF-Filt [the images are shown using the same display parameters]

8.4 Effects of the regularization

8.4.1 Optimization of the regularization parameters

The optimization was performed by reconstructing (including TOF information) an acquisition of a NEMA IEC Body Phantom (acquired statistics: 50.2 Mcounts, see section 8.3.1 for further details) using different sets of parameters (β, δ) and choosing the set which maximized the detectability index D (see section 6.3) relative to the smallest sphere (diameter 10 mm), i.e. a representative sphere of many oncological lesions. Larger lesions would suffer from less spatial resolution loss due to the regularization process, with consequent better results in terms of definition and contrast of the activity distribution.

The optimization was performed using 10 iterations, considered as a compromise between clinical requirements of time and recovery of spatial resolution thanks to the presence of TOF and PSF in the reconstruction algorithm.

In figure 8.13 the detectability for the smallest sphere is plotted as a function of the parameters β and δ for the proposed prior. The maximum detectability is obtained for

$$\boxed{\beta = 0.003, \quad \delta = 0.4} \tag{8.2}$$

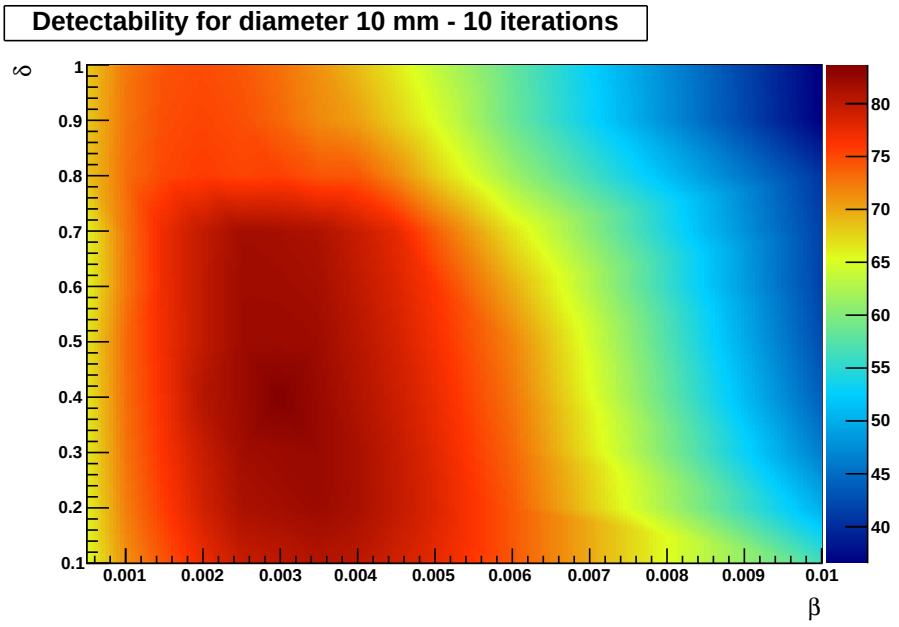


Figure 8.13: Optimization of the regularization parameters for the proposed prior with TOF by maximizing the D index

8.4.2 Quantitative accuracy

The quantitative validation was performed by using the NEMA IEC Body Phantom presented in section 8.3.1, with identical acquisition protocols. The chosen reconstruction parameters were: image matrix 256×256, 18 subsets, reconstruction FOV 50 cm, pixel size 1.95 mm, up to 50 iterations. The clinical post-filter used in TOF RnoPSF was coincident with the one used in the clinical practice, i.e. it was composed of a symmetric two-dimensional Gaussian filter with FWHM=5 mm in the transaxial planes and a three-point mean along the axial direction. The post-filter used in TOF RwPSF-Filt was composed of a symmetric two-dimensional Gaussian filter with FWHM=4 mm in the transaxial planes and a three-point mean along the axial direction. The regularization parameters for RwPSF-R were set equal to the optimal values $(0.003, 0.4)$ found in section 8.4.1. For TOF RwPSF-PR, the parameter β was set equal to the TOF RwPSF-R one, i.e. $\beta = 0.003$.

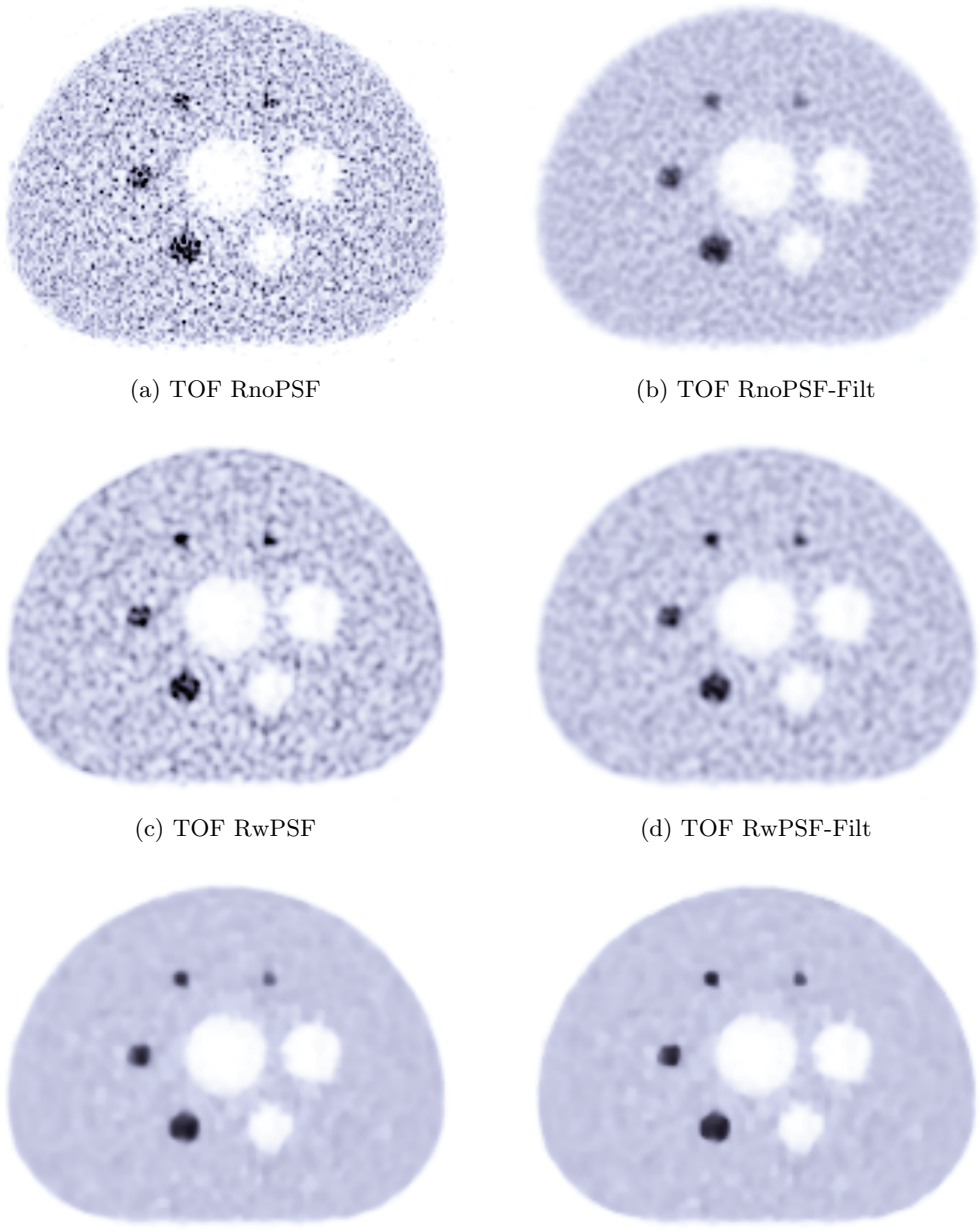
In table 8.3 the percent differences (with respect to TOF RnoPSF reconstructions) of the CR_{cold} and CR_{hot} coefficients, obtained with the different algorithms, for the largest sphere (37-mm diameter) and the smallest sphere (10-mm diameter) respectively and of the COV at 5, 10 and 50 iterations are reported. Similarly, in table 8.4 the percent differences of the contrast recovery coefficients and of the background variability, with respect to the clinical RwPSF-Filt reconstructions, obtained from the different regularized algorithms and from RwPSF are reported. In figure 8.15 the CR_{cold} coefficients for the largest sphere and the CR_{hot} coefficient for the smallest sphere are plotted versus the background coefficient of variation, while in figure 8.14 an example of the reconstructed images using the different algorithms is presented.

From tables 8.3 and 8.4 TOF RwPSF-R confirms — also in presence of the TOF information — that the proposed prior results in better quantification (at 10 iterations, +26.4% for the CR_{hot} on the smallest sphere and +3.2% for the CR_{cold} on the largest sphere) and smoother background (at 10 iterations, −27.6% for the COV) with respect to the clinical (TOF RwPSF-Filt) reconstruction. Even if for the smallest sphere the CR_{hot} results slightly lower than in TOF RnoPSF case, for the largest sphere TOF RwPSF-R converges to a CR_{cold} value very near to the TOF RwPSF one (−0.4%). For what concerns the background noise, the proposed prior reduces the COV with respect to all the reconstruction algorithms, in particular TOF RnoPSF and TOF RwPSF-Filt.

When compared to TOF RwPSF-PR, the proposed modification reveals very little differences in the cold regions; the cold contrast for the largest sphere is characterised by a very modest improvement introduced by TOF RwPSF-R (+0.1% at 5 iterations, +0.1% at 10 iterations and +0.5% at 50 iterations). The hot regions (e.g. the smallest sphere) benefit much more from the proposed modification, with an improvement in the quantification of 11.2% (5 iterations), 16.3%

(10 iterations) and 17.8% (50 iterations). Consequently, the drawbacks of the change in the prior (i.e. a slight increase of background COV — +1.0% at 5 iterations, +2.1% at 10 iterations and +3.2% at 50 iterations) are much less important than the obtained gains. Finally, as stated also in section 7.4.3, the introduction of the regularization accelerates the convergence of the algorithm.

The images (figure 8.14) confirm the numerical results. RwPSF-R efficiently reduces the noise content, preserving a natural appearance and retaining good definition and contrast of the spheres, both in the hot and in the cold cases.



(a) TOF RnoPSF

(b) TOF RnoPSF-Filt

(c) TOF RwPSF

(d) TOF RwPSF-Filt

(e) TOF RwPSF-PR

(f) TOF RwPSF-R

Figure 8.14: NEMA IEC Body Phantom at 10 iterations: comparison of TOF RnoPSF, TOF RnoPSF-Filt, TOF RwPSF, TOF RwPSF-Filt, TOF RwPSF-PR and TOF RwPSF-R [the images are shown using the same display parameters]

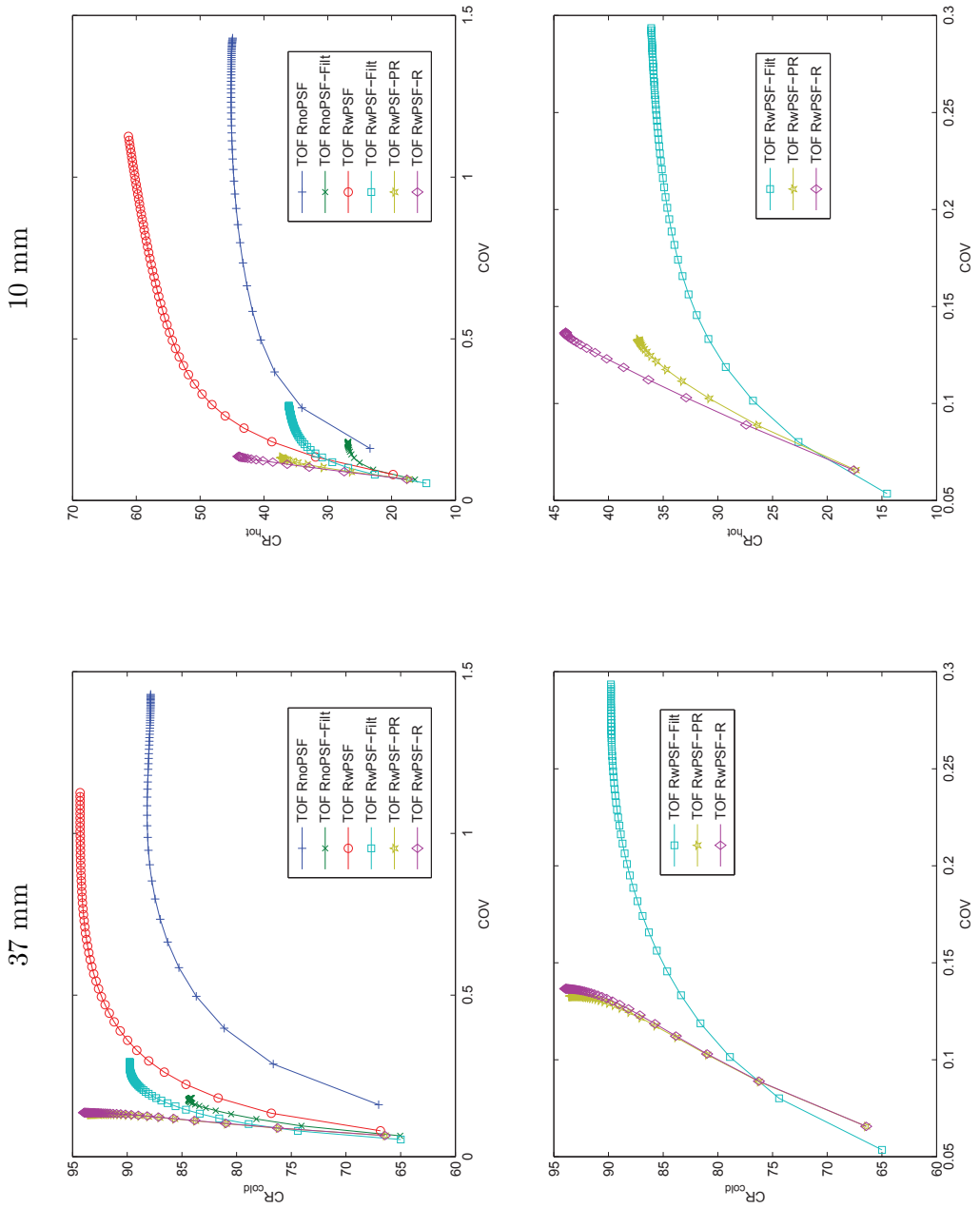


Figure 8.15: NEMA IEC phantom: comparison of the CR_{hot} for the 37-mm diameter sphere (left) and for the 10-mm diameter sphere (right) versus the background coefficient of variation for the reconstruction algorithms considered (top) and TOF RwPSF-Filt and the different regularization priors used (bottom) [each point represents one iteration]

Recon algorithm	**ΔCR_{cold} for the largest sphere**		
	5 it.	10 it.	50 it.
TOF RnoPSF-Filt	-3.9%	-4.2%	-4.3%
TOF RwPSF	+1.6%	+3.7%	+7.3%
TOF RwPSF-Filt	-2.2%	-0.7%	+2.2%
TOF RwPSF-PR	+0.5%	+2.4%	+6.4%
TOF RwPSF-R	+0.6%	+2.5%	+6.9%

Recon algorithm	**ΔCR_{hot} for the smallest sphere**		
	5 it.	10 it.	50 it.
TOF RnoPSF-Filt	-37.1%	-39.6%	-40.3%
TOF RwPSF	+10.2%	+18.7%	+36.4%
TOF RwPSF-Filt	-26.2%	-23.4%	-19.6%
TOF RwPSF-PR	-16.9%	-16.8%	-17.0%
TOF RwPSF-R	-7.6%	-3.2%	-2.2%

Recon algorithm	**ΔCOV**		
	5 it.	10 it.	50 it.
TOF RnoPSF-Filt	-75.6%	-81.4%	-87.3%
TOF RwPSF	-55.3%	-53.8%	-21.3%
TOF RwPSF-Filt	-77.2%	-79.9%	-79.5%
TOF RwPSF-PR	-79.9%	-85.7%	-90.7%
TOF RwPSF-R	-79.7%	-85.4%	-90.4%

Table 8.3: Percent differences of CR_{cold} coefficient for the largest sphere (37-mm diameter), of CR_{hot} coefficient for the smallest sphere (10-mm diameter) and of the background coefficient of variation COV with respect to TOF RnoPSF reconstructions

Recon algorithm	**ΔCR_{cold} for the largest sphere**		
	5 it.	10 it.	50 it.
TOF RwPSF	+3.9%	+4.4%	+5.0%
TOF RwPSF-PR	+2.8%	+3.1%	+4.1%
TOF RwPSF-R	+2.9%	+3.2%	+4.7%

Recon algorithm	**ΔCR_{hot} for the smallest sphere**		
	5 it.	10 it.	50 it.
TOF RwPSF	+49.3%	+54.9%	+69.7%
TOF RwPSF-PR	+12.5%	+8.7%	+3.3%
TOF RwPSF-R	+25.1%	+26.4%	+21.6%

Recon algorithm	**ΔCOV**		
	5 it.	10 it.	50 it.
TOF RwPSF	+96.3%	+129.7%	+283.7%
TOF RwPSF-PR	-11.8%	-28.9%	-54.7%
TOF RwPSF-R	-11.0%	-27.6%	-53.4%

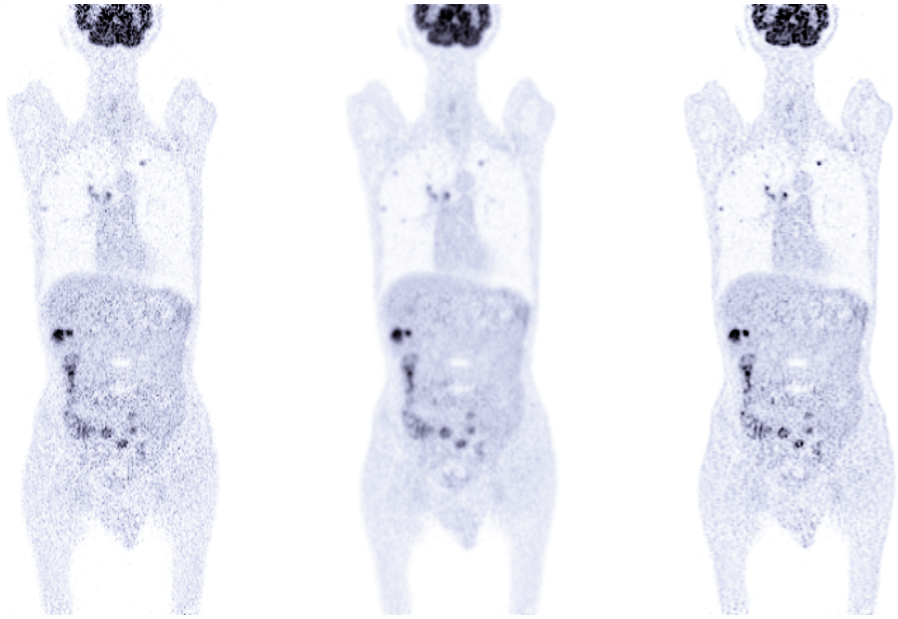
Table 8.4: Percent differences of CR_{cold} coefficient for the largest sphere (37-mm diameter), of CR_{hot} coefficient for the smallest sphere (10-mm diameter) and of the background coefficient of variation COV with respect to TOF RwPSF-Filt reconstructions

8.4.3 Qualitative improvements

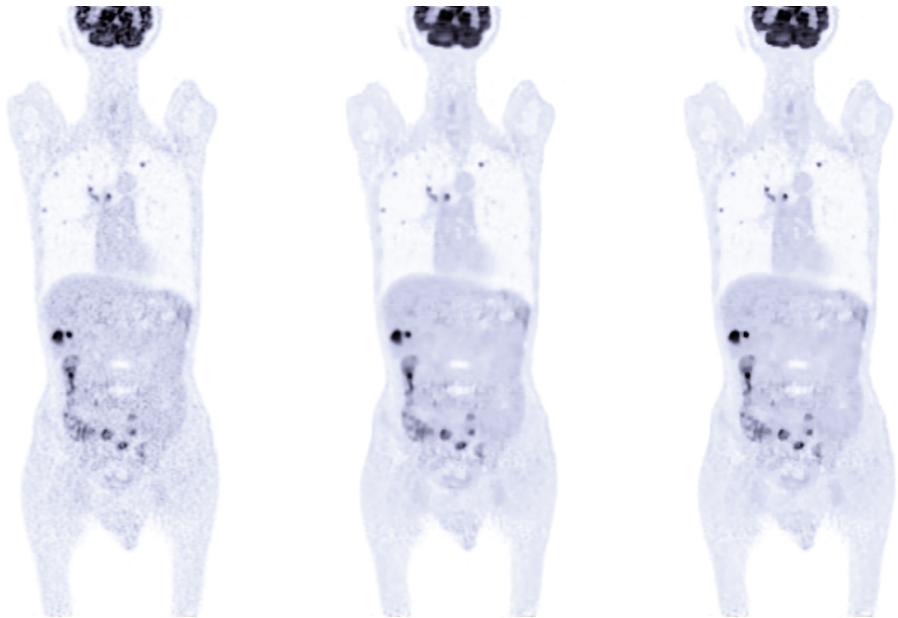
A qualitative assessment of the regularization effect was performed by reconstructing the acquired data of the oncological patients described in section 8.3.2 with the same reconstruction parameters. The regularization parameters for TOF RwPSF-R were set equal to the optimal values $(0.003, 0.4)$ found in section 8.4.1. For TOF RwPSF-PR, the parameter β was set equal to the TOF RwPSF-R one, i.e. $\beta = 0.003$.

As far as patient C is concerned, in figures 8.16 and 8.17 the comparison between the different reconstruction algorithms is performed on some coronal images, while one set of transaxial images is provided in figure 8.18. TOF RwPSF-R provides a very positive control of noise, maintaining good natural appearance and high enough spatial resolution and contrast, lower only when compared to TOF RwPSF (even if, in this case, with much better background uniformity). TOF RwPSF-R increases also the definition of non active regions, such as the external layer of the thighbones and the intervertebral disks, improving also the global quality of the image. TOF RwPSF-PR generates similar images to TOF RwPSF-R, with slightly less noisy images, but also with lower spatial resolution and definition of the details.

For what concerns patient D, in figures 8.19 and 8.20 the comparison between the different reconstruction algorithms is performed on a set of coronal images, while some transaxial images are provided in figure 8.21. The coronal images show how the proposed prior produces a very effective control of the noise, but maintains natural appearance, good contrast and spatial resolution, also for the smallest structures. Moreover, the global image quality and the definition of the activity distribution appear better when compared to TOF wPSF-Filt. Finally, also for patient D TOF RwPSF-PR generates similar images to TOF RwPSF-R, with the same differences as stated for patient C.

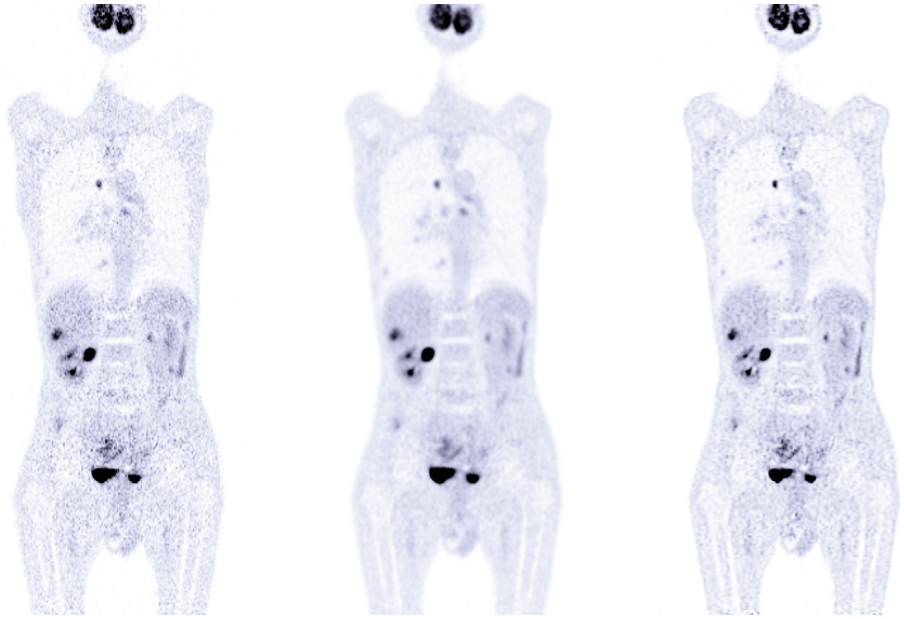


(a) TOF RnoPSF (b) TOF RnoPSF-Filt (c) TOF RwPSF

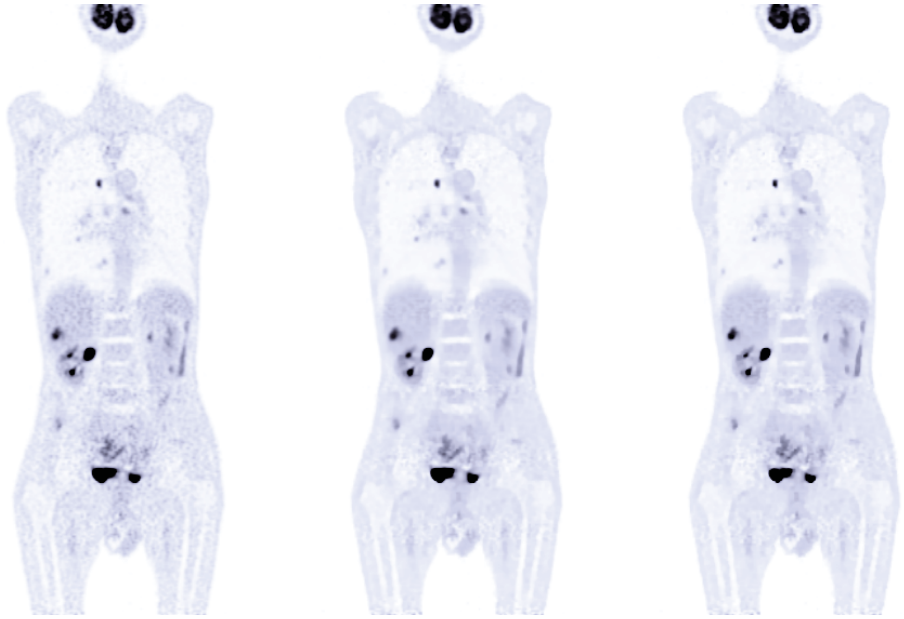


(d) TOF RwPSF-Filt (e) TOF RwPSF-PR (f) TOF RwPSF-R

Figure 8.16: Coronal images of oncological patient C, comparison of the different reconstruction algorithms [the images are shown using the same display parameters]

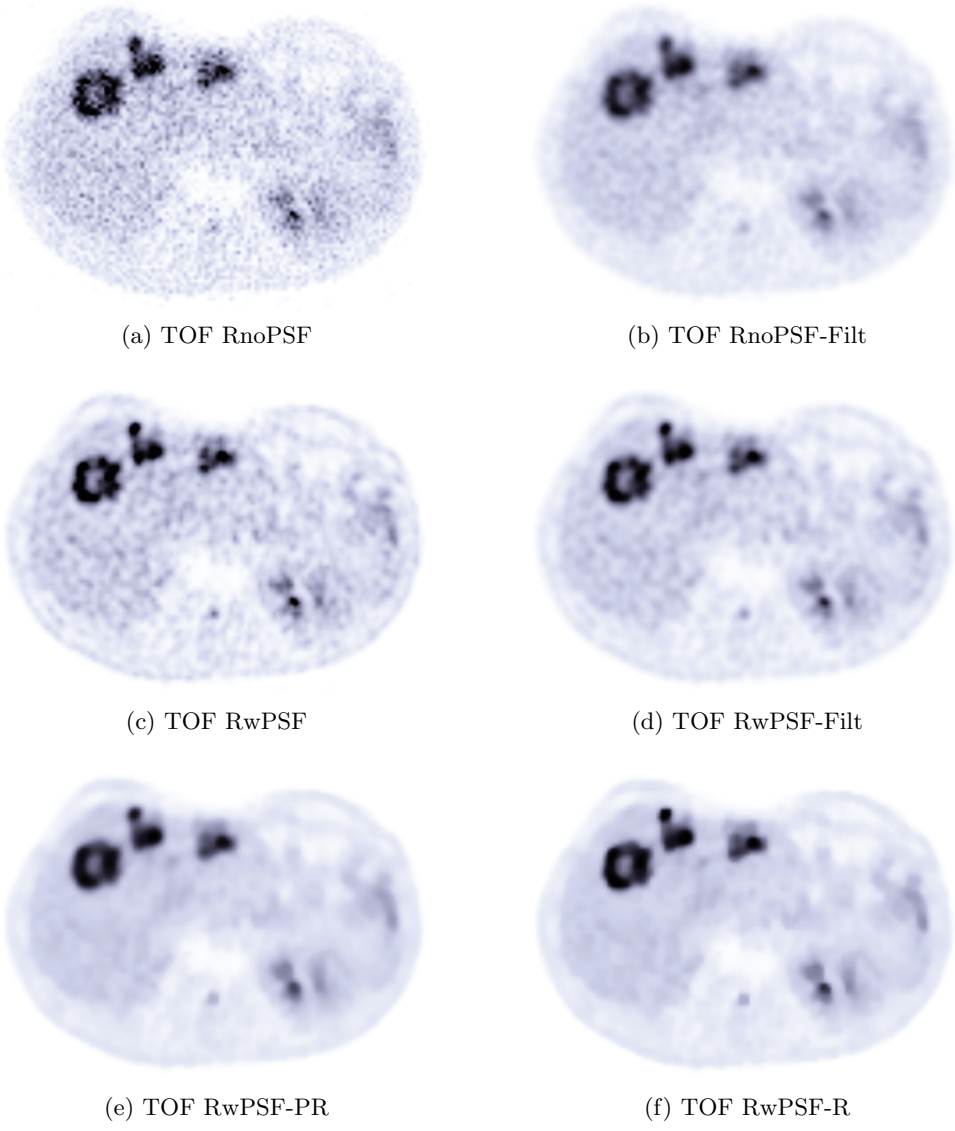


(a) TOF RnoPSF (b) TOF RnoPSF-Filt (c) TOF RwPSF



(d) TOF RwPSF-Filt (e) TOF RwPSF-PR (f) TOF RwPSF-R

Figure 8.17: Coronal images of oncological patient C, comparison of the different reconstruction algorithms [the images are shown using the same display parameters]



(a) TOF RnoPSF

(b) TOF RnoPSF-Filt

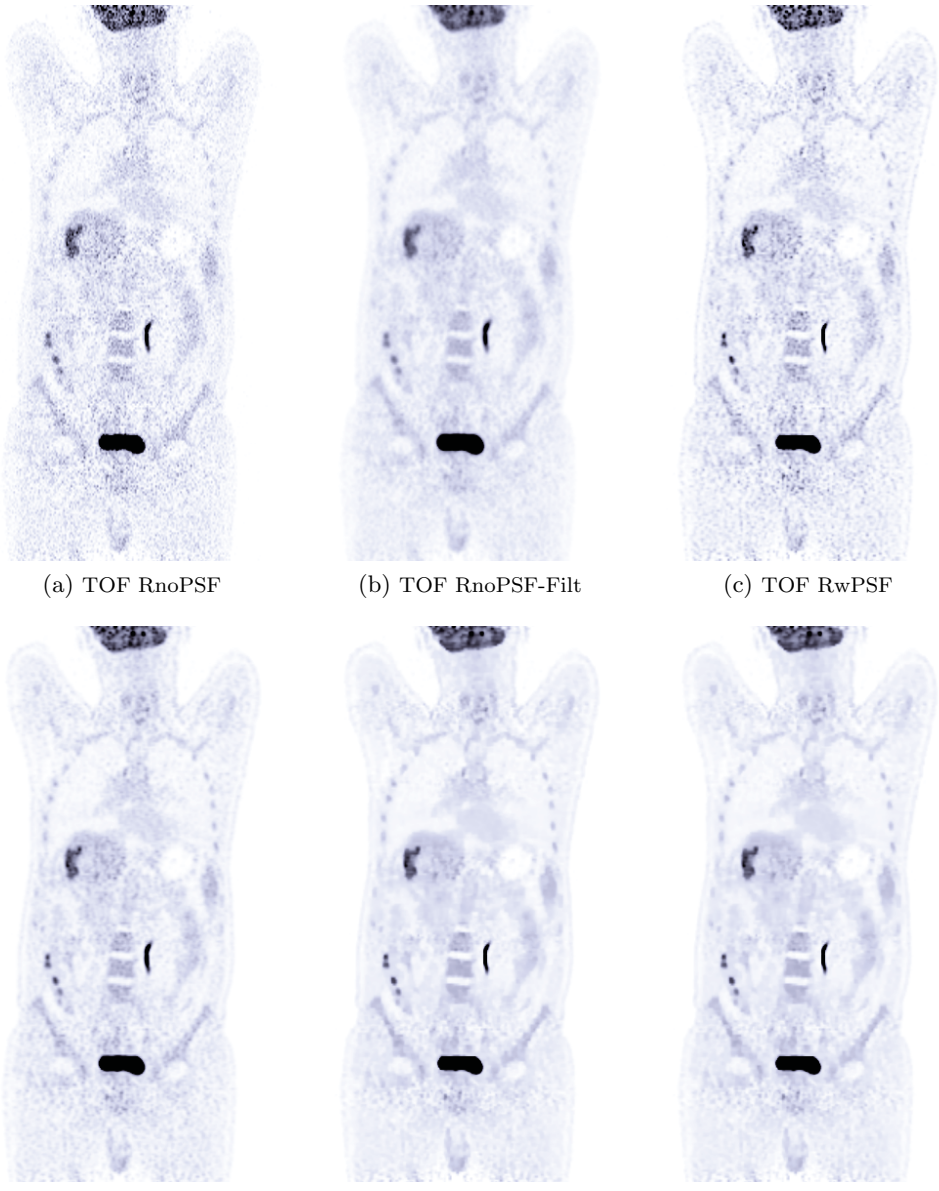
(c) TOF RwPSF

(d) TOF RwPSF-Filt

(e) TOF RwPSF-PR

(f) TOF RwPSF-R

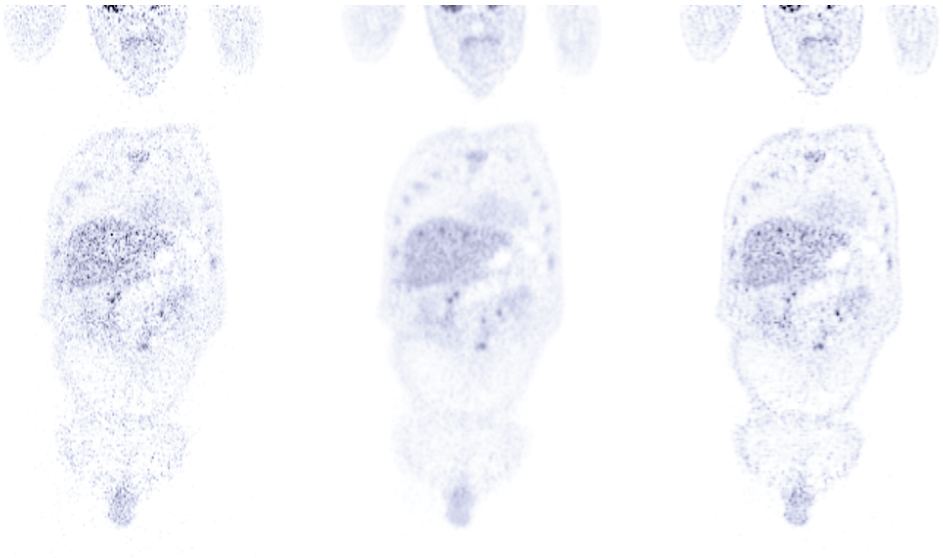
Figure 8.18: Transaxial images of oncological patient C, comparison of the different reconstruction algorithms [the images are shown using the same display parameters]



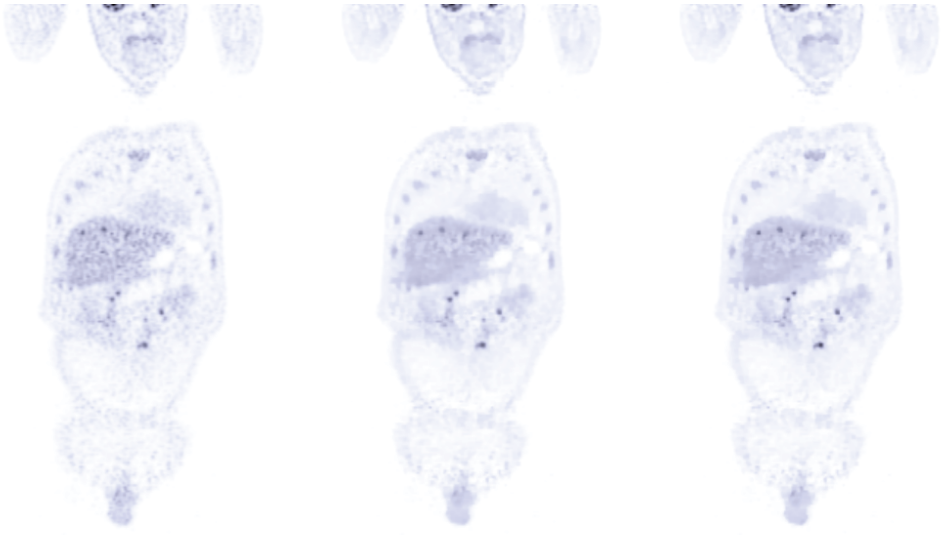
(a) TOF RnoPSF (b) TOF RnoPSF-Filt (c) TOF RwPSF

(d) TOF RwPSF-Filt (e) TOF RwPSF-PR (f) TOF RwPSF-R

Figure 8.19: Coronal images of oncological patient D, comparison of the different reconstruction algorithms [the images are shown using the same display parameters]

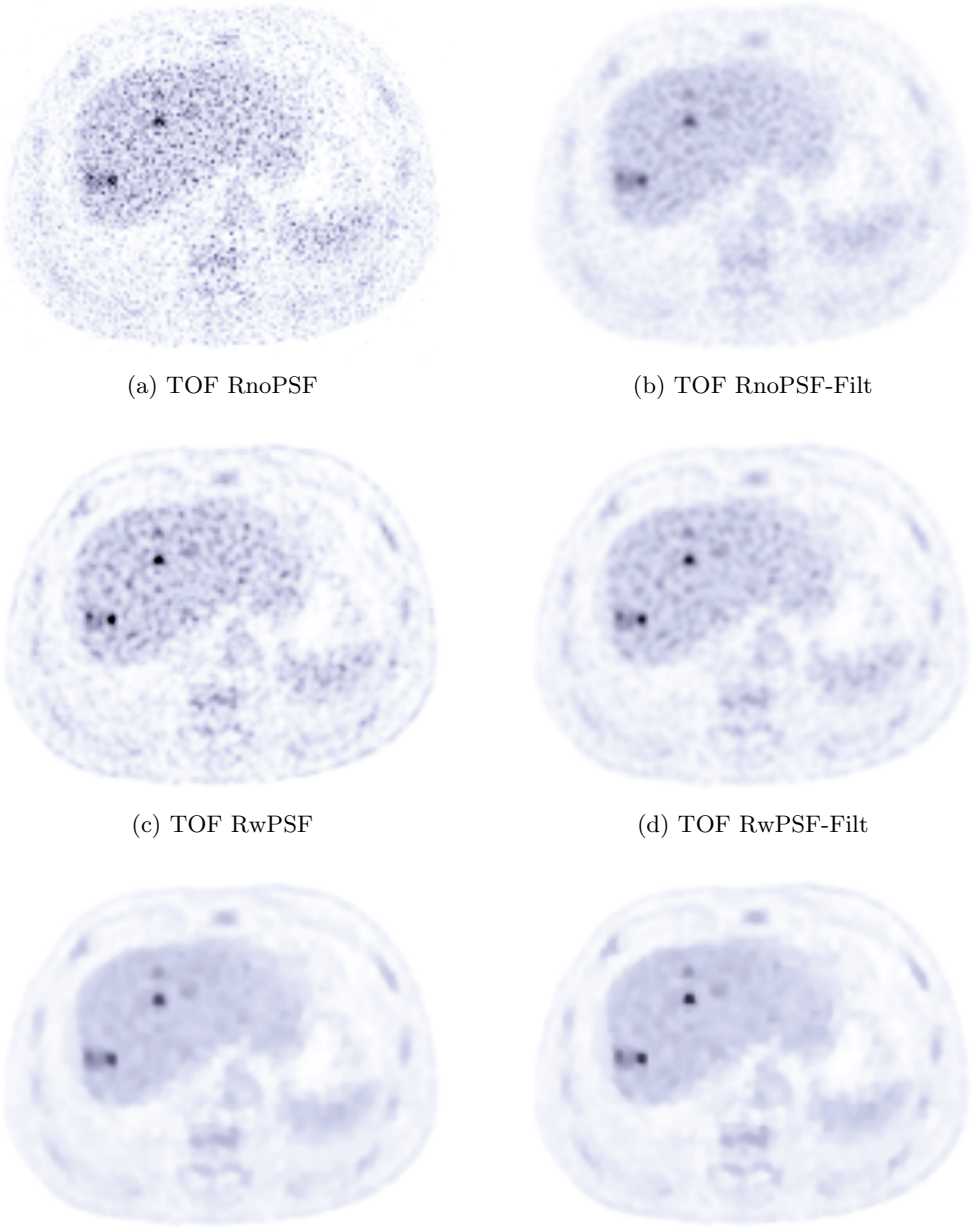


(a) TOF RnoPSF (b) TOF RnoPSF-Filt (c) TOF RwPSF



(d) TOF RwPSF-Filt (e) TOF RwPSF-PR (f) TOF RwPSF-R

Figure 8.20: Coronal images of oncological patient D, comparison of the different reconstruction algorithms [the images are shown using the same display parameters]



(a) TOF RnoPSF

(b) TOF RnoPSF-Filt

(c) TOF RwPSF

(d) TOF RwPSF-Filt

(e) TOF RwPSF-PR

(f) TOF RwPSF-R

Figure 8.21: Transaxial images of oncological patient D, comparison of the different reconstruction algorithms [the images are shown using the same display parameters]

Part IV

Conclusions

Chapter 9

Discussion

Positron Emission Tomography and, more recently, hybrid PET/CT systems have become increasingly important in the clinical practice thanks to their capability to provide complementary, spatially coregistered anatomical and functional information. However, PET imaging is still characterized by a "poor" spatial resolution due to the multiple physical effects linked with the 511 keV photon-matter interaction and with the detection apparatus. Low spatial resolution leads to a spread in the activity distribution, resulting in a degradation of the image quality (e.g. reduction of the contrast), in an underestimation of the tracer uptake (activity concentration) and in an overestimation of the lesion volume. In principle, each effect could be estimated by experimental measurements or by simulation studies and then included in the reconstruction scheme to be compensated. Unfortunately, some of these effects are very difficult or nearly impossible to be measured. Consequently, another approach may consist in accounting for all these effects by a "global" Point Spread Function (PSF), which describes how a point source is rendered by the system itself. The knowledge of the PSF, in principle, allows the correction of the resolution degradation and the recovery of the correct quantitative information; on the other hand, this approach requires the knowledge of the PSF in each point of the PET FOV since the response of the system is not uniform across the FOV of the scanner.

Several methods have been proposed to estimate the PSF, which can be generally categorized into one of the following classes: Monte Carlo simulations [34], analytical modelization [78, 79] and experimental measurements [4, 70, 97]. Once the PSF is known, different reconstruction methods have been proposed to take it into account in the reconstruction process, based on two different approaches, differentiating on the level at which the PSF is introduced: the sinogram approach (which assumes that the resolution leakage happens in the space of raw data) and the image approach (which models the PSF acting on the image to be recon-

structed). The first approach, more used up to now (e.g. [4, 70, 78]), evaluates the resolution spread acting on the acquired data, without any additional factor coming from the reconstruction process. Anyway, this is strictly dependent on the width of the sinogram bins and thus not linked with the chosen reconstruction FOV. On the other hand, if the PSF is taken into account at the image level (e.g. [83, 91]), its evaluation depends on the reconstruction algorithm used, but this method is simpler and allows a higher match between the pixel size and the precision in the PSF modelling. Both methods can only approximate the actual resolution model, which is complex due to the superposition of several different effects. Moreover, some of these effects are characteristic of the sinogram level (e.g. the detector blurring), while others are more relative to the image level (e.g. the positron range). On the other hand, from a theoretical point of view the two strategies should convey the same information about the response of the tomograph, especially in cases where the system response varies slowly across the FOV. Nevertheless, the implementation of the PSF in the image space appears to be more suitable for list–mode data, which is the basic acquisition modality used for TOF acquisition, as well as for 4D PET, i.e. a modality in which the acquired data, subject to some cyclic effects (usually the breathing process, even if also the cardiac functionality has been studied in the very last years), are divided into different *phases* of the cycle and are separately reconstructed in order to "freeze" the movements introduced by the effects. This suggests that the image–space approach might be more useful in the future developments of PET.

In this thesis it has been proposed a spatially variant PSF implementation in the image space of a 3D OSEM algorithm. Two different scanners from General Electric Medical Systems were considered, without (DSTE) and with (D690) TOF information. The PSF was chosen to be a 3D Gaussian function, which — thanks to the cylindrical symmetry of the scanners under study — was factorized into a transaxial two-dimensional Gaussian function and an axial one-dimensional Gaussian function. Consequently, in each point of the FOV the knowledge of the PSF coincides with the knowledge of four spread parameters, the *internal radial*, the *external radial*, the *tangential* and the *axial* ones.

Experimental measurements, therefore, were performed to determine the dependences of the PSF spread parameters on the position inside the scanner FOV. The way of measuring the response of the PET system is, in fact, an important point. Even in the simplest case, i.e. using a small point source in different positions of the scanner FOV, several factors have then to be taken into account to obtain a good and representative response of the PET system such as: dimension of the source, type of isotope, media surrounding the point source (air, water, warm radioactive background), number and position of the measurements to account for the spatial detector configuration [29, 70].

In the case here presented the cylindrical symmetry of the scanners suggested

considering dependences on the radial and axial distances from the scanner centre. To measure the PSF of the system a small radioactive ^{22}Na point source (a cylinder with diameter and height of 1 mm encapsulated in Lucite) in air was used. As demonstrated in [2] this type of source resembles well the characteristics of a ^{18}F point source in water, which is the most frequent condition in clinical practice. Other authors have used a line source in a warm radioactive background to measure the spatial resolution of a PET system [29]. This approach is interesting, in particular if a point source is considered instead of a line source, since it could resemble more accurately the response of a system in a "clinical" condition. Nevertheless, this way of measuring the PSF of the system is much more complex and the specific experimental conditions (e.g. ratio of the activity between warm background and point source, dimensions and shape of the point source container) and the following data processing (e.g. scatter correction, attenuation correction, estimation of the warm background etc.) should be carefully evaluated, as they could play a very important role, particularly in the estimation of the asymmetric tails of the PSF functions over the entire FOV of the PET system.

After establishing the source to be used for the PSF estimation, the number of point measurements is also another important factor. Not having a robotic system (as used e.g. in [70]), approximately 400 single measurements uniformly distributed within the scanner FOV were performed per scanner. Finally, since the PSF should act at the image level of an OSEM algorithm, it was considered important to determine the trend of the PSF widths in the same conditions, to match, as much as possible, the PSF evaluations and the algorithm and, so, the same algorithm was used to reconstruct the point source raw data.

Furthermore, it is known that the results obtained with iterative algorithms depend on the number of iterations used for the reconstruction of the data and this is particularly true when the activity distribution to be reconstructed is complex (e.g. whole body). On the other hand the convergence for a point source in air is fast: the number of total iterations (i.e. the product of the number of subsets and number of iterations) chosen (280 and 180, respectively for DSTE and D690) appeared to be adequate, since even increasing the number of iterations no significant variation in the numerical results of the resulting model was found (results not shown).

Moreover, the results from ^{22}Na acquisitions are equivalent to the ones obtainable with ^{18}F-FDG, as demonstrated in [2]. If another isotope (significantly different from ^{18}F for the positron emission energy) were used, a new set of measurements should be acquired to account for the contribution of the positron range specific of that isotope in the "global" PSF.

From each reconstructed image three orthogonal two-dimensional planes, passing through the voxel with the maximum intensity, were extracted along three

privileged directions (originating the radial-tangential, radial-axial and tangential-axial planes): each of them was then fitted by a two-dimensional function to determine the corresponding spread parameters.

Before fitting, the reconstructed images of the acquired point sources were post-filtered with a 2D Gaussian function (FWHM=5 mm in the transaxial planes) to reduce the noise and fluctuations in their peaks which otherwise would have made the fitting results unreliable to estimate the spread parameters. The post-filter did not invalidate the results of the fit, however, since its FWHM was included into the fitting function. Moreover, the fitting function took into account also the *actual* position of the point source (in order to correct the unavoidable imprecisions in the source positioning), the source dimensions (to avoid an overestimation of the spread parameters due to the approximation of point source) and the intrinsic discretization along the axial direction due to the finite dimension of slices.

The entire set of spread results obtained with ^{22}Na were then fitted with a two-dimensional function of the radial and axial distances from the scanner center to determine the required analytical dependences.

A validation of the proposed method of measurement showed its good accuracy in the building of the PSF model, justifying its use.

The implementation of the PSF consisted in a redefinition of the projector and backprojector of the OSEM algorithm. The practical implementation has been performed by a factorization of the 3D-PSF into a transaxial 2D PSF and an axial 1D PSF. Both the transaxial and the axial PSFs were stored into a three-dimensional matrix; the continuous model of the PSF has been discretized by calculating its *integral* for each voxel, allowing for a better adaptive implementation for each specific reconstruction FOV and pixel size. The much more common strategy of filling each voxel with the value of PSF in the middle point yields different shapes of the kernel, leading to possible inaccuracies. The dimension of the stored kernel was chosen about 4 times the spatial resolution of the scanners: this choice is conservative with respect to the possibility of generating artefacts due to the truncation of the PSF kernel.

The PSF was tested on some phantom and clinical data. The results showed improved quantitative accuracy, spatial resolution and image quality, with better results for higher number of iterations. The combined use of TOF and PSF appeared to allow them to take advantage of each other, leading to the best results. This is probably because TOF improves the SNR as a consequence of the more accurate events positioning but it does not improve the spatial resolution, which is the main goal of PSF. Thus, if TOF acts as an “accelerator” for the convergence of the signal, PSF can recover a better signal at a lower number of iterations, introducing, at the same time, a “filtering” effect.

Unfortunately, a common effect of iterative reconstruction techniques is the

increase of noise as iterations proceed, due to the ill-posed nature of the reconstruction problem [103]. Usually in clinical practice the quality of the images is privileged over their quantitative accuracy by stopping the iterative algorithm after few iterations and, consequently, far from convergence. This is true all the more if PSF is included in the algorithm, since the speed of convergence results lower than in non PSF algorithms [96]. Another important effect observed in PSF-based reconstructions is the enhancement of regions with sharp intensity transitions: this effect is recognizable throughout the literature (e.g. [10, 73, 83, 89]). In this thesis it was demonstrated, by means of 1D and 2D simulations, that it is strongly related to the PSF implementation and, even in presence of a perfectly matched kernel, unavoidable unless an unpractical number of iterations is used.

Regularization techniques have been demonstrated to be useful for taking noise under control during the reconstruction and improving the benefits from the use of the PSF information by increasing the number of iterations used. In particular, in this thesis a Bayesian variational regularization strategy [36, 84] has been tested and employed.

Two important topics should be taken into account when dealing with regularization techniques:

- since many regularization strategies exist (and each of them leads to different characteristics on the image), it is important to understand which effects are desired, in order to choose the "best" regularization strategy for the particular scope under study;

- since every regularization strategy depend on one or more parameters, it is useful to set some optimization criteria to select their values in order to obtain the best results.

As far as the first issue is concerned, in PET it is usually desirable to suppress the noise (to obtain smooth, uniform background regions) while retaining as much spatial resolution as possible, all the more if the PSF information is taken into account. Two good candidates are the Huber (or Gauss-Total Variation) [44, 50] and the generalized p-Gaussian [15] priors. The former provides good preservation of spatial resolution thanks to the TV component for high gradients, but the Gaussian component for low gradients might be insufficient in controlling very noisy environments (as often encountered in PET, in particular when the number of iterations is increased to exploit the PSF action) unless the regularization strength is set to a very high value, obtaining an unnatural reconstructed image. The p-Gaussian prior provides a very strong smoothing on background regions, resulting in good noise control, while it smoothes much less in signal regions. In this thesis a modification of the p-Gaussian prior was proposed to maintain the smoothing effect for low gradients (i.e. in background regions) and to reduce the

spatial resolution loss, while retaining "natural" transitions and appearance in the image. A 3D OSEM algorithm has been modified to include the proposed prior using a MAP OSL multiplicative approach. The value for the parameter p was chosen by qualitatively evaluating the results obtained with different values and choosing the one yielding the best compromise between noise suppression, preservation of spatial resolution and natural appearance of the image.

The second issue to be studied deals with the optimization criterion for the regularization parameters to be used. In this thesis a figure of merit, taking into account both the qualitative and the quantitative content, was proposed to evaluate the global "detectability" of a lesion. The validation of this detectability index showed a very good correlation with the human response and, thus, justified its use to set the regularization parameters.

The priors needed two regularization parameters: the regularization strength β (which controls the effect of the regularization strategy on the final image) and the signal-background threshold δ (which distinguishes the different behaviours to be applied in the different regions of the image). Among them, the latter parameter is, in principle, more intricate and delicate to set, since different regions in the same study might require different thresholds to maximize the general spatial resolution preservation. One conservative strategy may consist in setting the parameter for the worst conditions expected (e.g. for the smallest lesions and the lowest contrasts), assuming that larger lesions and/or higher contrasts lead to lower resolution loss introduced by the regularization.

The regularization parameters were determined by maximizing the detectability index for each prior. This optimization was performed for a sphere with diameter 10 mm and 10 OSEM iterations. The maximization results hinted a different behaviour of the two priors: the GTV prior showed a very narrow peak along the direction of the δ parameter, suggesting very different behaviours for small changes of this parameter and, consequently, indicating that setting this parameter could be a delicate process. For the proposed prior, small changes of the δ parameter around the maximum lead to very little differences in its behaviour, guaranteeing higher stability and "margins of security".

The validation of the proposed modifications was quantitative on data acquired with a NEMA IEC Body Phantom and qualitative on data relative to two oncological patients and consisted of a comparison between the standard reconstruction algorithms, the proposed algorithm, the results obtained with the p-Gaussian prior and with Gauss-Total Variation. This comparison showed an effective control of noise (but with natural appearance of the image) by the proposed prior with a contemporary good preservation of spatial resolution, contrast and definition of the activity distribution. Moreover, the proposed prior was shown to be able also to take the edge artefact under control, drastically reducing the overshoots originating at large transitions in the image.

Chapter 10

Future perspectives

The fitting procedure appears to be consistent and useful in determining the spread parameters. Like every fitting procedure, its reliability would probably benefit from larger datasets available. This could be obtained by reducing the reconstruction FOV and/or increasing the size of the matrix (e.g. 512×512). In case of a future improvement in the precision of the projectors and of the entire reconstruction process (in addition to an increased computational performance on the reconstruction console) the spread parameters might be determined with even higher precision and accuracy. In principle, the proposed fitting procedure should be already adequate to support (thanks to the inclusion of the source dimensions in the fit function itself) the reduced validity of the point source approximation due to the reduced size of the voxels.

The value used for the p-Gaussian prior ($p = 4/3$) — i.e. the starting point for the proposed modifications — resulted in the best compromise between preservation of spatial resolution, noise suppression and natural appearance of the clinical image. Whether this particular choice of the parameter gives the "globally" best results (i.e. considering all possible choices for the different parameters) is beyond the scope of this work. Anyway, the proposed strategy to modify the prior is independent of the parameter p and, thus, in the case of a different p this strategy is expected to remain valid.

It is important to underline also that the priors were optimized considering 10 OSEM iterations as a limit for the image reconstruction in clinically acceptable times. If, in the future and/or for different clinical requirements, the number of iterations could be increased, different regularization parameters and consequent numerical results would be probably obtained and the comparison between the reconstruction algorithms might yield different results. However, this goes beyond the scope of this work, which is based on the present situation of computational

power and was focused on the most used (and general) application of PET, i.e. oncology.

In other fields, e.g. neurology, the (even small) unavoidable spatial resolution loss introduced by the proposed prior might be excessive. On the other hand, in neurology the characteristics of the image (and, as a consequence, the requests for the prior) are very different to the oncology case: for example, the activity distribution is usually distributed on smaller dimensions, with generally lower noise. The less stringent constraint of noise control and, contemporarily, the need of lower loss of spatial resolution might require a completely different approach, but could be obtained also by simply reducing the β strength parameter. More studies on this topic might provide further hints.

The emissive regularization strategy here presented demonstrated good noise control and image quality. In some applications (of which neurology might be the most significative) the regularization characteristics could be more efficiently determined on an *anatomical* (or *morphological*) basis. The implementation of an anatomical regularization, either alone or in combination with the emissive one, could improve the definition of the structures and allow a better differentiation and definition of the effects to be applied on the image. Nevertheless, the use of an anatomical a-priori information could introduce a bias in the reconstructed images, since the functional information would be influenced by the a-priori assumptions; moreover, small (and unavoidable) mismatches between the anatomical and functional data would generally create artefacts, unless ad-hoc strategies are developed. Finally, a combined anatomical-functional prior would increase the complexity of the regularization strategy, e.g. for what concerns the set-up of the corresponding parameters and the evaluations of their respective effects.

Mathematics

M.1 Poisson statistics

Let X be a random variable following a Poisson distribution with mean λ. The probability of obtaining the value x is consequently

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

If n *independent* random variables $X_1, X_2, \ldots, X_n$ are considered, the joint probability of obtaining respectively the values $x_1, x_2, \ldots, x_n$ is the product of n Poisson distributions:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!}\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}\ldots\frac{e^{-\lambda_n}\lambda_n^{x_n}}{x_n!}$$

where λ_i is the mean relative to the variable X_i.

Now consider the variable $Y = X_1 + X_2$. The probability of obtaining the value y is then

$$P(Y = X_1 + X_2 = y) = \sum_{i=0}^{y} P(X_1 = i, X_2 = y - i)$$

since the different cases $(X_1 = i, X_2 = y - i)$ are independent and mutually exclusive. Expliciting the summation content,

$$\begin{aligned} P(Y = y) &= \sum_{i=0}^{y} \frac{e^{-\lambda_1}\lambda_1^i}{i!}\frac{e^{-\lambda_2}\lambda_2^{y-i}}{(y-i)!} = \frac{e^{-\lambda_1}e^{-\lambda_2}}{y!}\sum_{i=0}^{y}\frac{y!}{i!(y-i)!}\lambda_1^i\lambda_2^{y-i} = \\ &= \frac{e^{-\lambda_1}e^{-\lambda_2}}{y!}\sum_{i=0}^{y}\binom{y}{i}\lambda_1^i\lambda_2^{y-i} = \frac{e^{-(\lambda_1+\lambda_2)}}{y!}(\lambda_1+\lambda_2)^y = \frac{e^{-\lambda}\lambda^y}{y!} \end{aligned} \tag{M.1}$$

where $\lambda = \lambda_1 + \lambda_2$. Consequently, the sum of two Poisson distributions is a Poisson distribution, too, having as mean the sum of the means of the two starting distribution. The generalization to n Poisson distributions is then immediate showing that the sum of $n - 1$ Poisson distributions is a Poisson distribution

itself. In addition, if the variables $\tilde{X}_i = a_i X_i$ (where $a_i < 1$ is a multiplicative constant) are considered, the variable $\tilde{Y} = \sum_i a_i X_i$ describes a thinning of a Poisson process and therefore it follows a Poisson distribution [102] with mean value $\lambda = \sum_i a_i \lambda_i$. If the substitutions

$$\begin{cases} \tilde{Y} &= y_d \\ X_i &= n_b \\ a_i &= p_{bd} \\ \tilde{X}_i &= x_{bd} \end{cases}$$

are performed, the parallelisms with the EM algorithm are apparent.

The expectation value of X_1 conditioned to $Y = X_1 + X_2 = y$ can be written as

$$E\left[X_1 \left| Y = y\right.\right] = \sum_{x_1=0}^{y} x_1 P\left(X_1 = x_1 \left| Y = y\right.\right)$$

where

$$\begin{aligned} P\left(X_1 = x_1 \left| Y = y\right.\right) = \frac{P\left(X_1 = x_1, Y = y\right)}{P(Y = y)} = \frac{P\left(X_1 = x_1, X_2 = y - x_1\right)}{P(Y = y)} = \\ = \frac{\dfrac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!}\dfrac{e^{-\lambda_2}\lambda_2^{y-x_1}}{(y-x_1)!}}{\dfrac{e^{-(\lambda_1+\lambda_2)}}{y!}\left(\lambda_1+\lambda_2\right)^y} = \binom{y}{x_1}\frac{\lambda_1^{x_1}\lambda_2^{y-x_1}}{(\lambda_1+\lambda_2)^y} \end{aligned} \tag{M.2}$$

Consequently

$$\begin{aligned} E\left[X_1 \left| Y = y\right.\right] = \sum_{x_1=0}^{y} x_1 P\left(X_1 = x_1 \left| Y = y\right.\right) = \sum_{x_1=1}^{y} x_1 P\left(X_1 = x_1 \left| Y = y\right.\right) = \\ = \sum_{x_1=1}^{y} \frac{\lambda_1^{x_1}\lambda_2^{y-1-(x_1-1)}}{(\lambda_1+\lambda_2)^y}\frac{y!}{(x_1-1)!\,(y-x_1)!} = \\ = \frac{y\lambda_1}{(\lambda_1+\lambda_2)^y}\sum_{x_1'=0}^{y-1}\frac{\lambda_1^{x_1'}\lambda_2^{y-1-x_1'}(y-1)!}{x_1'!\,(y-1-x_1')!} = \frac{y\lambda_1}{(\lambda_1+\lambda_2)^y}\left(\lambda_1+\lambda_2\right)^{y-1} = \\ = y\frac{\lambda_1}{\lambda_1+\lambda_2} \end{aligned} \tag{M.3}$$

Considering the general case $Y = \sum_{i=1}^{N} X_i$, it is possible to introduce the auxiliary random variable $\tilde{X}_2 = \sum_{i=2}^{N} X_i$ and write $Y = X_1 + \tilde{X}_2$: the above statements

are then appliable in a similar way, obtaining

$$E\left[X_i \middle| Y = \sum_{i=1}^{N} X_i = y\right] = y\frac{\lambda_i}{\sum_{i=1}^{N}\lambda_i}$$

With the same substitutions presented above, the application to the expectation step of the MLEM algorithms (equation 3.11) is evident.

M.2 Point and non–point sources with Gaussian PSF

Consider two regular functions $f, g : \mathbb{R}^n \to \mathbb{R}$. The convolution operation $f \star g$ is defined as

$$(f \star g)(\vec{x}) = \int_{\mathbb{R}^n} f\left(\vec{h}\right) g\left(\vec{x} - \vec{h}\right) d\vec{h}$$

As a first step, consider the convolution between two *symmetric* monodimensional gaussians

$$f(x) = Ae^{-\frac{(x-b)^2}{2\sigma_1^2}} \qquad g(x) = Be^{-\frac{x^2}{2\sigma_2^2}}$$

Their convolution is

$$\begin{aligned}(f \star g)(x) &= \int_{-\infty}^{+\infty} Ae^{-\frac{(t'-b)^2}{2\sigma_1^2}} Be^{-\frac{(x-t')^2}{2\sigma_2^2}} dt' \overset{t=t'-b}{=} \int_{-\infty}^{+\infty} Ae^{-\frac{t^2}{2\sigma_1^2}} Be^{-\frac{(x-b-t)^2}{2\sigma_2^2}} dt \overset{y=x-b}{=} \\ &= AB\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2\sigma_1^2}} e^{-\frac{(y-t)^2}{2\sigma_2^2}} dt\end{aligned} \tag{M.4}$$

The sum in the exponential can be rearranged as follows:

$$\begin{aligned}\frac{t^2}{\sigma_1^2} + \frac{(y-t)^2}{\sigma_2^2} &= \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)t^2 - \frac{2y}{\sigma_2^2}t + \frac{y^2}{\sigma_2^2} = \frac{1}{\sigma_2^2}\left[\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2}t^2 - 2yt + y^2\right] = \\ &= \frac{1}{\sigma_2^2}\left[\left(\frac{\tilde{\sigma}}{\sigma_1}t - \frac{\sigma_1}{\tilde{\sigma}}y\right)^2 + \frac{\sigma_2^2 y^2}{\tilde{\sigma}^2}\right]\end{aligned}$$

where $\tilde{\sigma}^2 = \sigma_1^2 + \sigma_2^2$ has been introduced. Therefore, equation M.4 can be rewritten as

$$(f \star g)(x) = ABe^{-\frac{y^2}{2\tilde{\sigma}^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{\tilde{\sigma}}{\sigma_1\sigma_2}t - \frac{\sigma_1}{\tilde{\sigma}\sigma_2}y\right)^2} dt = AB\frac{\sigma_1\sigma_2}{\tilde{\sigma}}e^{-\frac{y^2}{2\tilde{\sigma}^2}} \underbrace{\int_{-\infty}^{+\infty} e^{-z^2/2}dz}_{\sqrt{2\pi}}$$

in which the substitution $z = \frac{\tilde{\sigma}}{\sigma_1\sigma_2}t - \frac{\sigma_1}{\tilde{\sigma}\sigma_2}y$ has been performed. In conclusion, then,

$$\left(Ae^{-\frac{(x-b)^2}{2\sigma_1^2}}\right) \star \left(Be^{-\frac{x^2}{2\sigma_2^2}}\right) = \sqrt{2\pi}AB\frac{\sigma_1\sigma_2}{\tilde{\sigma}}e^{-\frac{(x-b)^2}{2\tilde{\sigma}^2}} \tag{M.5}$$

Secondly, consider the convolution between two monodimensional gaussians, a symmetric one and an *asymmetric* one

$$f(x) = A\left[\theta(b-x)e^{-\frac{(x-b)^2}{2\sigma_l^2}} + \theta(x-b)e^{-\frac{(x-b)^2}{2\sigma_r^2}}\right] \qquad g(x) = Be^{-\frac{x^2}{2\sigma_c^2}}$$

where

$$\theta(t) = \begin{cases} 1 & \text{if } t > 0 \\ 1/2 & \text{if } t = 0 \\ 0 & \text{if } t < 0 \end{cases} \tag{M.6}$$

Following the same scheme of the previous example, their convolution is

$$\begin{aligned}(f \star g)(x) &= AB\left[\int_{-\infty}^{0} e^{-\frac{t^2}{2\sigma_l^2}} e^{-\frac{(y-t)^2}{2\sigma_c^2}} dt + \int_{0}^{+\infty} e^{-\frac{t^2}{2\sigma_r^2}} e^{-\frac{(y-t)^2}{2\sigma_c^2}} dt\right] = \\ &= AB\left[e^{-\frac{y^2}{2\tilde{\sigma}_l^2}} \int_{-\infty}^{0} e^{-\frac{1}{2\sigma_c^2}\left(\frac{\tilde{\sigma}_l}{\sigma_l}t - \frac{\sigma_l}{\tilde{\sigma}_l}y\right)^2} dt + e^{-\frac{y^2}{2\tilde{\sigma}_r^2}} \int_{0}^{+\infty} e^{-\frac{1}{2\sigma_c^2}\left(\frac{\tilde{\sigma}_r}{\sigma_r}t - \frac{\sigma_r}{\tilde{\sigma}_r}y\right)^2} dt\right]\end{aligned} \tag{M.7}$$

where, similarly to before, the symbols $\tilde{\sigma}_l^2 = \sigma_l^2 + \sigma_c^2$ and $\tilde{\sigma}_r^2 = \sigma_r^2 + \sigma_c^2$ have been used. Now, performing the substitutions

$$z = \frac{1}{\sqrt{2}\sigma_c}\left(\frac{\tilde{\sigma}_l}{\sigma_l}t - \frac{\sigma_l}{\tilde{\sigma}_l}y\right) \qquad w = \frac{1}{\sqrt{2}\sigma_c}\left(\frac{\tilde{\sigma}_r}{\sigma_r}t - \frac{\sigma_r}{\tilde{\sigma}_r}y\right)$$

equation M.7 changes into

$$(f \star g)(x) = AB\sqrt{2}\sigma_c\left[\frac{\sigma_l}{\tilde{\sigma}_l}e^{-\frac{y^2}{2\tilde{\sigma}_l^2}} \int_{-\infty}^{-\frac{1}{\sqrt{2}}\frac{\sigma_l}{\tilde{\sigma}_l\sigma_c}y} e^{-z^2} dz + \frac{\sigma_r}{\tilde{\sigma}_r}e^{-\frac{y^2}{2\tilde{\sigma}_r^2}} \int_{-\frac{1}{\sqrt{2}}\frac{\sigma_r}{\tilde{\sigma}_r\sigma_c}y}^{+\infty} e^{-w^2} dw\right]$$

Introducing the functions

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2} dt \qquad \operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_x^{+\infty} e^{-t^2} dt$$

the convolution can finally be written as

$$\begin{aligned}(f \star g)(x) = AB\frac{\sqrt{2\pi}\sigma_c}{2}\Bigg\{&\frac{\sigma_l}{\tilde{\sigma}_l}e^{-\frac{(x-b)^2}{2\tilde{\sigma}_l^2}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}}\frac{\sigma_l}{\tilde{\sigma}_l\sigma_c}(x-b)\right] + \\ &+\frac{\sigma_r}{\tilde{\sigma}_r}e^{-\frac{(x-b)^2}{2\tilde{\sigma}_r^2}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}}\frac{\sigma_r}{\tilde{\sigma}_r\sigma_c}(x-b)\right]\Bigg\}\end{aligned} \tag{M.8}$$

The last preliminary step is to calculate the convolution of two *two-dimensional* elliptical gaussians

$$f(x) = Ae^{-\frac{(x-a)^2}{2\sigma_a^2} - \frac{(y-b)^2}{2\sigma_b^2}} \qquad g(x) = Be^{-\frac{x^2}{2\sigma_1^2} - \frac{y^2}{2\sigma_2^2}}$$

An important point is the absence of mixed terms xy, i.e. the ellipses in the exponential have their axes *parallel* to the coordinate axes. The required convolution is

$$\begin{aligned}(f \star g)(x) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} ABe^{-\frac{(t-a)^2}{2\sigma_a^2}-\frac{(s-b)^2}{2\sigma_b^2}} e^{-\frac{(x-t)^2}{2\sigma_1^2}-\frac{(y-s)^2}{2\sigma_2^2}}\,dtds = \\ &= AB\left(\int_{-\infty}^{+\infty} e^{-\frac{(t-a)^2}{2\sigma_a^2}} e^{-\frac{(x-t)^2}{2\sigma_1^2}}\,dt\right)\left(\int_{-\infty}^{+\infty} e^{-\frac{(s-b)^2}{2\sigma_b^2}} e^{-\frac{(y-s)^2}{2\sigma_2^2}}\,ds\right)\end{aligned}$$

Therefore, the two-dimensional convolution — if no mixed terms are present in the product of the functions — is expressable as the product of two monodimensional convolutions (the statement is valid also for N–dimensional convolutions without mixed variables). Using equation M.5, then,

$$\left(Ae^{-\frac{(x-a)^2}{2\sigma_a^2}-\frac{(y-b)^2}{2\sigma_b^2}}\right)\star\left(Be^{-\frac{x^2}{2\sigma_1^2}-\frac{y^2}{2\sigma_2^2}}\right) = 2\pi AB\frac{\sigma_a\sigma_b\sigma_1\sigma_2}{\tilde{\sigma}_a\tilde{\sigma}_b}e^{-\frac{(x-a)^2}{2\tilde{\sigma}_a^2}-\frac{(y-b)^2}{2\tilde{\sigma}_b^2}} \quad \text{(M.9)}$$

where the symbols $\tilde{\sigma}_a^2 = \sigma_a^2 + \sigma_1^2$ and $\tilde{\sigma}_b^2 = \sigma_b^2 + \sigma_2^2$ have been used once again.

Now consider the physical case of a point source

$$S = A_s\delta^{(3)}(x-\mu_x, y-\mu_y, z-\mu_z)$$

where A_s is the activity of the source. This point source will be spread by the PSF

$$\begin{aligned}PSF_r(x) &= A_r\left[\theta(-x)e^{-\frac{x^2}{2\sigma_l^2}} + \theta(x)e^{-\frac{x^2}{2\sigma_r^2}}\right] \\ PSF_t(y) &= A_t e^{-\frac{y^2}{2\sigma_t^2}} \\ PSF_a(z) &= A_a e^{-\frac{z^2}{2\sigma_a^2}}\end{aligned}$$

$$\begin{aligned}PSF(x,y,z) &= PSF_r(x)PSF_t(y)PSF_a(z) = \\ &= A_rA_tA_a\left[\theta(-x)e^{-\frac{x^2}{2\sigma_l^2}} + \theta(x)e^{-\frac{x^2}{2\sigma_r^2}}\right]e^{-\frac{y^2}{2\sigma_t^2}}e^{-\frac{z^2}{2\sigma_a^2}} \quad \text{(M.10)}\end{aligned}$$

PSF must have unitary integral, since the counts cannot be created or destroyed:

$$\iiint_{\mathbb{R}^3} PSF(x,y,z)\,dx\,dy\,dz \stackrel{!}{=} 1$$

$$\underbrace{\left[\int_{-\infty}^{+\infty} PSF_r(x)\,dx\right]}_{A_r\sqrt{2\pi}(\sigma_l+\sigma_r)/2}\underbrace{\left[\int_{-\infty}^{+\infty} PSF_t(y)\,dy\right]}_{A_t\sqrt{2\pi}\sigma_t}\underbrace{\left[\int_{-\infty}^{+\infty} PSF_a(z)\,dz\right]}_{A_a\sqrt{2\pi}\sigma_a} \stackrel{!}{=} 1$$

$$\Longrightarrow A_r A_t A_a = \frac{2}{(2\pi)^{3/2}\,\sigma_a\sigma_t\,(\sigma_l+\sigma_r)} = k \tag{M.11}$$

Now introduce a *post–filter*

$$PF = \frac{1}{2\pi\sigma_c^2} e^{-\frac{x^2+y^2}{2\sigma_c^2}} \frac{1}{\sqrt{2\pi}\sigma_{ca}} e^{-\frac{z^2}{2\sigma_{ca}^2}}$$

where z refers to the scanner axis and (x, y) define transaxial planes. The post–filter, then, consists in a transaxially symmetric two-dimensional gaussian of sigma σ_c and an axially monodimensional gaussian* of sigma σ_{ca}. The filter has already been written adequately to have its space integral normalized to one.

The result of the physical measure M can then be schematized as

$$M = (S \star PSF) \star PF = A_s\, PSF(x-\mu_x, y-\mu_y, z-\mu_z) \star PF(x,y,z)$$

where the fundamental property of the δ distribution has been used in the last step. Substituting, then,

$$M = K \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left[\theta(\mu_x - t) e^{-\frac{(t-\mu_x)^2}{2\sigma_l^2}} + \theta(t-\mu_x) e^{-\frac{(t-\mu_x)^2}{2\sigma_r^2}}\right] e^{-\frac{(s-\mu_y)^2}{2\sigma_t^2}} e^{-\frac{(u-\mu_z)^2}{2\sigma_a^2}} \cdot$$
$$\cdot\, e^{-\frac{(x-t)^2+(y-s)^2}{2\sigma_c^2}} e^{-\frac{(z-u)^2}{2\sigma_{ca}^2}}\, dt\, ds\, du \tag{M.12}$$

with

$$K = \frac{2A_s}{(2\pi)^3\sigma_c^2\sigma_{ca}\sigma_a\sigma_t\,(\sigma_l+\sigma_r)}$$

Equation M.12 can be factorized in the product of the three monodimensional integrals. The results obtained above yield

$$\int_{-\infty}^{+\infty} \left[\theta(\mu_x - t) e^{-\frac{(t-\mu_x)^2}{2\sigma_l^2}} + \theta(t-\mu_x) e^{-\frac{(t-\mu_x)^2}{2\sigma_r^2}}\right] e^{-\frac{(x-t)^2}{2\sigma_c^2}}\, dt =$$
$$= \frac{\sqrt{2\pi}\sigma_c}{2}\left\{\frac{\sigma_l}{\tilde\sigma_l} e^{-\frac{(x-\mu_x)^2}{2\tilde\sigma_l^2}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}}\frac{\sigma_l}{\tilde\sigma_l\sigma_c}(x-\mu_x)\right] + \right.$$
$$\left. + \frac{\sigma_r}{\tilde\sigma_r} e^{-\frac{(x-\mu_x)^2}{2\tilde\sigma_r^2}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}}\frac{\sigma_r}{\tilde\sigma_r\sigma_c}(x-\mu_x)\right]\right\}$$

$$\int_{-\infty}^{+\infty} e^{-\frac{(s-\mu_y)^2}{2\sigma_t^2}} e^{-\frac{(y-s)^2}{2\sigma_c^2}}\, ds = \sqrt{2\pi}\frac{\sigma_t\sigma_c}{\tilde\sigma_t} e^{-\frac{(y-\mu_y)^2}{2\tilde\sigma_t^2}}$$

*Even if in this thesis the filter along z is not applied, the result obtained with $\sigma_{ca} = 0$ ($\tilde\sigma_a = \sigma_a$) is the same obtained with no axial filter at all.

$$\int_{-\infty}^{+\infty} e^{-\frac{(u-\mu_z)^2}{2\sigma_a^2}} e^{-\frac{(z-u)^2}{2\sigma_{ca}^2}}\, du = \sqrt{2\pi}\frac{\sigma_a\sigma_{ca}}{\tilde{\sigma}_a} e^{-\frac{(z-\mu_z)^2}{2\tilde{\sigma}_a^2}}$$

The physical measure with a point source can be expressed, then, as

$$M = \frac{A_s}{\sqrt{(2\pi)^3}\,(\sigma_l+\sigma_r)\,\tilde{\sigma}_a\tilde{\sigma}_t}\left\{\frac{\sigma_l}{\tilde{\sigma}_l} e^{-\frac{(x-\mu_x)^2}{2\tilde{\sigma}_l^2}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}}\frac{\sigma_l}{\tilde{\sigma}_l\sigma_c}(x-\mu_x)\right] + \right.$$
$$\left. + \frac{\sigma_r}{\tilde{\sigma}_r} e^{-\frac{(x-\mu_x)^2}{2\tilde{\sigma}_r^2}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}}\frac{\sigma_r}{\tilde{\sigma}_r\sigma_c}(x-\mu_x)\right]\right\} e^{-\frac{(y-\mu_y)^2}{2\tilde{\sigma}_t^2}}\, e^{-\frac{(z-\mu_z)^2}{2\tilde{\sigma}_a^2}} \quad \text{(M.13)}$$

or, splitting the formula into its components ($M = A_s M_r M_t M_a$),

$$M_r(x-\mu_x) = \frac{1}{\sqrt{2\pi}\,(\sigma_l+\sigma_r)}\left\{\frac{\sigma_l}{\tilde{\sigma}_l}\, e^{-\frac{(x-\mu_x)^2}{2\tilde{\sigma}_l^2}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}}\frac{\sigma_l}{\tilde{\sigma}_l\sigma_c}(x-\mu_x)\right] + \right.$$
$$\left. + \frac{\sigma_r}{\tilde{\sigma}_r} e^{-\frac{(x-\mu_x)^2}{2\tilde{\sigma}_r^2}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}}\frac{\sigma_r}{\tilde{\sigma}_r\sigma_c}(x-\mu_x)\right]\right\}$$
$$M_t(y-\mu_y) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_t} e^{-\frac{(y-\mu_y)^2}{2\tilde{\sigma}_t^2}}$$
$$M_a(z-\mu_z) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_a} e^{-\frac{(z-\mu_z)^2}{2\tilde{\sigma}_a^2}} \quad \text{(M.14)}$$

If a non-point source is used, the function describing the spatial extent of the source has to be convolved with the postfiltered PSF response of the point source[†] (as defined in equation M.14).

In this work a cylindrical source, with axis parallel to the tangential direction, is used (figure M.1). A cylinder[‡] with radius R and height H, centered in (μ_x, μ_y, μ_z), can be described by

$$\xi(x,y,z) = \begin{cases} (x-\mu_x)^2 + (z-\mu_z)^2 \le R^2 \\ |y-\mu_y| \le H/2 \end{cases}$$

or, using the function $\theta(t)$ defined in expression M.6,

$$\xi(x,y,z) = \theta\left(R^2 - (x-\mu_x)^2 + (z-\mu_z)^2\right)\left[\theta\left(y-\mu_y+\frac{H}{2}\right) - \theta\left(y-\mu_y-\frac{H}{2}\right)\right]$$

[†]Probably, the most immediate (even if algebraically more complex) sequence would be to first convolve the finite source with the PSF (this represents the natural resolution loss by the scanner and the reconstruction algorithm) and then to postfilter the resulting function. It is important to recall, however, that the convolution operation is commutative and associative: this guarantees the invariance of the final result on the order of application of the different steps.

[‡]For the sake of simplicity, in order to not increase the complexity of the notation, the *activity* in the cylinder is assumed to be unitary (or, in the general case, included in a global multiplicative constant); therefore, it will not be indicated hereafter.

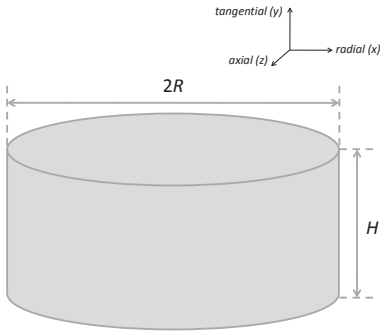


Figure M.1: Cylindrical source with axis parallel to the tangential (y) direction

The distribution in the final image would then be represented by

$$M' = (\xi \star PSF) \star PF = \xi \star (PSF \star PF) = \xi \star (M_r M_t M_a) \tag{M.15}$$

Each plane (radial-tangential, radial-axial, tangential-axial) may then be analitically described as follows.

Radial-tangential plane

Since the planes pass through the pixel with the maximum intensity (which, along the axial direction, is likely to be the center of the cylinder along the axial direction, being the diameter of the cylinder much smaller than the spacing between slices), both along the radial and tangential direction the profile is box-shaped and the two contributions are independent each other:

$$\begin{aligned} \xi_r(x) &= \theta(x - \mu_x + R) - \theta(x - \mu_x - R) \\ \xi_t(y) &= \theta(y - \mu_y + H/2) - \theta(y - \mu_y - H/2) \end{aligned}$$

The object in the radial-tangential plane is described by

$$\begin{aligned} P_{rt} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} M_r(t) M_t(s) \xi_r(x-t) \xi_t(y-s)\, dt\, ds = \\ &= \left(\int_{-\infty}^{+\infty} M_r(t) \xi_r(x-t)\, dt \right) \left(\int_{-\infty}^{+\infty} M_t(s) \xi_t(y-s)\, ds \right) = P_r(x) P_t(y) \end{aligned}$$

where the terms $P_r(x)$ and $P_t(y)$ can be calculated by expliciting the integrals:

$$\begin{aligned} P_r(x) &= \int_{-\infty}^{+\infty} M_r(t) \xi_r(x-t)\, dt = \int_{-\infty}^{+\infty} M_r(x-t) \xi_r(t)\, dt = \\ &= \int_{-R}^{R} M_r(x - \mu_x - t)\, dt \end{aligned}$$

$$P_r(x) = \frac{1}{\sqrt{2\pi}\,(\sigma_l + \sigma_r)} \int_{-R}^{R} \left\{ \frac{\sigma_l}{\tilde{\sigma}_l} e^{-\frac{(x-\mu_x-t)^2}{2\tilde{\sigma}_l^2}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}} \frac{\sigma_l}{\tilde{\sigma}_l \sigma_c} (x - \mu_x - t) \right] + \right.$$
$$\left. + \frac{\sigma_r}{\tilde{\sigma}_r} e^{-\frac{(x-\mu_x-t)^2}{2\tilde{\sigma}_r^2}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}} \frac{\sigma_r}{\tilde{\sigma}_r \sigma_c} (x - \mu_x - t) \right] \right\} dt \quad \text{(M.16)}$$

$$P_t(y) = \int_{-\infty}^{+\infty} M_t(s) \xi_r(y - s)\, ds = \int_{-H/2}^{H/2} M_t(y - \mu_y - s)\, ds$$

$$\begin{aligned} P_t(y) &= \frac{1}{\sqrt{2\pi}\tilde{\sigma}_t} \int_{-H/2}^{H/2} e^{-\frac{(y-\mu_y-s)^2}{2\tilde{\sigma}_t^2}}\, ds = \\ &= \frac{1}{2} \left[\operatorname{erf}\left(\frac{y - \mu_y + H/2}{\sqrt{2}\tilde{\sigma}_t} \right) - \operatorname{erf}\left(\frac{y - \mu_y - H/2}{\sqrt{2}\tilde{\sigma}_t} \right) \right] \end{aligned} \quad \text{(M.17)}$$

Radial-axial plane

In the radial-axial plane the radioactive source is described by a circle[§]:

$$\xi_{ra} = \theta\left(R^2 - (x - \mu_x)^2 + (z - \mu_z)^2\right) \quad \text{(M.18)}$$

In the radial-axial plane the object can be described as

$$\begin{aligned} P_{ra} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} M_r(t) M_a(s) \xi_{ra}(x - t, z - s)\, dt\, ds \\ &= \int_{-R}^{R} M_r(x - \mu_x - t) \left(\int_{-\sqrt{R^2-t^2}}^{\sqrt{R^2-t^2}} M_a(z - \mu_z - s)\, ds \right) dt = \\ &= \frac{1}{2} \int_{-R}^{R} M_r(x - \mu_x - t) \left[\operatorname{erf}\left(\frac{z - \mu_z + \sqrt{R^2 - t^2}}{\sqrt{2}\tilde{\sigma}_a} \right) - \right. \\ &\qquad \left. - \operatorname{erf}\left(\frac{z - \mu_z - \sqrt{R^2 - t^2}}{\sqrt{2}\tilde{\sigma}_a} \right) \right] dt \end{aligned}$$

[§]In the expressions presented hereafter a monometric Cartesian system of coordinates is assumed. In the scanner, however, the spacing along the axial direction is different from the spacing between pixels in the transaxial plane; consequently, in the resulting functions a scaling along the z axis — expression 4.2 — is necessary.

$$P_{ra} = \frac{1}{2\sqrt{2\pi}\,(\sigma_l + \sigma_r)} \int_{-R}^{R} \left\{ \frac{\sigma_l}{\tilde{\sigma}_l} e^{-\frac{(x-\mu_x-t)^2}{2\tilde{\sigma}_l^2}} \operatorname{erfc}\left[\frac{1}{\sqrt{2}} \frac{\sigma_l}{\tilde{\sigma}_l \sigma_c} (x - \mu_x - t) \right] + \right.$$
$$\left. + \frac{\sigma_r}{\tilde{\sigma}_r} e^{-\frac{(x-\mu_x-t)^2}{2\tilde{\sigma}_r^2}} \operatorname{erfc}\left[-\frac{1}{\sqrt{2}} \frac{\sigma_r}{\tilde{\sigma}_r \sigma_c} (x - \mu_x - t) \right] \right\} \cdot$$
$$\cdot \left[\operatorname{erf}\left(\frac{z - \mu_z + \sqrt{R^2 - t^2}}{\sqrt{2}\tilde{\sigma}_a} \right) - \operatorname{erf}\left(\frac{z - \mu_z - \sqrt{R^2 - t^2}}{\sqrt{2}\tilde{\sigma}_a} \right) \right] dt \quad \text{(M.19)}$$

Tangential-axial plane

With the same arguments presented for the radial-tangential plane, it is possible to assume that both along tangential and axial direction the profile is box-shaped and the two contributions are independent each other:

$$\begin{aligned} \xi_t(y) &= \theta(y - \mu_y + H/2) - \theta(y - \mu_y - H/2) \\ \xi_a(z) &= \theta(z - \mu_z + R) - \theta(z - \mu_z - R) \end{aligned}$$

The object in the tangential-axial plane can therefore be described by

$$P_{ta} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} M_t(s) M_a(t) \xi_t(y - s) \xi_a(z - t)\, dt\, ds =$$
$$= \left(\int_{-\infty}^{+\infty} M_t(s) \xi_t(y - s)\, ds \right) \left(\int_{-\infty}^{+\infty} M_a(t) \xi_a(z - t)\, dt \right) = P_t(y) P_a(z)$$

with $P_t(y)$ defined as in equation M.17 and

$$P_a(z) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_a} \int_{-R}^{R} e^{-\frac{(z-\mu_z-t)^2}{2\tilde{\sigma}_a^2}} dt = \frac{1}{2} \left[\operatorname{erf}\left(\frac{z - \mu_z + R}{\sqrt{2}\tilde{\sigma}_a} \right) - \operatorname{erf}\left(\frac{z - \mu_z - R}{\sqrt{2}\tilde{\sigma}_a} \right) \right] \quad \text{(M.20)}$$

M.3 The convolution operator

Since the image λ is described using a Cartesian coordinate system (x, y, z) with origin coincident with the scanner centre, the transaxial PSF — characterized by a circular symmetry and naturally described using a cylindrical coordinate system — should be referred to a Cartesian system, too (the axial component of the PSF is already referred to z axis, assuming for the PSF a cylindrical symmetry). A rotated elliptical, radially asymmetric, spatially-variant 2D Gaussian function

centred in (t,u) is described as

$$PSF(x,y|t,u) = \frac{1}{\pi\left[\sigma_i(t,u)+\sigma_e(t,u)\right]}\cdot \\ \cdot\left[\theta(-d)e^{-\left[A_1(x-t)^2+B_1(x-t)(y-u)+C_1(y-u)^2\right]}+ \right. \\ \left. +\theta(d)e^{-\left[A_2(x-t)^2+B_2(x-t)(y-u)+C_2(y-u)^2\right]}\right] \quad \text{(M.21)}$$

where

$$d = xt + yu - \left(t^2+u^2\right)$$

$$A_1 = \frac{\cos^2\varphi}{\left[\sigma_i\left(t,u\right)\right]^2} + \frac{\sin^2\varphi}{\left[\sigma_t\left(t,u\right)\right]^2} \qquad A_2 = \frac{\cos^2\varphi}{\left[\sigma_e\left(t,u\right)\right]^2} + \frac{\sin^2\varphi}{\left[\sigma_t\left(t,u\right)\right]^2}$$

$$B_1 = \frac{2\sin\varphi\cos\varphi}{\left[\sigma_i\left(t,u\right)\right]^2} - \frac{2\sin\varphi\cos\varphi}{\left[\sigma_t\left(t,u\right)\right]^2} \qquad B_2 = \frac{2\sin\varphi\cos\varphi}{\left[\sigma_e\left(t,u\right)\right]^2} - \frac{2\sin\varphi\cos\varphi}{\left[\sigma_t\left(t,u\right)\right]^2}$$

$$C_1 = \frac{\sin^2\varphi}{\left[\sigma_i\left(t,u\right)\right]^2} + \frac{\cos^2\varphi}{\left[\sigma_t\left(t,u\right)\right]^2} \qquad C_2 = \frac{\sin^2\varphi}{\left[\sigma_e\left(t,u\right)\right]^2} + \frac{\cos^2\varphi}{\left[\sigma_t\left(t,u\right)\right]^2}$$

$$\cos\varphi = \frac{t}{\sqrt{t^2+u^2}} \qquad \sin\varphi = \frac{u}{\sqrt{t^2+u^2}}$$

Given this representation of the PSF (and its spatial dependence), the operator * constitutes a modification of the standard convolution. The continuous representation of the operator * is

$$\left[\lambda * PSF\right](x,y,z) = \iiint_{-\infty}^{+\infty} N\lambda\left(t,u,v\right)\left[\theta(-d)e^{-\left[A_1(x-t)^2+B_1(x-t)(y-u)+C_1(y-u)^2\right]}+ \right. \\ \left. +\theta(d)e^{-\left[A_2(x-t)^2+B_2(x-t)(y-u)+C_2(y-u)^2\right]}\right]\cdot e^{-\frac{(z-v)^2}{2\left[\sigma_a(t,u,v)\right]^2}}\,dt\,du\,dv \quad \text{(M.22)}$$

where the normalization factor N is given by

$$N = \frac{2}{(2\pi)^{3/2}\left[\sigma_t\left(t,u,v\right)\right]\left[\sigma_a\left(t,u,v\right)\right]\left[\sigma_i\left(t,u,v\right)+\sigma_e\left(t,u,v\right)\right]}$$

and all coefficients are defined as stated above, performing only the substitutions

$$\sigma_i\left(t,u\right)\rightarrow\sigma_i\left(t,u,v\right) \qquad \sigma_e\left(t,u\right)\rightarrow\sigma_e\left(t,u,v\right) \qquad \sigma_t\left(t,u\right)\rightarrow\sigma_t\left(t,u,v\right)$$

Given the discrete representation of the image, also the operator should be represented in a discrete way. If the PSF is described by a kernel of dimensions $(2R+1)\times(2C+1)\times(2S+1)$, the discretized version of the * operator is

$$\left[\lambda * PSF\right]_{i,j,k} = \sum_{r=-R}^{R}\sum_{c=-C}^{C}\sum_{s=-S}^{S}\lambda_{i+r,j+c,k+s}\left[PSF(i+r,j+c,k+s)\right]_{i,j,k}$$

where $PSF(i+r,j+c,k+s)$ indicates the PSF relative to the voxel $(i+r,j+c,k+s)$.

M.4 The transposed PSF

Inside an iterative reconstruction algorithm of a tomographic imaging technique, a key concept is the projection operator (or *projector*), which links the image space with the space of the detected information (here called *sinogram*). Considering for simplicity only square images $N \times N$ and using some lexicographical ordering criteria, it is possible to represent the sinogram and the image as two vectors Y and X of length, respectively, D and $B = N^2$:

$$Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_d & \cdots & y_D \end{pmatrix}^T \qquad X = \begin{pmatrix} x_1 & x_2 & \cdots & x_b & \cdots & x_B \end{pmatrix}^T$$

Using this scheme, the projector may then be expressed as a matrix

$$P = \begin{pmatrix} p_{11} & p_{21} & \cdots & \cdots & p_{B1} \\ p_{12} & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & p_{bd} & & \ddots & \vdots \\ p_{1D} & \cdots & \cdots & \cdots & p_{BD} \end{pmatrix}$$

(where p_{bd} has the meaning indicated in the text, i.e. it represents the probability of detecting an event coming from pixel b into the sinogram bin d) so that the projection process may be written as a product between a vector and a matrix:

$$Y = PX \qquad y_d = \sum_{b=1}^{B} x_b p_{bd}$$

The inverse process (which consists in passing from the detection space to the image space) is called *backprojection*:

$$X = P^T Y \qquad x_b = \sum_{d=1}^{D} y_d p_{bd}$$

where the matrix P^T is the representation of the backprojector

$$P^T = \begin{pmatrix} p_{11} & p_{12} & \cdots & \cdots & p_{1D} \\ p_{21} & \ddots & & p_{bd} & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ p_{B1} & \cdots & \cdots & \cdots & p_{BD} \end{pmatrix}$$

The introduction of the PSF in the iterative reconstruction algorithm may be interpreted as a redefinition of the projector (and consequently the backprojector) operators:

$$Y = P'X \qquad X = \left(P'\right)^T Y$$

If the PSF is considered to act on the image and stored in a square kernel $M \times M$

$$K = \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1M} \\ k_{21} & k_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ k_{M1} & \cdots & \cdots & k_{MM} \end{pmatrix}$$

the convolution between the image and the PSF will result in a square matrix with $(N + M - 1)$ lines. Starting from this resulting matrix, a cutout of borders should be performed to obtain a matrix of dimensions $N \times N$.

Using the vector X, the convolution process can be written as the product between a rectangular matrix C with $(N + M - 1)^2$ rows and N^2 columns: the following cutout phase, in this case, would consist in removing proper rows from the matrix C and so defining a new matrix C'. The following example should help clarify these steps.

Consider the case with $N = 4$, $M = 3$ and a spatially invariant kernel K:

$$\text{Image} = \begin{pmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{pmatrix} \equiv X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \dots & 16 \end{pmatrix}^T$$

$$K = \begin{pmatrix} \mathrm{a} & \mathrm{d} & \mathrm{g} \\ \mathrm{b} & \mathrm{e} & \mathrm{h} \\ \mathrm{c} & \mathrm{f} & \mathrm{i} \end{pmatrix}$$

$$\text{Image}*\text{PSF} = \left(\begin{array}{c:ccc:c} 1' & 7' & 13' & 19' & 25' & 31' \\ \hdashline 2' & 8' & 14' & 20' & 26' & 32' \\ 3' & 9' & 15' & 21' & 27' & 33' \\ 4' & 10' & 16' & 22' & 28' & 34' \\ 5' & 11' & 17' & 23' & 29' & 35' \\ \hdashline 6' & 12' & 18' & 24' & 30' & 36' \end{array}\right) \equiv X' = \begin{pmatrix} 1' & 2' & 3' & \dots & 36' \end{pmatrix}^T$$

(the region inside the dashed lines corresponds to the original image), where

$$\begin{cases} 1' = 1 \cdot \mathrm{a} \\ 2' = 1 \cdot \mathrm{b} + 2 \cdot \mathrm{a} \\ \vdots \\ 15' = 1 \cdot \mathrm{i} + 2 \cdot \mathrm{h} + 3 \cdot \mathrm{g} + 5 \cdot \mathrm{f} + 6 \cdot \mathrm{e} + 7 \cdot \mathrm{d} + 9 \cdot \mathrm{c} + 10 \cdot \mathrm{b} + 11 \cdot \mathrm{a} \\ \vdots \end{cases}$$

Introducing the matrix C, it is then possible to write

$$X' = CX$$

where C is organized as follows:

$C =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1'	a															
2'	b	a														
3'	c	b	a													
4'		c	b	a												
5'			c	b												
6'				c												
7'	d				a											
8'	e	d			b	a										
9'	f	e	d		c	b	a									
10'		f	e	d		c	b	a								
11'			f	e			c	b								
12'				f				c								
13'	g				d				a							
14'	h	g			e	d			b	a						
15'	i	h	g		f	e	d		c	b	a					
16'		i	h	g		f	e	d		c	b	a				
17'			i	h			f	e			c	b				
18'				i				f				c				
19'					g				d				a			
20'					h	g			e	d			b	a		
21'					i	h	g		f	e	d		c	b	a	
22'						i	h	g		f	e	d		c	b	a
23'							i	h			f	e			c	b
24'								i				f				c
25'									g				d			
26'									h	g			e	d		
27'									i	h	g		f	e	d	
28'										i	h	g		f	e	d
29'											i	h			f	e
30'												i				f
31'													g			
32'													h	g		
33'													i	h	g	
34'														i	h	g
35'															i	h
36'																i

(the empty cells are equal to zero). The grey rows correspond to the resulting image elements to be removed in order to obtain an image with the same dimensions

of the original ones:

$$\text{Cropped image} = \begin{pmatrix} 8' & 14' & 20' & 26' \\ 9' & 15' & 21' & 27' \\ 10' & 16' & 22' & 28' \\ 11' & 17' & 23' & 29' \end{pmatrix} \equiv X'' = \begin{pmatrix} 8' & 9' & 10' & 11' & \dots & 29' \end{pmatrix}^T$$

So, it is possible to write

$$X'' = C'X$$

where

$C' =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
8'	e	d			b	a										
9'	f	e	d		c	b	a									
10'		f	e	d		c	b	a								
11'			f	e			c	b								
14'	h	g			e	d			b	a						
15'	i	h	g		f	e	d		c	b	a					
16'		i	h	g		f	e	d		c	b	a				
17'			i	h			f	e			c	b				
20'					h	g			e	d			b	a		
21'					i	h	g		f	e	d		c	b	a	
22'						i	h	g		f	e	d		c	b	a
23'							i	h			f	e			c	b
26'									h	g			e	d		
27'									i	h	g		f	e	d	
28'										i	h	g		f	e	d
29'											i	h			f	e

The "PSF-aware" projector may be defined, then, as

$$P' = PC'$$

and, consequently, the corresponding backprojector is

$$\left(P'\right)^T = \left(PC'\right)^T = \left(C'\right)^T P^T$$

where $(C')^T$ can be written using the above representation of C:

$(C')^T =$

	8'	9'	10'	11'	14'	15'	16'	17'	20'	21'	22'	23'	26'	27'	28'	29'
1	e	f			h	i										
2	d	e	f		g	h	i									
3		d	e	f		g	h	i								
4			d	e			g	h								
5	b	c			e	f			h	i						
6	a	b	c		d	e	f		g	h	i					
7		a	b	c		d	e	f		g	h	i				
8			a	b			d	e			g	h				
9					b	c			e	f			h	i		
10					a	b	c		d	e	f		g	h	i	
11						a	b	c		d	e	f		g	h	i
12							a	b			d	e			g	h
13									b	c			e	f		
14									a	b	c		d	e	f	
15										a	b	c		d	e	f
16											a	b			d	e

Comparing C' (and how the kernel K is organized inside it) and $(C')^T$, it is then possible to explicit how the transposed PSF kernel K' should be organized (consider for example the greyed column in C'):

$$K' = \begin{pmatrix} \mathrm{i} & \mathrm{f} & \mathrm{c} \\ \mathrm{h} & \mathrm{e} & \mathrm{b} \\ \mathrm{g} & \mathrm{d} & \mathrm{a} \end{pmatrix}$$

Not all the columns in $(C')^T$ contain the complete kernel: the missing rows, however, correspond to the pixels created by the convolution externally to the dimensions of the original matrix (see above). Consequently, when translating the PSF matrix in the PSF kernel, these missing rows can be considered filled with any number, since they will not have any influence on the final image.

Consequently, if the PSF is assumed spatially invariant inside the FOV (as chosen up to this point),

$$PSF^T(\vec{x}) = PSF(-\vec{x})$$

as stated also in [32].

If, instead, a spatially variant PSF is considered

$$K_i = \begin{pmatrix} \mathrm{a}_i & \mathrm{d}_i & \mathrm{g}_i \\ \mathrm{b}_i & \mathrm{e}_i & \mathrm{h}_i \\ \mathrm{c}_i & \mathrm{f}_i & \mathrm{i}_i \end{pmatrix}$$

(where i labels the pixel the PSF refers to), the matrices C and C' have to be changed (for simplicity, only the latter is reported):

$C' =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
8'	e_1	d_2			b_5	a_6										
9'	f_1	e_2	d_3		c_5	b_6	a_7									
10'		f_2	e_3	d_4		c_6	b_7	a_8								
11'			f_3	e_4			c_7	b_8								
14'	h_1	g_2			e_5	d_6			b_9	a_{10}						
15'	i_1	h_2	g_3		f_5	e_6	d_7		c_9	b_{10}	a_{11}					
16'		i_2	h_3	g_4		f_6	e_7	d_8		c_{10}	b_{11}	a_{12}				
17'			i_3	h_4			f_7	e_8			c_{11}	b_{12}				
20'					h_5	g_6			e_9	d_{10}			b_{13}	a_{14}		
21'					i_5	h_6	g_7		f_9	e_{10}	d_{11}		c_{13}	b_{14}	a_{15}	
22'						i_6	h_7	g_8		f_{10}	e_{11}	d_{12}		c_{14}	b_{15}	a_{16}
23'							i_7	h_8			f_{11}	e_{12}			c_{15}	b_{16}
26'									h_9	g_{10}			e_{13}	d_{14}		
27'									i_9	h_{10}	g_{11}		f_{13}	e_{14}	d_{15}	
28'										i_{10}	h_{11}	g_{12}		f_{14}	e_{15}	d_{16}
29'											i_{11}	h_{12}			f_{15}	e_{16}

and, consequently, its transpose will be

$(C')^T =$

	8'	9'	10'	11'	14'	15'	16'	17'	20'	21'	22'	23'	26'	27'	28'	29'
1	e_1	f_1			h_1	i_1										
2	d_2	e_2	f_2		g_2	h_2	i_2									
3		d_3	e_3	f_3		g_3	h_3	i_3								
4			d_4	e_4			g_4	h_4								
5	b_5	c_5			e_5	f_5			h_5	i_5						
6	a_6	b_6	c_6		d_6	e_6	f_6		g_6	h_6	i_6					
7		a_7	b_7	c_7		d_7	e_7	f_7		g_7	h_7	i_7				
8			a_8	b_8			d_8	e_8			g_8	h_8				
9					b_9	c_9			e_9	f_9			h_9	i_9		
10					a_{10}	b_{10}	c_{10}		d_{10}	e_{10}	f_{10}		g_{10}	h_{10}	i_{10}	
11						a_{11}	b_{11}	c_{11}		d_{11}	e_{11}	f_{11}		g_{11}	h_{11}	i_{11}
12							a_{12}	b_{12}			d_{12}	e_{12}			g_{12}	h_{12}
13									b_{13}	c_{13}			e_{13}	f_{13}		
14									a_{14}	b_{14}	c_{14}		d_{14}	e_{14}	f_{14}	
15										a_{15}	b_{15}	c_{15}		d_{15}	e_{15}	f_{15}
16											a_{16}	b_{16}			d_{16}	e_{16}

The kernel (K_i)' should then be organized as follows:

$$(K_i)' = \begin{pmatrix} i_{i-(4)-1} & f_{i-1} & c_{i+(4)-1} \\ h_{i-(4)} & e_i & b_{i+(4)} \\ g_{i-(4)+1} & d_{i+1} & a_{i+(4)+1} \end{pmatrix}$$

or, referring to the row and column indexes r and c,

$$(K_{r,c})' = \begin{pmatrix} \mathrm{i}_{-1,-1} & \mathrm{f}_{-1,0} & \mathrm{c}_{-1,+1} \\ \mathrm{h}_{0,-1} & \mathrm{e}_{0,0} & \mathrm{b}_{0,+1} \\ \mathrm{g}_{+1,-1} & \mathrm{d}_{+1,0} & \mathrm{a}_{+1,+1} \end{pmatrix}$$

where $a_{i,j}$ refers to the coefficient a relative to the pixel located at row $r+i$ and column $c+j$.

Considering general values for N and M, if – with a necessary change in notation – the kernel is indicated with

$$K_{r,c} = \begin{pmatrix} a_1 & a_{M+1} & \cdots & \\ a_2 & \ddots & & \\ \vdots & & \ddots & a_{M^2-1} \\ a_M & \cdots & a_{M^2-M} & a_{M^2} \end{pmatrix}_{r,c}$$

the transposed kernel will be represented as

$$(K_{r,c})' = \begin{pmatrix} a_{M^2}^{\left[-\frac{M-1}{2},-\frac{M-1}{2}\right]} & a_{M^2-M}^{\left[-\frac{M-1}{2},-\frac{M-1}{2}+1\right]} & \cdots \\ a_{M^2-1}^{\left[-\frac{M-1}{2}+1,-\frac{M-1}{2}\right]} & & \\ & \ddots & a_2^{\left[+\frac{M-1}{2}-1,+\frac{M-1}{2}\right]} \\ \vdots & & \\ & \cdots & a_1^{\left[+\frac{M-1}{2},+\frac{M-1}{2}\right]} \end{pmatrix}$$

where $a_n[i,j]$ refers to the coefficient a_n relative to the pixel located at row $r+i$ and column $c+j$ (with this notation, the generic element a_n of the kernel $K_{r,c}$ may also be written as $a_n[0,0]$). Once again, if the PSF is considered spatially invariant, for a generic term a_n in the kernel holds $a_n[i,j] = a_n[0,0] = a_n$ and, as a consequence, the previously shown result

$$(K_{r,c})' = \begin{pmatrix} a_{M^2} & a_{M^2-M} & \cdots & a_M \\ a_{M^2-1} & \ddots & & \vdots \\ \vdots & & \ddots & a_2 \\ \vdots & \cdots & a_{M+1} & a_1 \end{pmatrix}_{r,c}$$

$$PSF^T(\vec{x}) = PSF(-\vec{x})$$

is obtained.

Analogously as above, if a one-dimensional kernel, relative to slice s, is represented by

$$H_s = \begin{pmatrix} a_1 & a_2 & \cdots & a_M \end{pmatrix}_s$$

the transposed kernel will be represented by

$$(H_s)' = \begin{pmatrix} a_M\left[-\frac{M-1}{2}\right] & a_{M-1}\left[-\frac{M-1}{2}+1\right] & \cdots & a_1\left[+\frac{M-1}{2}\right] \end{pmatrix}$$

where $a_n[i]$ refers to the coefficient a_n relative to the pixel located at slice $s+i$.

M.5 Derivation of the variational regularization term

In this section the derivative of the energy function in the variational regularization approach, with respect to the image (equation 6.8) is derived. By introducing the notation

$$\langle f, g\rangle = \langle g, f\rangle = \int (fg)(t)\, dt$$

(where the integral is calculated over the entire space in which f and g are defined), the following equality holds:

$$\left\langle \frac{\partial U}{\partial \lambda}, s \right\rangle = \lim_{\rho\to 0} \frac{U(\lambda+\rho s) - U(\lambda)}{\rho} \tag{M.23}$$

where the energy function is

$$U(\lambda+\rho s) = \int_\Omega \phi\left(|\nabla(\lambda(\omega)+\rho s(\omega))|\right) d\omega = \int_\Omega \phi\left(|\nabla\lambda(\omega)+\rho\nabla s(\omega)|\right) d\omega \tag{M.24}$$

as defined in equation M.24.

Assuming $\rho \ll 1$, since

$$\begin{aligned} |\nabla\lambda+\rho\nabla s| =& \sqrt{(\nabla\lambda+\rho\nabla s)\cdot(\nabla\lambda+\rho\nabla s)} \approx \sqrt{|\nabla\lambda|^2+2\rho\nabla s\cdot\nabla\lambda} \approx \\ \approx& |\nabla\lambda|\left(1+\frac{\rho\nabla s\cdot\nabla\lambda}{|\nabla\lambda|^2}\right) = |\nabla\lambda|+\frac{\rho\nabla s\cdot\nabla\lambda}{|\nabla\lambda|} \end{aligned}$$

equation M.24 can be rewritten as

$$\begin{aligned} U(\lambda+\rho s) &\approx \int_\Omega \phi\left(|\nabla\lambda|+\frac{\rho\nabla s\cdot\nabla\lambda}{|\nabla\lambda|}\right) d\omega \approx \\ &\approx \int_\Omega \left[\phi(|\nabla\lambda|)+\frac{\rho\nabla s\cdot\nabla\lambda}{|\nabla\lambda|}\phi'(|\nabla\lambda|)\right] d\omega = \\ &= U(\lambda)+\int_\Omega \frac{\rho\nabla s\cdot\nabla\lambda}{|\nabla\lambda|}\,\phi'(|\nabla\lambda|)\, d\omega = \\ &= U(\lambda)+\left\langle \phi'(|\nabla\lambda|)\frac{\nabla\lambda}{|\nabla\lambda|}, \rho\nabla s\right\rangle = \\ &= U(\lambda)-\rho\left\langle \nabla\cdot\left[\phi'(|\nabla\lambda|)\frac{\nabla\lambda}{|\nabla\lambda|}\right], s\right\rangle \end{aligned}$$

where it has been assumed the Neumann boundary condition

$$\phi'\left(|\nabla\lambda|\right)\big|_{\partial\Omega} = 0$$

with $\partial\Omega$ indicating the boundary of the image domain Ω.

Consequently, by comparison with equation M.23,

$$\frac{\partial U\left(\lambda\right)}{\partial\lambda} = -\nabla\cdot\left(\phi'\left(|\nabla\lambda|\right)\frac{\nabla\lambda}{|\nabla\lambda|}\right)$$

and, given the definition of $\mathrm{D}\left[\lambda\right]$ in equation 6.8,

$$\mathrm{D}\left[\lambda\right] = -\frac{\partial U\left(\lambda\right)}{\partial\lambda} = \nabla\cdot\left(\phi'\left(|\nabla\lambda|\right)\frac{\nabla\lambda}{|\nabla\lambda|}\right)$$

An alternative derivation for anisotropic diffusion can be found in [111].

Bibliography

[1] S. Alenius. *On noise reduction in iterative image reconstruction algorithms for emission tomography: median root prior.* PhD thesis, Tampere University of Technology, http://research.nokia.com/files/thesis.pdf, October 1999. [cited at p. 46, 48, 49, 89]

[2] A. Alessio, P. Kinahan, R. Harrison, and T. Lewellen. Measured spatially variant system response for PET image reconstruction. In *Nuclear Science Symposium Conference Record*, volume 4, pages 1986–1990. IEEE, 2005. [cited at p. 63, 177]

[3] A. Alessio, P. Kinahan, and T. Lewellen. Modeling and incorporation of system response functions in 3-D whole body PET. *IEEE Trans. Med. Imag.*, 25:828–837, 2006. [cited at p. 76]

[4] A. M. Alessio, C. W. Stearns, S. Tong, S. G. Ross, A. Ganin S. Kohlmyer, and P. E. Kinahan. Application and evaluation of a measured spatially variant system model for PET image reconstruction. *IEEE Trans. Med. Imag.*, 29:938–949, 2010. [cited at p. 79, 139, 175, 176]

[5] H. Anger. Scintillation camera with multichannel collimators. *J. Nucl. Med.*, 5: 515–531, 1964. [cited at p. 32]

[6] E. Asma and R. Manjeshwar. Analysis of organ uniformity in low count density penalized likelihood PET images. In *Nuclear Science Symposium Conference Record*, volume 6, pages 4426–4432. IEEE, 2007. [cited at p. 93, 94]

[7] G. Aubert and L. Vese. A variational method in image recovery. *SIAM J. Numer. Anal.*, 34:1948–1979, 1997. [cited at p. 90]

[8] R. Badawi and P. Marsden. Self–normalization of emission data in 3D PET. *IEEE Trans. Nucl. Sci.*, 46:709–712, 1999. [cited at p. 35]

[9] R. Badawi, M. Miller, D. Bailey, and P. Marsden. Randoms variance reduction in 3D PET. *Phys. Med. Biol.*, 44:941–954, 1999. [cited at p. 24]

[10] B. Bai and P. D. Esser. Effect of edge artifact on quantification of positron emission tomography. In *Nuclear Science Symposium Conference Record*, pages 2263–2266. IEEE, 2010. [cited at p. 179]

[11] D. L. Bailey and S. R. Meikle. A convolution–subtraction scatter correction method for 3D PET. *Phys. Med. Biol.*, 39:411–424, 1994. [cited at p. 25]

[12] E. Beckmann. CT scanning the early days. *The British Journal of Radiology*, 79: 5–8, 2006. [cited at p. 4]

[13] M. Bentourkia, P. Msaki, J. Cadorette, and R. Lecompte. Assessment of scatter components in high-resolution PET: Correction by nonstationary convolution subtraction. *J. Nucl. Med.*, 36:121–130, 1995. [cited at p. 25]

[14] T. Beyer, D. W. Townsend, T. Brun, P. E. Kinahan, M. Charron, R. Roddy, J. Jerin, J. Young, L. Byars, and R. Nutt. A combined PET/CT scanner for clinical oncology. *J. Nucl. Med.*, 41:1369–1379, 2000. [cited at p. 7]

[15] C. Bouman and K. Sauer. A Generalized Gaussian Image Model for Edge-Preserving MAP Estimation. *IEEE Trans. Im. Proc.*, 2:296–310, 1993. [cited at p. 89, 92, 179]

[16] G. Brix, J. Doll, M. Bellemann, H. Trojan, U. Haberkorn, P. Schmidlin, and H. Ostertag. Use of scanner characteristics in iterative image reconstruction for high-resolution positron emission tomography studies of small animals. *Eur. J. Nucl. Med.*, 24:779–786, 1997. [cited at p. 76]

[17] N. Cao, R. H. Huesman, W. W. Moses, and J. Qi. Detection performance analysis for time-of-flight PET. *Phys. Med. Biol.*, 55:6931–6950, 2010. [cited at p. 18]

[18] M. Casey, H. Gadagkar, and D. Newport. A component based method for normalisation in volume PET. In *3rd International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Aix-les-Bains, France*, 1995. [cited at p. 35]

[19] S. R. Cherry, S. R. Meikle, and E. J. Hoffman. Correction and characterization of scattered events in three-dimensional PET using scanners with retractable septa. *J. Nucl. Med.*, 34:671–678, 1993. [cited at p. 25]

[20] S. Cho. *Fast iterative image reconstruction for three-dimensional PET and its extension to time-of-flight PET.* PhD thesis, University of Southern California, http://digitallibrary.usc.edu/assetserver/controller/item/etd-CHO-2369.pdf, 2008. [cited at p. 60]

[21] S. Cho, S. Ahn, Q. Li, and R. M. Leahy. Analytical properties of time-of-flight PET data. *Phys. Med. Biol.*, 53:2809–2821, 2008. [cited at p. 59]

[22] Arthur H. Compton. A quantum theory of the scattering of X–rays by light elements. *Phys. Rev.*, 21:483–502, 1923. [cited at p. 22]

[23] M. Conti. State of the art and challenges of time-of-flight PET. *Physica Medica*, 25:1–11, 2009. [cited at p. 18, 60]

[24] M. Conti. Why is TOF PET reconstruction a more robust method in presence of inconsistent data? *Phys. Med. Biol.*, 56:155–168, 2011. [cited at p. 18, 60]

[25] M. Conti, B. Bendriem, M. Casey, M. Chen, F. Kehren, C. Michel, and V. Panin. First experimental results of time-of-flight reconstruction on an LSO PET scanner. *Phys. Med. Biol.*, 50:4507–4526, 2005. [cited at p. 59, 60]

[26] Data Spectrum Corporation. `http://www.spect.com/pub/NEMA_IEC_Body_Phantom_Set.pdf`. [cited at p. 112, 115]

[27] M. Couceiro, A. Blanco, N. C. Ferreira, R. Ferreira Marques, P. Fonte, and L. Lopes. RPC-PET: status and perspectives. *Nucl. Instr. Meth. A*, 580:915–918, 2007. [cited at p. 16]

[28] M. E. Daube-Witherspoon, S. Surti, A. Perkins, C. C. M. Kyba, R. Wiener, M. E. Werner, R. Kulp, and J. S. Karp. The imaging performance of a $LaBr_3$-based PET scanner. *Phys. Med. Biol.*, 55:45–64, 2010. [cited at p. 14, 144]

[29] H. W. A. M. De Jong, F. H. P. Van Velden, R. W. Kloet, F. L. Buijs, R. Boellaard, and A. A. Lammertsma. Performance evaluation of the ECAT HRRT: an LSO-LYSO double layer high resolution, high sensitivity scanner. *Phys. Med. Biol.*, 52: 1505–1526, 2007. [cited at p. 176, 177]

[30] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49:2463–2475, 2004. [cited at p. 45, 64]

[31] M. Defrise and L. Xuan. Fast and exact Fourier rebinning using John's equation. In *Nuclear Science Symposium Conference Record*, volume 2, pages 869–873. IEEE, 1999. [cited at p. 45]

[32] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J. Olivo-Marin, and J. Zerubia. *3D Microscopy Deconvolution using Richardson-Lucy Algorithm with Total Variation Regularization.* INRIA, July 2004. [cited at p. 79, 198]

[33] G. Dougherty. *Digital Image Processing for medical applications.* Cambridge, 2009. [cited at p. 48]

[34] J. Engle and D. Kadrmas. Modeling the spatially-variant point spread function in a fast projector for improved fully-3D PET reconstruction. *J. Nucl. Med.*, 48: 417P, 2007. [cited at p. 175]

[35] G. T. Fechner. *Elemente der Psychophysik — vol. 2, Leipzig: Breitkopf und Haertel.* 1860. [cited at p. 97]

[36] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 6:721–741, 1984. [cited at p. 88, 92, 179]

[37] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Prentice Hall, second edition, 2002. [cited at p. 48]

[38] P. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imag.*, 9:84–93, 1990. [cited at p. 89]

[39] S. Grootoonk, T. J. Spinks, D. Sashin, N. M. Spryou, and T. Jones. Correction for scatter in 3D brain PET using a dual energy window method. *Phys. Med. Biol.*, 41:2757–2774, 1996. [cited at p. 25]

[40] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imag.*, 8:194–202, 1989. [cited at p. 88]

[41] W. R. Hendee and E. R. Ritenour. *Medical Imaging Physics*. John Wiley & Sons, fourth edition, 2002. [cited at p. 3, 4]

[42] E. Hewitt and R. Hewitt. The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis. *Arch. Hist. Exact Sci.*, 21:129–160, 1980. [cited at p. 127]

[43] E. Hoffman, T. Guerrero, G. Germano, W. Digby, and M. Dahlborn. PET system calibrations and corrections for quantitative and spatially accurate images. *IEEE Trans. Nucl. Sci.*, 36:1108–1112, 1989. [cited at p. 36]

[44] P. J. Huber. *Robust statistics*. John Wiley and Sons, 1981. [cited at p. 92, 179]

[45] H. Hudson and R. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 13:601–609, 1994. [cited at p. 50, 55, 75]

[46] M. Iatrou, R. Manjeshwar, S. Ross, K. Thielemans, and C. Stearns. 3D implementation of scatter estimation in 3D PET. In *Nuclear Science Symposium Conference Record*, pages 2142–2145. IEEE, 2006. [cited at p. 26]

[47] L. Jaroff. A winning combination. *Time Magazine*, 156(23), 4 Dec. 2000. [cited at p. 8]

[48] A. R. Karimian and C. J. Thompson. Assessment of a new scintillation crystal $LaBr_3$ in PET scanners using Monte Carlo method. *Nukleonika*, 53:3–6, 2008. [cited at p. 14]

[49] J. S. Karp, S. Surti, M. E. Daube-Witherspoon, and G. Muehllehner. Benefit of time-of flight in PET: experimental and clinical results. *J. Nucl. Med.*, 49:462–470, 2008. [cited at p. 144]

[50] S. L. Keeling. Total Variation based convex filters for medical imaging. *Appl. Math. and Comput.*, 139:101–119, 2003. [cited at p. 90, 92, 93, 127, 130, 179]

[51] P. E. Kinahan and J. G. Rogers. Analytic 3D image reconstruction using all detected events. *IEEE Trans. Nucl. Sci.*, 36:964–968, 1990. [cited at p. 45]

[52] G. F. Knoll. *Radiation Detection and Measurement*. John Wiley & Sons, third edition, 2000. [cited at p. 14, 22, 32]

[53] Isotope Products Laboratories. http://www.ipl.isotopeproducts.com/new_ipl_site/pdf/IPL_Multimodal_Sources.pdf. [cited at p. 63]

[54] T. F. Lang, B. H. Hasegawa, S. C. Liew, J. K. Brown, S. Blankespoor, S. M. Reilly, E. L. Gingold, and C. E. Cann. A prototype emission-transmission imaging system. In *Nuclear Science Symposium Conference Record*, volume 3, pages 1902–1906. IEEE, 1991. [cited at p. 7]

[55] K. Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Trans. Med. Imag.*, 9:439–446, 1990. Corrections ITMI 10(2) 1991 p. 228. [cited at p. 89]

[56] P. Lecoq, A. Annenkov, A. Gektin, M. Korzhik, and C. Pedrini. *Inorganic Scintillators for Detector Systems*. Springer Berlin, 2006. [cited at p. 15]

[57] C. S. Levin, M. Dahlbom, and E. J. Hoffman. A Monte-Carlo correction for the effect of Compton scattering in 3D PET brain imaging. *IEEE Trans. Nucl. Sci.*, 42:1181–1185, 1995. [cited at p. 25]

[58] T. K. Lewellen. The challenge of detector designs for PET. *Am. J. Roentgenol.*, 195:301–309, 2010. [cited at p. 15]

[59] Z. Liang. Detector response restoration in image reconstruction of high resolution Positron Emission Tomography. *IEEE Trans. Med. Imag.*, 13:314–321, 1994. [cited at p. 33]

[60] C. Lois, B. W. Jakoby, M. J. Long, K. F. Hubner, D. W. Barker, M. E. Casey, M. Conti, V. Y. Panin, D. J. Kadrmas, and D. W. Townsend. An assessment of the impact of incorporating time-of-flight information into clinical PET/CT imaging. *J. Nucl. Med.*, 51:237–245, 2010. [cited at p. 18]

[61] M. Lubberink, T. Kosugi, H. Schneider, H. Ohba, and M. Bergström. Non–stationary convolution subtraction scatter correction with a dual–exponential scatter kernel for the Hamamatsu SHR–7700 animal PET scanner. *Phys. Med. Biol.*, 49:833–842, 2004. [cited at p. 25]

[62] A. Mallon and P. Grangeat. Three-dimensional PET reconstruction with time-of-flight measurements. *Phys. Med. Biol.*, 37:717–729, 1992. [cited at p. 59]

[63] M. Marengo. *La fisica in medicina nucleare*. Pàtron Editore, 2001. [cited at p. 14, 21, 37, 54]

[64] T. Moon. The expectation–maximization algorithm. *IEEE Signal Processing Magazine*, pages 47–60, 1996. [cited at p. 50]

[65] W. Moses and S. Derenzo. Empirical observation of resolution degradation in positron emission tomographs using block detectors. *J. Nucl. Med.*, 34:101P, 1993. [cited at p. 37]

[66] G. Muehllehner and J. S. Karp. Positron emission tomography. *Phys. Med. Biol.*, 51:R117–R137, 2006. [cited at p. 12]

[67] E. Mumcuoğlu, R. Leahy, S. Cherry, and E. Hoffman. Accurate geometric and physical response modelling for statistical image reconstruction in high resolution PET. In *Nuclear Science Symposium Conference Record*, volume 3, pages 1569–1573. IEEE, 1996. [cited at p. 56, 79]

[68] S. Mustafovic and K. Thielemans. Additive and multiplicative versions of the maximum a posteriori algorithm with median root prior. In *Nuclear Science Symposium Conference Record*, volume 3, pages 1783–1785. IEEE, 2001. [cited at p. 104]

[69] J. M. Ollinger. Model-based scatter correction for fully 3D PET. *Phys. Med. Biol.*, 41:153–176, 1996. [cited at p. 26]

[70] V. Panin, F. Kehren, C. Michel, and M. Casey. Fully 3–D PET reconstruction with system matrix derived from point source measurements. *IEEE Trans. Med. Imag.*, 25:907–921, 2006. [cited at p. 76, 79, 175, 176, 177]

[71] K. Parodi. *On the feasibility of dose quantification with in-beam PET data in radiotherapy with ^{12}C and proton beams.* PhD thesis, Fakultät Mathematik und Naturwissenschaften der Technischen Universität Dresden, 2004. [cited at p. 13]

[72] M. Partridge, A. Spinelli, W. Ryder, and C. Hindorf. The effect of β^+ energy on performance of a small animal PET camera. *Nucl. Instr. Meth. A*, 568:933–936, 2006. [cited at p. 21]

[73] D. G. Politte and D. L. Snyder. The use of constraints to eliminate artifacts in maximum-likelihood image estimation for emission tomography. *IEEE Trans. Nucl. Sci.*, 35:608–610, 1988. [cited at p. 179]

[74] G. Pratx, S. Surti, and C. Levin. Fast list-mode reconstruction for Time-of-Flight PET using graphics hardware. *IEEE Trans. Nucl. Sci.*, 58:105–109, 2011. [cited at p. 60]

[75] J. Qi. Investigation of Lesion Detection in MAP reconstruction with Non Gaussian Priors. In *Nuclear Science Symposium Conference Record*, volume 3, pages 1704–1708. IEEE, 2005. [cited at p. 89]

[76] J. Qi and R. Leahy. Iterative reconstruction techniques in emission computed tomography. *Phys. Med. Biol.*, 51:R541–R578, 2006. [cited at p. 89]

[77] J. Qi, R. Leahy, C. Hsu, T. Farquhar, and S. Cherry. Fully 3D Bayesian image reconstruction for the ECAT EXACT HR+. *IEEE Trans. Nucl. Sci.*, 45:1096–1103, 1998. [cited at p. 89]

[78] J. Qi, R. M. Leahy, S. R. Cherry, A. Chatziioannou, and T. H. Farquhar. High resolution 3D bayesian image reconstruction using the microPET small animal scanner. *Phys. Med. Biol.*, 43:1001–1013, 1998. [cited at p. 56, 76, 79, 175, 176]

[79] A. Rahmim, J. Tang, M. A. Lodge, S. Lashkari, M. R. Ay, R. Lautamäki, B. M. W. Tsui, and F. M. Bengel. Analytic system matrix resolution modeling in PET: an application to Rb-82 cardiac imaging. *Phys. Med. Biol.*, 53:5947–5965, 2008. [cited at p. 175]

[80] A. Reader. The promise of new PET image reconstruction. *Physica Medica*, 24: 49–56, 2008. [cited at p. 43, 44]

[81] A. Reader and H. Zaidi. Advances in PET image reconstruction. *PET Clin.*, 2: 173–190, 2007. [cited at p. 44, 49]

[82] A. Reader, F. Sureau, C. Comtat, I. Buvat, and R. Trébossen. List-mode reconstruction with system modeling derived from projections. In *Nuclear Science Symposium Conference Record*, pages 1832–1836. IEEE, 2005. [cited at p. 56, 76]

[83] A. J. Reader, P. J. Julyan, H. Williams, D. L. Hastings, and J. Zweit. EM algorithm system modeling by image–space techniques for PET reconstruction. *IEEE Trans. Nucl. Sci.*, 50:1392–1397, 2003. [cited at p. 76, 79, 83, 176, 179]

[84] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, 2009. [cited at p. 90, 179]

[85] V. Selivanov, Y. Picard, J. Cadorette, S. Rodrigue, and R. Lecomte. Detector response models for statistical iterative image reconstruction in high resolution PET. *IEEE Trans. Nucl. Sci.*, 47:1168–1175, 2000. [cited at p. 56]

[86] L. Shao, R. Freifelder, and J. S. Karp. Triple energy window scatter correction technique in PET. *IEEE Trans. Med. Imag.*, 13:641–648, 1994. [cited at p. 25]

[87] L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, MI-2:113–122, 1982. [cited at p. 50, 75]

[88] D. L. Snyder, L. J. Thomas, and M. M. Ter-Pogossian. A mathematical model for positron-emission tomography systems having time-of-flight measurements. *IEEE Trans. Nucl. Sci.*, 28:3575–3583, 1981. [cited at p. 59]

[89] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte. Noise and edge artifacts in maximum-likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, 6:228–238, 1987. [cited at p. 79, 81, 179]

[90] C. W. Stearns. Scatter correction method for 3D PET using 2D fitted gaussian functions. *J. Nucl. Med.*, 36:105P, 1995. [cited at p. 26]

[91] F. C. Sureau, A. J. Reader, C. Comtat, C. Leroy, M. Ribeiro, I. Buvat, and R. Trébossen. Impact of image-space resolution modeling for studies with the High-Resolution Research Tomograph. *J. Nucl. Med.*, 49:1000–1008, 2008. [cited at p. 79, 176]

[92] S. Surti, A. Kuhn, M.E. Werner, A.E. Perkins, J. Kolthammer, and J.S. Karp. Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. *J. Nucl. Med.*, 48:471–480, 2007. [cited at p. 144]

[93] E. Tapscott and A. Gottschalk. Nuclear Medicine Pioneer, Hal O. Anger, 1920-2005. *J. Nucl. Med. Technol.*, 33:250–253, 2005. [cited at p. 6]

[94] M. M. Ter-Pogossian, M. E. Phelps, and E. J. Hoffman. A positron-emission transaxial tomograph for nuclear imaging (PETT). *Radiology*, 114:89–98, 1975. [cited at p. 7]

[95] M. Teräs, T. Tolvanen, J. Johansson, J. Williams, and J. Knuuti. Performance of the new generation of whole-body PET/CT scanners: Discovery STE and Discovery VCT. *Eur. J. Nucl. Mol. Imaging*, 34:1683–1692, 2007. [cited at p. 107]

[96] K. Thielemans, E. Asma, S. Ahn, R. M. Manjeshwar, T. Deller, S. G. Ross, C. W. Stearns, and A. Ganin. Impact of PSF modeling on the convergence rate and edge behavior of EM images in PET. In *Nuclear Science Symposium Conference Record*, pages 3267–3272. IEEE, 2010. [cited at p. 179]

[97] M. S. Tohme and J. Qi. Iterative image reconstruction for positron emission tomography based on a detector response function estimated from point source measurements. *Phys. Med. Biol.*, 54:3709–3725, 2009. [cited at p. 64, 175]

[98] T. Tomitani. Image reconstruction and noise evaluation in photon time-of-flight assisted positron emission tomography. *IEEE Trans. Nucl. Sci.*, 28:4582–4588, 1981. [cited at p. 59]

[99] D. W. Townsend. Multimodality imaging of structure and function. *Phys. Med. Biol.*, 53:R1–R39, 2008. [cited at p. 14]

[100] D. W. Townsend, T. Beyer, P. E. Kinahan, T. Brun, R. Roddy, R. Nutt, and L. G. Bryars. The SMART scanner: a combined PET/CT tomography for clinical oncology. In *Nuclear Science Symposium Conference Record*, volume 2, pages 1170–1174. IEEE, 1998. [cited at p. 7]

[101] S. Vandenberghe and J. Karp. Rebinning and reconstruction techniques for 3D TOF-PET. *Nucl. Instr. Meth. Phys. Res. A*, 569:421–424, 2006. [cited at p. 60]

[102] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for Positron Emission Tomography. *J. Am. Stat. Assoc.*, 80:8–20, 1985. [cited at p. 184]

[103] C. R. Vogel. *Computational Methods for Inverse Problems.* SIAM, 2002. [cited at p. 87, 179]

[104] O. Warburg. On the origin of cancer cells. *Science*, 123(3191):309–314, 1956. [cited at p. 12]

[105] C. C. Watson. Extension of single scatter simulation to scatter correction of time-of-flight PET. *IEEE Trans. Nucl. Sci.*, 54:1679–1686, 2007. [cited at p. 60]

[106] E. H. Weber. Der Tastsinn und das Gemeingefuhl. In *Handwörterbuch der Physiologie — vol. 3, Braunschweig.* 1850. [cited at p. 97]

[107] M. N. Wernick and J. N. Aarsvold. *Emission tomography: the fundamentals of PET and SPECT.* Elsevier, 2004. [cited at p. 14]

[108] H. Wilbraham. On a certain periodic function. *Cambridge and Dublin Math. J.*, 3:198–201, 1848. [cited at p. 127]

[109] J. M. Wilson, S. G. Ross, T. Deller, E. Asma, R. Manjeshwar, and T. G. Turkington. A phantom study of regularized image reconstruction in PET. In *Nuclear Science Symposium Conference Record*, pages 3661–3665. IEEE, 2010. [cited at p. 94]

[110] W. H. Wong, N. A. Mullani, E. A. Philippe, R. Hartz, and K. L. Gould. Image improvement and design optimization of the time-of-flight PET. *J. Nucl. Med.*, 24: 52–60, 1983. [cited at p. 144]

[111] Y. You, W. Xu, A. Tannenbaum, and M. Kaveh. Behavioral analysis of anisotropic diffusion in image processing. *IEEE Trans. Im. Proc.*, 5:1539–1553, 1996. [cited at p. 202]

[112] H. Zaidi and K. F. Koral. Scatter modelling and compensation in emission tomography. *Eur. J. Nucl. Med. Mol. Imaging*, 31:761–782, 2004. [cited at p. 30]

List of Figures

1.1 The first X-ray image . . . 4
1.2 The first CT image . . . 4
1.3 Hal Anger and Benedict Cassen . . . 6
1.4 Some applications of scintigraphy . . . 6
1.5 The PETT II prototype . . . 7
1.6 Example of PET-CT exam . . . 8

2.1 Chemical structure of FDG . . . 13
2.2 Different metabolism between glucose and FDG . . . 13
2.3 Scheme of a crystal block read by PMTs . . . 16
2.4 Difference between 2D and 3D acquisitions due to septa usage . . . 17
2.5 Scheme of pixelated detector design . . . 17
2.6 Comparison between the positioning of annihilation events in conventional and TOF PET . . . 19
2.7 Compton effect . . . 23
2.8 Energy of a scattered photon for an incoming 511 keV photon . . . 23
2.9 Polar plot of the Klein–Nishina distribution . . . 27
2.10 Example of convertion from Hounsfield units into attenuation coefficients at 511 keV energy . . . 29
2.11 Example of a photon energy spectrum acquired in a PET scanner . . . 31
2.12 Possible wrong evaluation of pulse emission due to multicrystal reading 32
2.13 Effects on bin amplitude due to radial geometry . . . 33
2.14 Different path lengths inside the scintillator crystals depending on incidence angle of the photon . . . 34
2.15 Response of a scintillator crystal to a point source located on the crystal axis . . . 34
2.16 Depth–of–interaction effect . . . 34
2.17 Scatter inside scintillator crystals . . . 35
2.18 General radial repositioning scheme . . . 36

2.19 Principal components of PET resolution . . . 38
2.20 Increasing radial asymmetry with higher distance from FOV center . . 39
2.21 Source located off-center and radially asymmetric two-dimensional Gaussian . . . 39

3.1 Variables used in the data representation . . . 42
3.2 Three methods to store data — an example of a point source . . . 43
3.3 Some of the projectors used in tomography . . . 45
3.4 Radon transform and Fourier slice theorem . . . 46
3.5 Ramp filter . . . 47
3.6 Common filters used in FBP . . . 48
3.7 Example of FBP artifacts . . . 49
3.8 Time-binned sinograms for TOF reconstruction . . . 59
3.9 Variables for 3D TOF . . . 60

4.1 Example of OSEM reconstruction of a point source . . . 65
4.2 Extraction of the three planes for PSF fit . . . 66
4.3 Introduction of the equivalent radius in the fitting procedure . . . 69
4.4 Example of simulated PSF acquisition . . . 72
4.5 Example of fits on simulated PSF acquisition . . . 72
4.6 Validation of PSF fitting procedure – 1 . . . 73
4.7 Validation of PSF fitting procedure – 2 . . . 74

5.1 Organization of the transaxial PSF kernel . . . 77
5.2 Organization of the axial PSF kernel . . . 78
5.3 Storage of integral value of the PSF . . . 79
5.4 Edge effect in a reconstruction of three uniform spheres . . . 79
5.5 One-dimensional simulation without PSF recovery . . . 81
5.6 Main results of data reconstruction one-dimensional simulation with PSF recovery . . . 82
5.7 Data used for two-dimensional simulation of edge effect . . . 83
5.8 Main results of data reconstruction two-dimensional simulation with PSF recovery . . . 84
5.9 Removal of edge artefact with increasing iterations using the correct PSF . . . 84

6.1 Effects of different p values in the p-Gaussian prior . . . 95
6.2 $\phi(x)$ and $\phi''(x)$ with different values of the parameter δ for the Gauss-Total Variation, the p-Gaussian and the proposed priors . . . 96
6.3 Deviation from pure Poisson noise in reconstructed images . . . 99
6.4 Example of simulated PET image . . . 100
6.5 Comparison of the voxel content in simulated and real PET image . . . 101
6.6 Validation of the detectability index . . . 103

7.1 Example of PSF fits on DSTE . . . 108
7.2 Dependences of the spread parameters on the radial and axial distances — DSTE . . . 109
7.3 Comparison of real, fitted and kernel PSF profiles — DSTE . . . 110
7.4 Images of the ^{22}Na sources . . . 112
7.5 ^{22}Na sources: RnoPSF, RnoPSF-Filt and RwPSF . . . 113
7.6 ^{22}Na sources: invariant PSF . . . 114
7.7 NEMA IEC Body Phantom Set . . . 115
7.8 NEMA IEC phantom: CR_{hot} vs. COV with PSF . . . 117
7.9 Images of the NEMA IEC phantom . . . 117
7.10 NEMA IEC phantom: CR_{hot} vs. COV with invariant PSF . . . 118
7.11 Cylindrical phantom: CR_{hot} coefficients . . . 119
7.12 Images of the cylindrical phantom . . . 119
7.13 Coronal MIP images of the oncology patient . . . 121
7.14 Transaxial images of the oncology patient . . . 122
7.15 Transaxial images of the neurological patient . . . 123
7.16 Sagittal images of the neurological patient . . . 124
7.17 Optimization of the regularization parameters for the proposed and the Gauss-Total Variation priors . . . 126
7.18 Comparison of the different reconstruction algorithms concerning the edge effect . . . 128
7.19 Activity line profiles in the uniform spheres obtained with the different regularization priors . . . 129
7.20 Images of the NEMA IEC phantom with different regularization strategies . . . 132
7.21 NEMA IEC phantom: CR_{hot} vs. COV with PSF and regularization . . . 133
7.22 Oncological patient A: comparison of the different reconstruction algorithms . . . 136
7.23 Oncological patient B: comparison of the different reconstruction algorithms . . . 137

8.1 Example of PSF fits on D690 . . . 140
8.2 Dependences of the spread parameters on the radial and axial distances — D690 . . . 141
8.3 Comparison of real, fitted and kernel PSF profiles — D690 . . . 142
8.4 Comparison of the NEMA IEC Body Phantom images with/without PSF and/or TOF . . . 148
8.5 CR coefficients versus background COV for non TOF and TOF reconstruction algorithms . . . 149
8.6 CR coefficients and COV versus the number of iterations for non TOF and TOF reconstruction algorithms . . . 150

8.7 Coronal images of oncological patient C: comparison of the different reconstruction algorithms . . . 152
8.8 Coronal images of oncological patient C: comparison of the different reconstruction algorithms . . . 153
8.9 Transaxial images of oncological patient C: comparison of the different reconstruction algorithms . . . 154
8.10 Coronal images of oncological patient D: comparison of the different reconstruction algorithms . . . 156
8.11 Coronal images of oncological patient D: comparison of the different reconstruction algorithms . . . 157
8.12 Transaxial images of oncological patient D: comparison of the different reconstruction algorithms . . . 158
8.13 Optimization of the regularization parameters for the proposed prior with TOF . . . 159
8.14 Comparison of the NEMA IEC Body Phantom images with TOF and regularization . . . 162
8.15 NEMA IEC phantom: CR_{hot} vs. COV with PSF, TOF and regularization . . . 163
8.16 Coronal images of oncological patient C: comparison of the different reconstruction algorithms and regularization . . . 167
8.17 Coronal images of oncological patient C: comparison of the different reconstruction algorithms and regularization . . . 168
8.18 Transaxial images of oncological patient C: comparison of the different reconstruction algorithms and regularization . . . 169
8.19 Coronal images of oncological patient D: comparison of the different reconstruction algorithms and regularization . . . 170
8.20 Coronal images of oncological patient D: comparison of the different reconstruction algorithms and regularization . . . 171
8.21 Transaxial images of oncological patient D: comparison of the different reconstruction algorithms and regularization . . . 172

M.1 Cylindrical source with axis parallel to the tangential direction . . . 190

List of Tables

2.1 Some β^+-emitting isotopes of potential interest in PET imaging . . . 12
2.2 Some of the non-FDG radiopharmaceuticals used in PET imaging . . 13
2.3 Principal crystals used in PET 14
2.4 Summary of the most recent PET systems available with their main characteristics 20
2.5 End–point energy, modal energy and positron ranges for commonly used PET radioisotopes 21

4.1 Choice of sigma parameters for the simulation of PSF fitting 71

6.1 Characteristics of the proposed prior compared to Gauss-Total Variation and p-Gaussian priors 97

7.1 Percent differences of CR_{hot} and COV with respect to RnoPSF 131
7.2 Percent differences of CR_{hot} and COV with respect to RwPSF-Filt . . 134

8.1 Percent differences for non TOF and TOF reconstructions of CR coefficients and COV with respect to RnoPSF 146
8.2 Percent differences for non TOF and TOF reconstructions of CR coefficients and COV with respect to TOF RnoPSF 147
8.3 Percent differences, with respect to RnoPSF, of CR_{hot}, CR_{cold} and COV with TOF and regularization 164
8.4 Percent differences, with respect to RnoPSF, of CR_{hot}, CR_{cold} and COV with TOF and regularization 165

List of Symbols and Abbreviations

Abbreviation	Description	Definition
ACF	attenuation correction factor	page 56
APD	avalanche photodiode	page 15
BGO	bismuth germanate	page 14
CLT	central limit theorem	page 37
COV	coefficient of variation	page 116
CT	computerized tomography	page 3
D690	Discovery 690	page 139
DSTE	Discovery STE	page 107
DOI	depth of interaction	page 33
FBP	filtered backprojection	page 47
FDG	2-[^{18}F]fluoro-2-deoxy-D-glucose	page 12
FORE	Fourier rebinning	page 45
FOV	field of view	page 15
FFT	fast Fourier transform	page 47
FT	Fourier transform	page 47
FWHM	full width at half maximum	page 30
GF	Gaussian function	page 37
GSO	gadolinium orthosilicate	page 15
GTV	Gauss-Total Variation	page 92

Abbreviation	Description	Definition
IR	iterative reconstruction	page 50
LBR	lesion-to-background ration	page 37
LOR	line of response	page 11
LSO	lutetium orthosilicate	page 15
LUT	lookup table	page 5
LYSO	lutetium–yttrium orthosilicate	page 15
MAP	maximum a posteriori	page 88
MIP	maximum intensity projection	page 120
MLEM	maximum likelihood expectation maximization	page 50
MRI	magnetic resonance imaging	page 7
OSEM	ordered subsets expectation maximization	page 50
OSL	one-step-late	page 89
PDF	probability density function	page 50
PMT	photomultiplier tube	page 15
PSF	point spread function	page 9
PVE	partial volume effect	page 37
RPC	resistive plate chambers	page 16
ROI	region of interest	page 21
QED	quantum electrodynamics	page 14
SNR	signal to noise ratio	page 18
SPECT	single photon emission computerized tomography	page 5
TOF	time of flight	page 15
TSF	time spread function	page 60
VOI	volume of interest	page 21

Symbol	Description
β	regularization strength
δ	regularization threshold
$\lambda_b^{[k]}$	voxel b of image λ at iteration k
σ_a	axial sigma of PSF
σ_e	external (towards the crystals) radial sigma of PSF
σ_i	internal (towards the scanner center) radial sigma of PSF
σ_t	tangential sigma of PSF
ϕ	regularization prior
A_d	attenuation along LOR d
B	number of voxels in the image
B	backprojector
D	number of LORs in the sinogram
D	regularization term, derivative of the energy function
P	projector
p_{bd}	probability of detecting an event from voxel b in LOR d
R_d	random coincidences along LOR d
S_d	scattered coincidences along LOR d
$U(\lambda)$	regularization energy function
y_d	events detected in LOR d
RnoPSF	OSEM reconstruction without PSF
RnoPSF-Filt	OSEM reconstruction without PSF and a postfilter
RwPSF	OSEM reconstruction with PSF
RwPSF-Filt	OSEM reconstruction with PSF and a postfilter
RwPSF-GTVR	OSEM reconstruction with PSF and Gauss-Total Variation regularization prior
RwPSF-PR	OSEM reconstruction with PSF and p-Gaussian regularization prior
RwPSF-R	OSEM reconstruction with PSF and the herein proposed regularization prior