



King's Research Portal

DOI:

[10.1088/1361-6560/aa7670](https://doi.org/10.1088/1361-6560/aa7670)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Mehranian, A., Belzunce, M. A., Niccolini, F., Politis, M., Prieto, C., Turkheimer, F., Hammers, A., & Reader, A. J. (2017). PET image reconstruction using multi-parametric anato-functional priors. *Physics in Medicine and Biology*, 62(15), 5975–6007. <https://doi.org/10.1088/1361-6560/aa7670>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

PET image reconstruction using multi-parametric anato-functional priors

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 Phys. Med. Biol. 62 5975

(<http://iopscience.iop.org/0031-9155/62/15/5975>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 159.92.151.144

This content was downloaded on 12/07/2017 at 16:25

Please note that [terms and conditions apply](#).

You may also be interested in:

[Anatomy-guided brain PET imaging incorporating a joint prior model](#)

Lijun Lu, Jianhua Ma, Qianjin Feng et al.

[Anatomically-aided PET reconstruction using the kernel method](#)

Will Hutchcroft, Guobao Wang, Kevin T Chen et al.

[Regularized image reconstruction with an anatomically adaptive prior for PET](#)

Chung Chan, Roger Fulton, David Dagan Feng et al.

[Patch-based image reconstruction for PET using prior-image derived dictionaries](#)

Marzieh S Tahaei and Andrew J Reader

[Bayesian PET image reconstruction](#)

Jing Tang and Arman Rahmim

[PET image reconstruction with anatomical edge guided level set prior](#)

Jinxu Cheng-Liao and Jinyi Qi

[Accelerated time-of-flight \(TOF\) PET image reconstruction using TOF bin subsetization and TOF weighting matrix pre-computation](#)

Abolfazl Mehranian, Fotis Kotasidis and Habib Zaidi



**FILL IN THE
MISSING PIECES.**

When every piece
counts, get the
whole picture with
LinacView and Adaptive.

VISIT US AT AAPM BOOTH 4051

STANDARD IMAGING



PET image reconstruction using multi-parametric anato-functional priors

Abolfazl Mehranian^{1,4} , Martin A Belzunce¹,
Flavia Niccolini², Marios Politis², Claudia Prieto¹,
Federico Turkheimer³, Alexander Hammers¹
and Andrew J Reader¹

¹ Department of Biomedical Engineering, Division of Imaging Sciences and Biomedical Engineering, King's College London, St. Thomas' Hospital, London, United Kingdom

² Neurodegeneration Imaging Group, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London, London, United Kingdom

³ Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London, London, United Kingdom

E-mail: Abolfazl.Mehranian@kcl.ac.uk

Received 11 October 2016, revised 18 April 2017

Accepted for publication 1 June 2017

Published 6 July 2017



CrossMark

Abstract

In this study, we investigate the application of multi-parametric anato-functional (MR-PET) priors for the maximum *a posteriori* (MAP) reconstruction of brain PET data in order to address the limitations of the conventional anatomical priors in the presence of PET-MR mismatches. In addition to partial volume correction benefits, the suitability of these priors for reconstruction of low-count PET data is also introduced and demonstrated, comparing to standard maximum-likelihood (ML) reconstruction of high-count data. The conventional local Tikhonov and total variation (TV) priors and current state-of-the-art anatomical priors including the Kaipio, non-local Tikhonov prior with Bowsher and Gaussian similarity kernels are investigated and presented in a unified framework. The Gaussian kernels are calculated using both voxel- and patch-based feature vectors. To cope with PET and MR mismatches, the Bowsher and Gaussian priors are extended to multi-parametric priors. In addition, we propose a modified joint Burg entropy prior that by definition exploits all parametric information in the MAP reconstruction of



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

⁴ Division of Imaging Sciences & Biomedical Engineering, King's College London, 3rd Floor, Lambeth Wing, St Thomas' Hospital, London, SE1 7EH, United Kingdom

PET data. The performance of the priors was extensively evaluated using 3D simulations and two clinical brain datasets of [^{18}F]florbetaben and [^{18}F]FDG radiotracers. For simulations, several anato-functional mismatches were intentionally introduced between the PET and MR images, and furthermore, for the FDG clinical dataset, two PET-unique active tumours were embedded in the PET data. Our simulation results showed that the joint Burg entropy prior far outperformed the conventional anatomical priors in terms of preserving PET unique lesions, while still reconstructing functional boundaries with corresponding MR boundaries. In addition, the multi-parametric extension of the Gaussian and Bowsher priors led to enhanced preservation of edge and PET unique features and also an improved bias-variance performance. In agreement with the simulation results, the clinical results also showed that the Gaussian prior with voxel-based feature vectors, the Bowsher and the joint Burg entropy priors were the best performing priors. However, for the FDG dataset with simulated tumours, the TV and proposed priors were capable of preserving the PET-unique tumours. Finally, an important outcome was the demonstration that the MAP reconstruction of a low-count FDG PET dataset using the proposed joint entropy prior can lead to comparable image quality to a conventional ML reconstruction with up to 5 times more counts. In conclusion, multi-parametric anato-functional priors provide a solution to address the pitfalls of the conventional priors and are therefore likely to increase the diagnostic confidence in MR-guided PET image reconstructions.

Keywords: PET-MRI, image reconstruction, anatomical priors, regularization

 Supplementary material for this article is available [online](#)

(Some figures may appear in colour only in the online journal)

1. Introduction

The recently introduced simultaneous clinical PET-MR (positron emission tomography—magnetic resonance) imaging systems are able to provide molecular imaging data and complementary multi-parametric (MP) MRI information. The MRI data can be exploited to guide PET image reconstruction and therefore help reduce the noise and resolution blurring that usually degrade the quality of PET images. In current clinical practice, the PET data are mainly reconstructed using the maximum likelihood expectation maximization (MLEM) algorithm and point spread function (PSF) resolution modelling (Reader *et al* 2002). However, for routine PET scans, the true ML solution is very noisy and inclusion of PSF during reconstruction can lead to Gibbs artefacts near edges. Hence, Bayesian maximum *a posteriori* (MAP) image reconstruction has been explored to reduce noise and stabilize the ML solution using *a priori* knowledge of the unknown image, such as how smooth it is or what it should look like. In a smooth image, it is highly probable that neighbouring voxels have similar intensity values compared to distant voxels. A prior that encourages this property, such as a quadratic Markov random field (MRF) prior, attempts to suppress large local differences between voxels on the basis that they are probably due to noise. However, some of the local differences are associated with legitimate image boundaries which should be preserved during image reconstruction.

Over the last two decades, various MRF priors have been designed to improve upon the quadratic prior using non-quadratic edge preserving potential functions that assign a lower penalty to large local differences on the assumption that they are probably associated with valid boundaries. The priors that utilize non-quadratic potential functions, such as total variation (TV) (Huber 1981, Lange 1990, Rudin *et al* 1992), can preserve edges; however they are usually limited in distinguishing valid edges from noisy fluctuations as they rely mainly on short-range voxel interactions. In addition, due to the low-pass filtering nature of the PET scanners and the resulting partial volume averaging effect (PVE), some boundaries are lost and cannot be restored purely from the PET data. To enhance the performance of such priors, two main approaches have been explored in the literature that extend the local priors to non-local ones for the robust identification of edges using either the long-range interaction of image voxels or their intensity similarities (Yu and Fessler 2002, Chen *et al* 2008). These approaches, which have their origins from bilateral filtering (Tomasi and Manduchi 1998) and non-local means (Buades *et al* 2005), have been successfully employed in MAP PET image reconstruction (Chen *et al* 2008, Wang and Qi 2012) as well as post-reconstruction filtering (Dutta *et al* 2013, Chan *et al* 2014). However, as these methods rely only the PET information, they might not be able to handle PVE and to fully recover the lost boundary information. Another promising approach is to employ anatomical boundaries available in hybrid PET-MR or PET-CT to assign a lower weight on the smoothing of the PET image boundaries (Gindi *et al* 1993).

Bowsher *et al* (2004) proposed to calculate the interaction (local differences) of each image voxel with its most similar neighbours predefined from an anatomical image. Essentially, the Bowsher method weights the local differences using zero-one weighting factors, thereby disabling the smoothing across boundaries, but encouraging it within anatomical regions. The main advantages of this method are that it does not require the pre-segmentation of anatomical images, and it can be easily incorporated into different MRF priors; however, as it relies on the 'binary' selection of each voxel's neighbours, it is particularly vulnerable to mismatches between the anatomical image and the true activity distribution. To improve the performance of the Bowsher prior in the case of the anato-functional inconsistencies, Kazantsev *et al* (2011) investigated the combination of the Bowsher's weights and those calculated from the PET estimate. Nguyen and Lee (2013) calculated the non-local weights from both PET and anatomical images to address the mismatches between PET and MR or CT information.

Based on the Lysaker–Osher–Tai (LOT) model (Lysaker *et al* 2004), another approach is to improve MRF priors using the normalized gradient vector fields (normal vectors) obtained from anatomical side information. Kazantsev *et al* (2014) included the normal vectors obtained from a complementary CT image into the LOT model to improve the TV prior based on the alignment of the PET and normal vectors. However, this proposed prior depends on the orientations of the anatomical and functional edges (e.g. in the case of hot and cold lesions). More recently, Ehrhardt *et al* (2016) derived a prior based on the structural similarity of the PET and MR images measured by the alignment of the PET and MR gradients. In fact, this prior generalizes the prior proposed by Kaipio *et al* (1999) for improving the quadratic prior. As will be discussed in section 2.1.4, the Kaipio prior in practice encourages edge preservation by weighting the magnitude of PET image gradients (i.e. local differences) with the angle between the PET gradients and MR normal vectors. One of the advantages of these methods is that the prior is reduced to a conventional MRF prior, if PET image boundaries do not have corresponding anatomical boundaries in MR images. However, similar to the Bowsher prior, they might induce false boundaries in the reconstructed PET images if MR boundaries do not have corresponding functional boundaries in the true PET image. Moreover, the quality of the MR image is an important factor in deriving accurate normal vectors, as clinical MR images

typically suffer from noise and/or partial volume effects, especially when they are down-sampled to match the voxel size and field of view of the PET images.

All of the above-mentioned anatomical priors put less weight or penalty on the smoothing of PET boundaries based on boundary side information. A different class of priors instead aims to maximize an information-based similarity metric, such as mutual information or joint entropy (JE), between PET and anatomical images (Nuyts 2007), based on the assumption that they are mutually informative and the images can be inferred from each other. Nuyts found that the JE prior results in lower bias in the reconstructed PET images compared to those reconstructed using mutual information (Nuyts 2007); similar results were also observed by Somayajula *et al* (2011). The JE prior does not require the pre-segmentation of the anatomical image and can be robust to inconsistencies between PET and anatomical images, as it relies on their joint probability distribution; however, it also ignores the spatial correlation between neighbouring voxels, which is an important feature for PET tracers that are confined in anatomical regions.

The availability of MP images in simultaneous PET-MR systems can provide a unique opportunity for improving PET image quality. These parametric maps are mutually complementary and informative, for instance, an early-stage lesion might not have any morphological manifestations in T1- or T2-weighted MR images but might appear metabolically active in PET images and at the same time show restricted/elevated diffusion patterns in diffusion-weighted MR images. In this regard, the aim of this study was fourfold: (i) unification of the state-of-the-art anatomical priors in a common reconstruction framework, which is of importance in unveiling the similarity between different priors and therefore in developing robust and reliable reconstruction setups. The studied priors include the local Tikhonov and TV priors, non-local anatomically-guided Tikhonov prior, the Kaipio and Bowsher priors. In addition, we proposed a modified joint Burg entropy that outperforms the state-of-the-art Bowsher prior in the case of PET-MR mismatches; (ii) extension of the anatomical priors to MP anato-functional ones. Given the simultaneity and complementary nature of PET-MR data, we proposed the novel idea of MP priors to cope with PET-MR mismatches; (iii) evaluation of the MAP reconstructions using the same optimization algorithm and the same simulations and clinical datasets, given that a comparison of these priors is still missing in the literature; and (iv) application of anatomically guided PET reconstruction for not only partial volume correction (PVC) but also for reducing the PET scan time or injected dose. To the best of our knowledge, for the first time, the application of anato-functional priors is demonstrated for reducing scan time or injected dose while achieving the same image quality obtained using conventional reconstruction of high-count data. In this study, we focused on the application and evaluation of different priors for brain studies using simulated brain phantoms and clinical brain datasets using two different radiotracers. However, the developed methodologies should also be applicable to whole body studies.

2. Material and methods

2.1. MAP PET image reconstruction

Let $\mathbf{y} \in \mathbb{Z}_+^M$ be the set of independently measured emission data collected during the PET scan of an object with the activity distribution represented by $\mathbf{u} \in \mathbb{R}^N$. In Bayesian PET image reconstruction, the MAP reconstruction consists of maximizing the posterior probability of \mathbf{u} given \mathbf{y} , that is:

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmax}} \left\{ \sum_{i=1}^M \log p(y_i|\mathbf{u}) - \beta R(\mathbf{u}) \right\} \quad (1)$$

where the probability distribution $p(y_i|\mathbf{u})$ of the i th measured data value given an image \mathbf{u} is often best modelled using a Poisson distribution with an expected value of

$$\mathbb{E}[y_i] = n_i a_i \sum_{j=1}^N g_{ij} u_j + \mathbb{E}[r_i] \quad (2)$$

where g_{ij} is the geometric probability detection of annihilation events emitted from the j th voxel in the i th PET detector, n_i and a_i are detector normalization factors and attenuation factors and $\mathbb{E}[r_i]$ is the expected number of scatter and random events in the i th detector. In the MAP framework, the PET image is also treated as a random vector with a probability distribution. As mentioned earlier, in PET imaging a smooth or noise-free image is always sought, in which local differences (interactions) between neighbouring voxels are small. The probability distribution of interacting voxels can be well modelled using a MRF and a Gibbs distribution as $P(\mathbf{u}) \propto \exp(-\beta R(\mathbf{u}))$, where β is a regularization parameter and $R(\mathbf{u})$ is the Gibbs energy function, commonly known as a regularizer or prior. In this study, we considered the following prior, which generalizes most of the included MRF priors:

$$R(\mathbf{u}) = \sum_j^N \phi \left(\sum_{b \in \mathcal{N}_j} \xi_{jb} \omega_{jb} \psi(u_j - u_b) \right), \quad \xi_{jb} = \frac{1}{\sqrt{\sum_{i=1}^3 (j^{(i)} - b^{(i)})^2}} \quad (3)$$

ψ and ϕ are potential functions, where ψ operates on the intensity differences between the j th voxel and its neighbouring voxels, and ϕ operates on the neighbourhood (\mathcal{N}_j) sum of those local differences. ξ_{bj} and ω_{bj} are coefficients that weight the intensity differences between voxels j and b based on their proximity (in this study the inverse of their Euclidean distance) and based on their similarity, respectively (see figure 1). $\{j^{(i)}\}$ and $\{b^{(i)}\}$ are the Cartesian coordinates of the j th and b th voxel. The Tikhonov prior is defined by setting $\psi(s) = s^2$ and $\phi(t) = t$, while the smoothed isotropic TV prior is defined for $\psi(s) = s^2$ and $\phi(t) = \sqrt{t + \delta^2}$, where $\delta > 0$ is a smoothing factor that makes the prior continuously differentiable and reduces the stair-casing and patchy artefacts usually observed in non-smooth TV regularization (the derivative of this smoothed TV prior is given in appendix A).

In the regularization of a 3D image volume, the prior R with a neighbourhood of 6 nearest neighbours (i.e. a first-order neighbourhood), and $\omega_{jb} = 1$, is referred to as a local regularization, while for a higher-order neighbourhood (e.g. of the size 342 for a $7 \times 7 \times 7$ patch) and spatially variant ω_{jb} the prior is referred to as a non-local regularization method (Wang and Qi 2012). The ω_{jb} coefficients encourage the smoothness along boundaries but discourage it across them. Generally, there are three approaches in the calculation of these coefficients, based on: (i) the PET image reconstructed from a previous iteration (Chen *et al* 2008), (ii) the MRI or CT side anatomical information as in the Bowsher method, which can be aptly referred to as side-similarity and (iii) both the PET and side anatomical information (Kazantsev *et al* 2011, Nguyen and Lee 2013). In this study, we extend the latter approach to calculate the similarity kernels based on all available MP MRI data as well as the PET data.

Specifically, we evaluate the conventional local quadratic (Tikhonov) and TV priors, the non-local quadratic prior with Bowsher and Gaussian kernels (Chen *et al* 2008), the Kaipio prior and an modified joint Burg entropy prior. For the optimization in (1), we employed Green's one step late (OSL) MAP-EM algorithm (Green 1990), which extends the MLEM

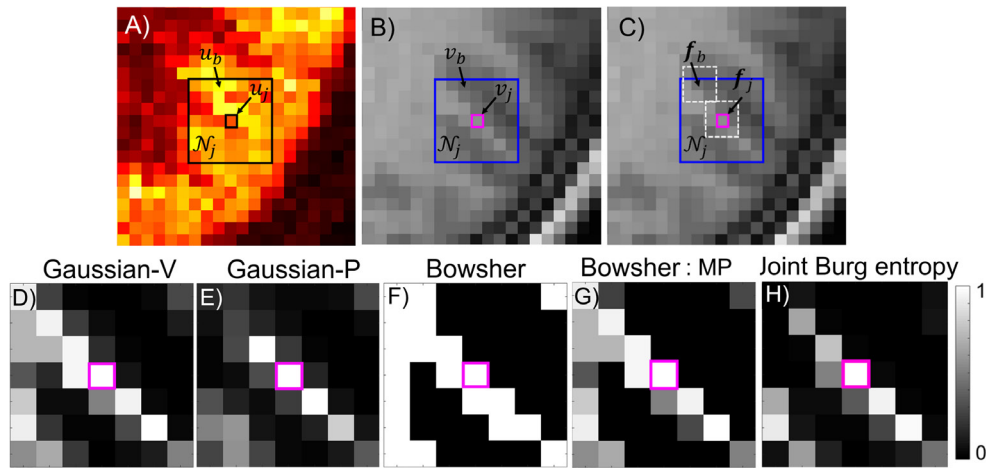


Figure 1. (A) Illustration of a 7×7 neighbourhood (\mathcal{N}_j) around the j th voxel of the PET image (u_j). (B) and (C) The corresponding neighbourhoods in a co-registered T1-MR image, indicating individual voxels and patches of voxels, respectively. The similarity coefficients of the j th voxel of the PET image calculated from the MR image using the Gaussian kernel method with (D) voxel-based (Gaussian-V) and (E) patch-based (Gaussian-P) feature vectors, (F) using the Bowsher method (using the 20 most similar neighbours). (G) and (H) The same coefficients calculated from both the PET and MR images using a MP extension of the Bowsher prior and using the proposed joint Burg entropy prior, respectively.

algorithm by adding the first derivative of the prior, evaluated at the previous iteration, to the sensitivity image. In the following we elaborate the above priors, their first-derivative and the extension of the priors amenable to the MP case.

2.1.1. Non-local quadratic prior with MP Gaussian similarity kernels. For this prior, the similarity coefficients between a given voxel of the PET image, u_j , and its neighbouring voxels, u_b , in the neighbourhood \mathcal{N}_j are calculated from a reference image \mathbf{x} (e.g. a complementary MRI, a prior PET image or the PET image under reconstruction) using the following Gaussian functions:

$$\omega_{jb}^V(\mathbf{x}) = \frac{1}{z} \exp\left(-\frac{(\mathbf{x}_j - \mathbf{x}_b)^2}{2\sigma^2}\right), \quad \omega_{jb}^P(\mathbf{x}) = \frac{1}{z} \exp\left(-\frac{\|\mathbf{f}_j(\mathbf{x}) - \mathbf{f}_b(\mathbf{x})\|_2^2}{2\sigma^2}\right), \quad z = \sum_b \omega_{jb} \quad (4)$$

where in ω_{jb}^V , the similarities are calculated based on their individual voxel intensity values (namely, Gaussian-V), while in ω_{jb}^P , they are calculated based on the intensity values of a patch of voxels, $\mathbf{f}_l(\mathbf{x})$, centred on voxels j and b (namely, Gaussian-P), see figures 1(B) and (C). The σ and z are the shape (standard deviation) and normalization parameters of the Gaussian functions, respectively. Figures 1(D) and (E) shows the similarity kernels calculated using the Gaussian-V and -P methods from the T1-MR image. As can be seen, the Gaussian-P method is comparably stricter in identifying the similar voxels to the highlighted central voxel, therefore it is more robust to noise, especially when the similarity coefficients are calculated from the PET image itself. Note that the proximity coefficients in (3) can also be calculated using a Gaussian kernel as in bilateral Gaussian filtering (Tomasi and Manduchi 1998); however, this kernel will add an additional σ parameter that requires tuning by the user. In this study,

we therefore calculated these weights based on the inverse of the Euclidean distance between voxels.

The Gaussian kernels were extended to MP ones based on the geometric mean of the individual kernels calculated from different parametric images. Let $\{\mathbf{v}^{(i)} \in \mathbb{R}^N\}_{i=1}^L$ be a set of L MP MR images, the MP similarity kernels were then defined as:

$$\omega_{jb} = \frac{1}{z} \sqrt{(L+1)} \sqrt{\omega_{jb}^p(\mathbf{u}) \times \omega_{jb}^x(\mathbf{v}^{(1)}) \dots \times \omega_{jb}^x(\mathbf{v}^{(L)})} \tag{5}$$

where $\omega_{jb}^p(\mathbf{u})$ is calculated from the PET image using Gaussian-P kernels, which makes the resulting kernels robust to noise in the PET image, $\omega_{jb}^x(\mathbf{v}^{(i)})$ is calculated from the i th parametric image using either Gaussian-P or Gaussian-V methods and z is a normalization factor. The derivative of the resulting quadratic prior is given by:

$$\frac{\partial R(\mathbf{u})}{\partial u_j} = 2 \sum_{b \in \mathcal{N}_j} \xi_{jb} \omega_{jb} (u_j - u_b). \tag{6}$$

2.1.2. Non-local quadratic prior with MP Bowsher similarity kernels. In the Bowsher method, the most similar neighbouring voxels of a given voxel in the PET image (u_j) in the neighbourhood \mathcal{N}_j are identified from the same neighbourhood in a complementary MR image, as shown in figure 1(F). In this approach, the top B most similar neighbours are calculated based on their absolute differences in the MR image, i.e. $|v_b - v_j| \geq \dots \geq |v_B - v_j|$, thereby the resulting similarity coefficients will be 0 or 1, where 0 is assigned to the voxels not within the top B most similar values. For MP extension of the Bowsher method, we followed the approach proposed by Kazantsev *et al* (2011) in which the similarity kernels are obtained by the product of the Bowsher’s binary coefficients, calculated from the anatomical image, and the Gaussian-P kernels calculated from the current estimate of the PET image (see figure 1(G)). The derivative of the resulting prior was calculated as in equation (6), which would correspond to an *asymmetric* Bowsher prior, as proposed in Vunckx and Nuyts (2010).

2.1.3. MP local joint Burg entropy. In this section, we present a MP JE prior with local spatial interaction of voxels based on the Burg entropy (Byrne 1993). Let the vectors $\mathbf{u}, \mathbf{v}^{(1)}, \dots, \mathbf{v}^{(L)}$ be a realization of a set of random variables $\mathbf{U}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(L)}$. The JE is an information-theoretic criterion that measures the information gained from this set of variables with the joint probability distribution function (PDF) of $p(\mathbf{U}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(L)})$. The joint Burg entropy is defined as the integration of the logarithm of the joint PDF of the variables:

$$H(\mathbf{U}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(L)}) = \sum_{l,m,\dots,n}^S \log p(u_l, v_m^{(1)}, \dots, v_n^{(L)}) \delta u \delta v^{(1)} \dots \delta v^{(L)} \tag{7}$$

where S is the number of bins at which the PDF is integrated with δ bin widths. Note the difference between this prior and the widely used Shannon JE is that the latter weights the logarithm of the joint PDF with the negative of the joint PDF in order to provide a weighted average of information content (Pluim *et al* 2003). The Burg entropy in fact measures the dispersion of the joint PDF of the variables. If the variables are mutually informative and can be inferred from each other, their joint PDF is more compact and therefore their entropy is low, whereas if they are independent, their joint PDF is broad and thus the entropy is high. Therefore, the aim is to minimize the joint Burg entropy in order to increase the correlation

of the variables. The joint PDF can be approximated using the non-parametric Parzen density estimation with a multivariate Gaussian window function with a diagonal covariance matrix $\Sigma = \text{diag} \{ \sigma_u^2, \sigma_{v^{(1)}}^2, \dots, \sigma_{v^{(L)}}^2 \}$, as follows:

$$p(x, y, \dots, z) = \frac{1}{N} \sum_b^N \mathcal{G}(x, u_b, \sigma_u) \mathcal{G}(y, v_b^{(1)}, \sigma_{v^{(1)}}) \dots \mathcal{G}(z, v_b^{(L)}, \sigma_{v^{(L)}}) \tag{8}$$

$$\mathcal{G}(q, r, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(q-r)^2}{2\sigma^2}\right)$$

where N is the number of samples used to calculate the PDF, which similar to Somayajula *et al* (2011) was set equal to the number of voxels in the images and \mathcal{G} is Gaussian function as in (4). Since the logarithm tends to flatten the joint PDF, the integration in (7) can be approximated by the mean of the log-distribution, as follows:

$$H(\mathbf{U}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(L)}) \approx -\frac{1}{N} \sum_j^N \log p(u_j, v_j^{(1)}, \dots, v_j^{(L)}) \tag{9}$$

As pointed out in Somayajula *et al* (2011), the Shannon JE can also be approximated in the same way such that it can be defined as the above equation. Note that to be consistent with the notion of maximum entropy and the derivative of the quadratic priors used in this study, similar to Shannon entropy, we introduced a negative sign in (9). Thus, the aim would be to maximize the negative of joint Burg entropy. The OSL MAP-EM reconstruction requires the calculation of the first derivative of the employed prior. The derivative of JE evaluated at j th bin (or voxel) depends on all voxel intensities simultaneously, which leads to discarding regional information. As suggested in Somayajula *et al* (2011), the derivative can be approximated by evaluating the summation for the voxels that are in the neighbourhood of the j th voxel, that is (see appendix B for more details):

$$\frac{\partial H}{\partial u_j} = 2 \sum_{b \in \mathcal{N}_j} \xi_u \omega_{jb} (u_j - u_b), \quad \xi_u = \frac{1}{2N\sigma_u^2} \tag{10}$$

where ω_{jb} act as similarity weighting coefficients given by:

$$\omega_{jb} = \frac{1}{p(u_j, v_j^{(1)}, \dots, v_j^{(L)})} \exp\left(-\frac{(u_j - u_b)^2}{2\sigma^2}\right) \prod_{i=1}^L \exp\left(-\frac{(v_j^{(i)} - v_b^{(i)})^2}{2\sigma_{v^{(i)}}^2}\right) \tag{11}$$

In figure 1(H), the coefficients calculated from both the PET and MR images have been shown. We replaced ξ_u in (10) with the proximity coefficients ξ_{jb} in (3) in order to further improve the performance of this prior in spatial neighbourhoods. As a result of this modification, the range of the regularization parameter, β , for this prior will be approximately the same as for the other Tikhonov-based priors.

2.1.4. Kaipio prior. Based on the formulations in Ehrhardt *et al* (2016), the Kaipio prior is a quadratic prior improved by inclusion of the squared inner product of the PET image gradient and MR normal vectors. Let $\mathbf{g}_j \in \mathbb{R}^D$ and $\mathbf{n}_j \in \mathbb{R}^D$ be the gradient vector of the PET image \mathbf{u} and the normalized gradient vector of an anatomical image \mathbf{v} at the j th voxel, respectively, where D is the size of the neighbourhood of the j th voxel. The b th elements of these vectors are given by:

$$g_{jb} = u_j - u_b, \quad n_{jb} = \frac{v_j - v_b}{\sqrt{\sum_{b \in \mathcal{N}_j} (v_j - v_b)^2}}. \quad (12)$$

The Kaipio prior can be defined as (see appendix C for more details):

$$R(\mathbf{u}) = \sum_j^N \sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b)^2 - \left(\sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}} (u_j - u_b) \right)^2. \quad (13)$$

This prior can also be described based on the alignment of the vectors \mathbf{g}_j and \mathbf{n}_j and the angle θ_j formed between them, as follows (see appendix B for more details):

$$R(\mathbf{u}) = \sum_j^N \sin^2(\theta_j) \sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b)^2. \quad (14)$$

Basically, a non-zero \mathbf{n}_j indicates the presence of an edge in the anatomical image, therefore if the vectors \mathbf{g}_j and \mathbf{n}_j are aligned, that is, $\theta_j = 0$ (parallel) or $\theta_j = 180^\circ$ (anti-parallel), the regularization of the corresponding PET edge is switched off leading to its preservation. Note in the cases that $\mathbf{n}_j = 0$ (e.g. where there is no anatomical edge in the j th voxel of the MR image), the prior is reduced to the quadratic prior. The main advantage of the Kaipio prior over the Bowsher and Gaussians priors is that it is free of any shape parameter (e.g. B and σ), however, it has two disadvantages: firstly, as the prior relies on MR normal vectors, it is very susceptible to noise and PVE artefacts in the MR images (which are usually down-sampled to the size of PET images) due to the normalization in (12), therefore even low-amplitude noise can indicate the presence of an edge, which makes the prior unable to discern valid edges. Therefore, the de-noising and de-blurring of MR images is of high importance for the performance of this type of prior (see figure 9 in the *Results* section). Secondly, its extension to a MP prior is not straightforward as one needs to calculate the resultant normal vectors from MP MR images and more importantly to estimate the normal vectors from the noisy PET image itself, which requires designing an additional regularized estimation approach (Lysaker *et al* 2004). For the above reasons, we evaluated this prior for only T1-weighted MR guided PET image reconstruction. The derivative of the Kaipio prior is the same as in (6) with the weights given by (see appendix C for details):

$$\omega_{jb} = 1 - \frac{n_{jb}}{\sqrt{\xi_{jb}}} \sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}}. \quad (15)$$

In table 1, the derivatives of the aforementioned priors and their similarity kernels are summarized. Note the similarity kernels of anatomical priors are shown for conventional MR-guided priors.

2.2. Optimization and parameter selection

The MAP problem in (1) was optimized using the Green’s one-step-late (OSL) MAP-EM algorithm, as follows:

$$u_j^{n+1} = \frac{u_j^n}{\sum_{i=1}^M g_{ij} n_i a_i + \beta \frac{\partial R(\mathbf{u}^n)}{\partial u_j}} \sum_{i=1}^M g_{ij} n_i a_i \frac{y_i}{n_i a_i \sum_b g_{ib} u_b^n + r_i} \quad (16)$$

Table 1. The similarity kernels of the conventional and anatomical priors included in this study, with the partial derivative of $\frac{\partial R}{\partial u_j} = 2 \sum_{b \in \mathcal{N}_j} \xi_{jb} \omega_{jb} (u_j - u_b)$.

Prior	Similarity kernels: ω_{jb}	Description
Local Tikhonov	1	Results in uniform smoothing
Local TV	$\frac{1}{2\sqrt{\sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b)^2 + \delta^2}}$	Encourages piece-wise smoothing based on the inverse of the magnitude of the local differences
^a Gaussian-V	$\frac{1}{z} \exp\left(-\frac{(x_j - x_b)^2}{2\sigma^2}\right)$	Improves upon the Tikhonov prior by weighting the local differences based on their voxel-wise Gaussian similarity in the reference image \mathbf{x}
^a Gaussian-P	$\frac{1}{z} \exp\left(-\frac{f_j(x) - f_b(x)^2}{2\sigma^2}\right)$	Calculates the Gaussian similarity of voxel intensities of reference image \mathbf{x} based on their local neighborhood (patch)
Kaipio	$1 - \frac{n_{jb}}{\sqrt{\xi_{jb}}} \sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}}$	Preserves the edges wherein the vector of local differences and MR normal vectors (n_{jb}) are (anti-) parallel
Bowsher	$ v_b - v_j \geq \dots \geq v_B - v_j $ subject to a binary selection	Selects the B most similar neighbours of a given voxel in the MR image, \mathbf{v} , based on their absolute differences
Joint Burg entropy	$\frac{\mathcal{G}(u_j, \sigma_u) \mathcal{G}(v_j, \sigma_v)}{\sum_b \mathcal{G}(u_j, \sigma_u) \mathcal{G}(v_j, \sigma_v)}$	Calculates the similarity of voxel intensities as the normalized product of their Gaussian similarities in both the PET, \mathbf{u} , and the MR image \mathbf{v}

^a Gaussian similarity kernels with voxel (V) and patch (P) based feature vectors.

\mathbf{u} and \mathbf{v} represent PET and MR images.

σ_u and σ_v are the standard deviation of the PET and MRI Gaussian kernels.

\mathcal{G} : the standard Gaussian kernel defined in equation (8).

where r_i is expected value of random and scatter coincidences, obtained from a delayed coincidence window (Noll 1999) and model-based scatter simulation (Watson 2000). This iterative algorithm differs from the ordinary Poisson (OP) ML-EM algorithm principally in that the first derivative of the prior R , evaluated at the previous (latest) image update, is added to the sensitivity image (which is obtained from the back-projection of the product of the attenuation and normalization sinograms). However, it should be noted that the OSL MAP-EM algorithm has been shown to converge to the MAP solution only for certain types of priors, particularly, those for which the potential functions have a bounded derivative (Lange 1990). Alternative approaches to OSL include using De Pierro’s convexity lemma (De Pierro 1995) to define a separable surrogate for the prior R , which can guarantee the convergence to the maximizer of the a-posterior density.

The priors included in the present work have a number of user-defined parameters that determine their performance. Table 2 summarizes these key parameters, along with those with pre-defined values commonly used in our simulation and clinical data reconstructions. For the local priors including Tikhonov, TV, a first-order neighborhood consisting of 6 immediate neighbors was chosen, while for the anatomical priors a higher-order neighborhood (a search window of size $7 \times 7 \times 7$) consisting of 342 neighbors was chosen. A local window size of $3 \times 3 \times 3$ was chosen for the calculation of Gaussian kernels with patch-based feature vectors.

The ML-EM and MAP-EM image reconstructions were declared generally converged when the relative difference between successive image estimates fell below a tolerance of $\tau = 1 \times 10^{-4}$, otherwise, if the total number of iterations reached a maximum number of $n = 150$. The relative difference was defined as the normalized Euclidian distance between the current, \mathbf{u}^{n+1} , and previous, \mathbf{u}^n , image estimates, as follows:

Table 2. List of the parameters of the regularization methods included in this study. The pre-defined values of some of the parameters commonly used in this study are also presented.

Priors	Parameters
Local Tikhonov	$ \mathcal{N}_3 = 6, \beta$
Local TV	$ \mathcal{N}_3 = 6, \delta = 1 \times 10^{-3}, \beta$
Kaipio	$ \mathcal{N}_7 = 342, \beta$
^a Gaussian-V	$ \mathcal{N}_7 = 342, \sigma_u, \beta$
^a Gaussian-P	$ \mathcal{N}_7 = 342, \mathcal{F}_3 = 26, \sigma_u, \beta$
Bowsher	$ \mathcal{N}_7 = 70, \beta$
Joint Burg entropy	$ \mathcal{N}_7 = 342, \sigma_u, \sigma_v, \beta$

^a Gaussian similarity kernels with voxel (V) and patch (P) based feature vectors.

$|\mathcal{N}_n|, |\mathcal{F}_n|$: number of included neighbors in a $n \times n \times n$ neighborhood (\mathcal{N}) or feature vector (\mathcal{F}).

β : Prior's weighting factor (regularization parameter).

δ : TV smoothness parameter.

σ_u, σ_v : Standard deviation of Gaussian kernels used for PET and MR images.

$$\eta(k) = \frac{\|\mathbf{u}^{n+1} - \mathbf{u}^n\|_2}{\|\mathbf{u}^n\|_2}. \quad (17)$$

2.3. Datasets

2.3.1. Numerical PET-MR simulations. To objectively evaluate the performance of the studied priors and their MP versions with respect to a ground truth, extensive 3D realistic simulation studies were performed. The *BrainWeb* phantom (Cocosco et al 1997) was utilized to simulate the bio-distribution of [¹⁸F]fluorodeoxyglucose (FDG) radiotracer in the brain together with T1- and T2-weighted MR images and attenuation maps, all with a matrix size of $344 \times 344 \times 127$ and voxel of size $2 \times 2 \times 2 \text{ mm}^3$. A representative sagittal slice of this phantom is shown in figure 2. To further simulate the existence of mismatched boundaries or abnormalities between PET and MR images, a few regions of the T1 and T2-MR images were uniquely and commonly removed in such a way that the anatomical inconsistencies were simulated among all three parametric images (see arrows 1 and 3 in figure 2). Moreover, four unique and shared lesions were considered in the PET activity and T2-weighted MR images (each contains a small and a large lesion, see arrows 2 and 4 in figure 2). The only shared lesion is shown in the true activity map in figure 2. For the true activity map, the ratio of mean activity concentration in grey matter (GM) to white matter (WM) and that of lesions to WM were simulated to 3:1 and 7:1, respectively. The signal intensity of the T1- and T2-weighted MR images were simulated according to the *BrainWeb*'s simulated brain database. To simulate PVE and noise, the MR images were smoothed using a Gaussian filter, 3 mm full width at half maximum (FWHM) and Gaussian noise with standard deviation equal to 1% of the largest voxel intensity was added to the images. The simulated attenuation map consists of air, soft tissue and bone tissue classes, with linear attenuation coefficients of 0, 0.098, 0.13 cm^{-1} , respectively. For this phantom, two sets of reconstructions were performed including the conventional T1-MR guided PET image reconstruction and the MP MR-PET-guided PET image reconstruction. In the latter, we considered the priors that could be extended to include MP data such as Bowsher, Gaussian kernel methods and joint Burg entropy.

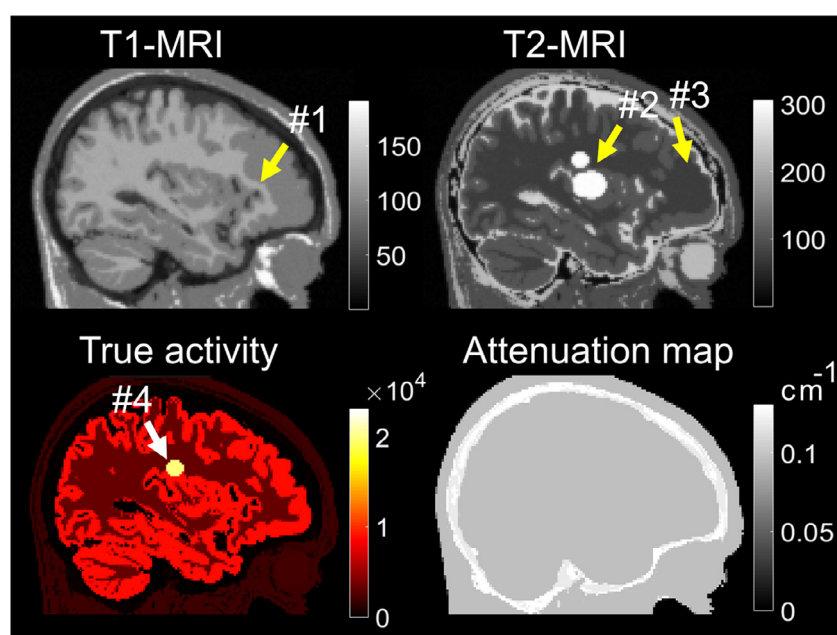


Figure 2. The 3D MP brain numerical phantom used in this study. The arrows point to the boundary mismatches among T1-MR, T2-MR and PET activity images (#1 and #3) as well as unique and common lesions in T2-MR and PET activity images (#2 and #4). The simulated attenuation map consists of air, soft tissue and bone tissue classes.

In our 3D simulations, we modelled a forward projector based on the native geometry of the Siemens Biograph mMR scanner (Siemens Healthcare, Erlangen), which included attenuation, normalization factors, randoms and scatter coincidences, as well as PSF modelling. The normalization factors were estimated from a 5 h ^{68}Ge phantom scan using a component-based approach (Belzunce and Reader 2016). The scatter sinograms were simulated by Gaussian smoothing in the radial and angular directions (with Gaussian kernels of 20 cm and 8 radians in FWHM) of the direct planes (segment zero) of net-true emission sinograms (obtained from forward projection of the true activity which was then attenuated and normalized by the attenuation and normalization factors). The resulting simulated 2D scatter sinograms were then extended to 3D sinograms using the so-called inverse single slice rebinning and scaled to simulate a scatter fraction of 50%. A randoms fraction of 30% was also simulated using Poisson noise derived from a uniform mean value. The PET system's PSF was modelled using an image-space shift-invariant Gaussian convolution kernel of 4 mm in FWHM. This approximation is appropriate for brain imaging in a large-bore clinical PET scanner, where the degradation of the scanner's resolution is mainly an issue towards the periphery of the FOV (Rahmim *et al* 2013). In our numerical simulations, a low-count PET scan with 10 million prompt coincidences was simulated to demonstrate the application of anato-functional priors in especially low-dose or short-time PET scans, where the need for regularization is greater. The geometric component of the PET forward projector was implemented based on a single-ray Siddon's algorithm in C++ with MATLAB (The MathWorks, Inc) interface. The reconstructions were performed on a 12-core Intel Xeon 3.5 GHz workstation with 128 GB RAM with multithreading capability. Nonetheless, owing to the high computational burden of 3D MAP reconstructions, particularly for bias-variance evaluations (see next section), the mMR PET simulations were performed with axial compression (span) of 35, leading to 40%

speed-up in each EM update compared to span 11, which is the default configuration of the mMR scanner. However, it should be emphasized that the simulated sinograms were generated by a PET projector built for span 35 and reconstructed with the same projector. Therefore, there was no resolution and image quality loss due to axial compression of sinogram planes. Note that if the simulated sinograms had been created using a span 1 projector and then compressed to span 35, there would have been resolution loss, however, in our simulation the sinograms were created using a span 35 projector and reconstructed without any axial compression. In supplementary figure 1 (stacks.iop.org/PMB/62/5975/mmedia), simulations with sinograms created using span 11 and span 35 projectors and sinograms created with span 1 and axially compressed to span 11 and 35 has been compared, showing that resolution loss occurs only for axially compressed sinograms. It should be emphasized that these simulations are an approximation of the real data case, but the reconstruction of the idealised span 35 is not approximate, so no resolution degradation occurs. In our workstation, the calculation time for the derivative of the prior (which is performed using a single core) is around 6.5 s, while the reconstruction time for one MAP-EM update for span 11 sinograms is about 2 min, which is reduced by 40% for span 35 sinograms.

2.3.2. Clinical PET-MR datasets. The performance of the included priors was also evaluated for the reconstruction of clinical PET-MR brain datasets, acquired from the Siemens mMR scanner at King's College London & Guy's and St Thomas' PET Centre. In this study, two datasets consisting of [^{18}F]florbetaben and [^{18}F]FDG PET scans were included. The study was approved by the institutional review boards and the research ethics committee. Written informed consent was obtained from all study participants. The [^{18}F]florbetaben was used as a myelin marker for MS patients as these tracers have significant affinity for myelin (Bodini *et al* 2016). The patient referred for myelin imaging was administered 300 MBq [^{18}F]florbetaben and a single-bed dynamic PET scan was immediately started and lasted for about 95 min. A 3D T1-magnetization prepared rapid acquisition gradient recalled echo (MPRAGE) and a dual-point Dixon MR sequence were acquired simultaneously with the PET data and respectively used for diagnostic anatomical imaging and attenuation map generation for PET attenuation correction. The T1-MPRAGE data were acquired using a 5-channel head and neck coil with the following parameters: repetition time (TR): 1700 ms, echo time (TE): 2.63 ms, inversion time (TI): 900 ms, number of averages (NEX): 1, flip angle: 9° , pixel bandwidth of 199 Hz, reconstruction matrix size of $224 \times 256 \times 176$ and voxel size of $1.05 \times 1.05 \times 1.1 \text{ mm}^3$ and. The Dixon data were acquired using the spoiled gradient recalled (SPGR) sequence with the following parameters: T1: 3.6 ms, TE: 2.46 ms, NEX: 1, flip angle: 10° , pixel bandwidth of 946 Hz, reconstruction matrix size of $192 \times 126 \times 128$ and voxel size of $2.06 \times 2.06 \times 3.12 \text{ mm}^3$. In this study, we histogrammed the list-mode data and used the last 5 min of the whole scan with the aim of evaluating the MAP reconstructions in low-count and short-frame PET scans. In the resulting PET emission sinogram, a total count of about 37 million prompts (25 million net trues and 12 million delayed) were recorded. In the second dataset, the patient referred for an [^{18}F]FDG PET scan was administered ~ 200 MBq of [^{18}F]FDG and after 75 min' uptake underwent a 30 min single-bed static PET scan. A total number of counts of about 580 million prompts including 450 million net trues and 130 million delayed were recorded in this scan. Similar to the [^{18}F]florbetaben scan, the T1-MPRAGE and Dixon MR sequences were acquired for this dataset. Figure 3 shows the representative coronal slices of the PET and MR images of the datasets. The florbetaben and FDG images shown in figure 3 were reconstructed using both the vendor-provided reconstruction software (Siemens e7 tools, 3 iterations, 24 subsets, with and without sinogram-based PSF) as well as

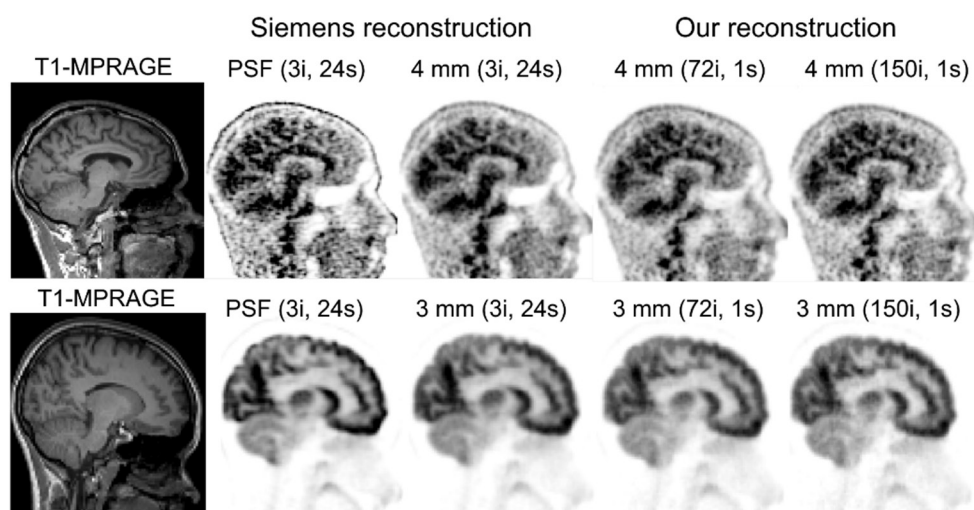


Figure 3. The PET clinical datasets used in this study included [^{18}F]florbetaben (top) and [^{18}F]FDG (bottom) reconstructed by the vendor-provided software (siemens e7 tools, OSEM, 3 iterations 24 subsets (72 updates), with PSF (no filtering) and without PSF (with 4/3 mm post-reconstruction Gaussian filtering) and our own software (MLEM, 72 and 150 iterations, no PSF with 4/3 mm post-reconstruction Gaussian filtering). The T1-MPRAGE MR image of each dataset has also been shown.

our own software (with 72 iterations, 1 subset, and 150 iterations and 1 subset of MLEM, with 4/3 mm post-reconstruction Gaussian kernel).

To evaluate the performance of the MP priors in the case of mismatches between the PET and MR images, two artificial hot lesions with different sizes (small and large) and activity levels were simulated in the FDG PET dataset (see figure 13 in the *Results* section). For this simulation, two lesion masks were drawn on the original lesion-free PET images. The activity of the voxels within the lesion masks were set to that of the corresponding voxels in the original PET images increased by 40% and 80%. The simulated lesions were smoothed by a 2.5 mm Gaussian kernel and then forward projected to obtain a lesion sinogram, which was then attenuated and normalized. A 10 million-count Poisson noise realization was generated from this sinogram. Finally the noisy lesion sinogram was added to the original lesion-free sinogram.

To demonstrate the performance of the anato-functional priors using the proposed joint Burg entropy in low count scans, the full-statistics sinogram of the FDG PET dataset was randomly down-sampled to achieve datasets with 20%, 8%, 3% and 1% of the data, while preserving the Poisson statistics of the data.

2.4. Evaluation metrics

To objectively evaluate the performance of the regularization methods in the reconstruction of simulated datasets, a region-of-interest (ROI) based bias-variance analysis was performed using 10 noise realizations. Three ROIs were considered, each one included all voxels belonging to WM, GM and tumours in the true activity map. The bias was calculated in each ROI as follows:

$$\text{Bias} = 100 \times \frac{1}{N_{\text{ROI}}} \sum_{j \in \text{ROI}} \frac{|\bar{x}_j - x_j^{\text{true}}|}{x_j^{\text{true}}} \quad (18)$$

where $\bar{x}_j = \frac{1}{N_r} \sum_{r=1}^{N_r} x_j^r$ represents the ensemble mean value of each voxel calculated for all N_r noise realizations. N_{ROI} is the total number of voxels in the ROI. \mathbf{x}^{true} is the PET ground truth image. The variance was calculated using the average of the pixel-level coefficient of variation (COV) for each ROI:

$$\text{COV} = 100 \times \frac{1}{N_{\text{ROI}}} \sum_{j \in \text{ROI}} \frac{\sqrt{\frac{1}{N_r-1} \sum_{r=1}^{N_r} (x_j^r - \bar{x}_j)^2}}{\bar{x}_j}. \quad (19)$$

For a single-noise realization, the reconstruction methods were also compared based on the normalized mean square difference (NRMSD) between an image estimate x and the ground truth as follows:

$$\text{NRMSD} = 100 \times \frac{\|\mathbf{x} - \mathbf{x}^{\text{true}}\|_2}{\|\mathbf{x}^{\text{true}}\|_2}. \quad (20)$$

As in clinical datasets, there is no ground truth image we qualitatively evaluated the performance of T1-MR guided and MP MR guided PET image reconstruction compared to the standard MLEM reconstruction.

3. Results

3.1. Simulations

Figure 4 compares the reconstruction results of the brain phantom using the studied algorithms including: the MLEM algorithm with 4 mm post-reconstruction Gaussian filtering, Tikhonov and TV priors and T1-MR anatomical priors, i.e. the Kaipio, Gaussian-P/V, Bowsher and the joint Burg entropy. This figure shows a representative sagittal slice comparing the performance of methods in the recovery of a metabolically active lesion and a region of cortex with mismatched MR anatomical boundaries (see arrows), respectively. In supplementary materials figure 2, the reconstruction results are also presented for other views. As can be seen, the MAP reconstruction methods have substantially reduced noise compared to the MLEM algorithm as the included priors are especially designed to enforce smoothness and, if applicable, to preserve edges. The results show that the Gaussian-V, Bowsher and Burg methods achieve the best performance for the region where the PET and MR image have common boundaries. However, as shown, it is noticeable that the Gaussian-P/V and Bowsher priors tend to completely suppress the PET-unique lesion compared to the Kaipio and JE priors, which also exploit MR anatomical information. As mentioned earlier, for PET regions that do not have corresponding MRI regions, the Kaipio degenerates to the Tikhonov prior, therefore for this simulated lesion, they perform similarly. On the other hand, the Burg prior in essence relies of both MRI and PET information, therefore it is able to preserve PET unique feature. As shown in figure 4, the Gaussian-P prior leads to the blurring of the edges compared to Gaussian-V, which should be attributed to the fact that the similarity kernels are calculated based on patches centred on each voxel, which can rigorously exclude dissimilar voxels (see figures 1(D) and (E)). By proper selection of the σ_u parameter, the Gaussian-V prior can approach the Bowsher

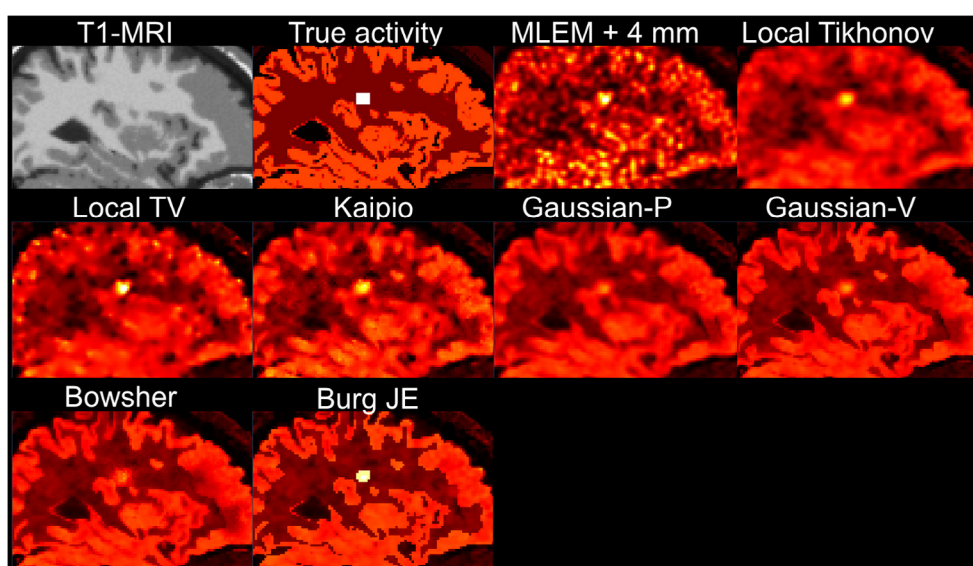


Figure 4. The results of the MAP image reconstructions of the brain phantom using the conventional local and non-local anatomical priors, compared to the MLEM reconstruction followed by Gaussian filtering with kernel width of 4mm. All PET images are shown with the same colour scale.

prior with a given B number. In this experiment, the regularization parameter β was selected based on minimizing the NRMSE over the whole brain.

Figure 5 shows the NRMSE performance of the algorithms versus the regularization parameter for the whole brain as well as the tumours. The optimal β values of 1, 200, 0.1, 8, 8, 10 and 10 were found for Tikhonov, TV, Kaipio, Gaussian-P, Gaussian-V, Bowsher and JE methods, respectively. Supplementary materials figure 3 shows the NRMSE versus the actual values of β . The σ_u parameter of the Gaussian-P/V priors were heuristically set to 0.08. For the Bowsher prior, 70 most similar neighbours out of 342 ($B = 70$) were considered. The σ_u and σ_v parameters of the JE prior were heuristically set to 5 and 1×10^3 . The combination of these parameters were selected such that the reconstructed images are free of artefacts (such as the individual or group of voxels that get extremely higher values and noise-like artefacts), especially in the case of Kaipio and Gaussian priors.

Figure 6 show the results of MP guided PET image reconstructions in the brain phantom dataset for the Gaussian-P/V, Bowsher and Burg JE priors in comparison with T1-MR guided reconstructions. Figure 4 in supplementary materials also shows other views. As shown, there are a missing tumour and a mismatched anatomical region that is partly complemented by the T2-MR image. In the T2-MR image however there are two lesions of which the smaller one matches the PET lesion. The results showed that the inclusion of T2-MR and as well as PET information can lead improved recovery of the lesion in the PET images. As can be seen, the large lesion in T2-MR image has not significantly induced false edges in the MP-guided PET images, however, in all reconstructions there is a faint trace of the inferior edge of the lesion, especially in the JE prior, where noise has been sparsely preserved. Figure 4(B) in supplementary materials also shows that the MP-guided reconstruction improves the performance of the methods in the case of anatomical mismatches, especially for the Bowsher method. Supplementary materials figure 5 compares the activity profiles of the reconstruction methods shown in figures 4 and 6.

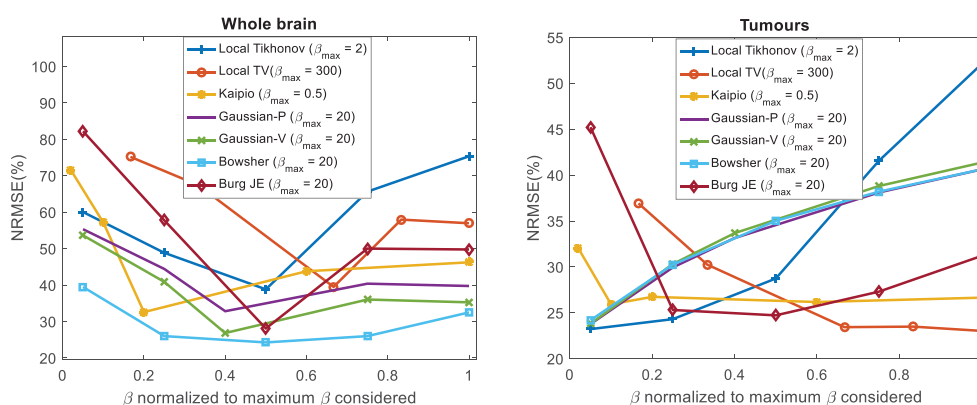


Figure 5. The selection of optimal regularization parameter for the brain phantom based on minimization of the NRMSE over the whole brain (left). The NRMSE results are also shown for the tumours (right).

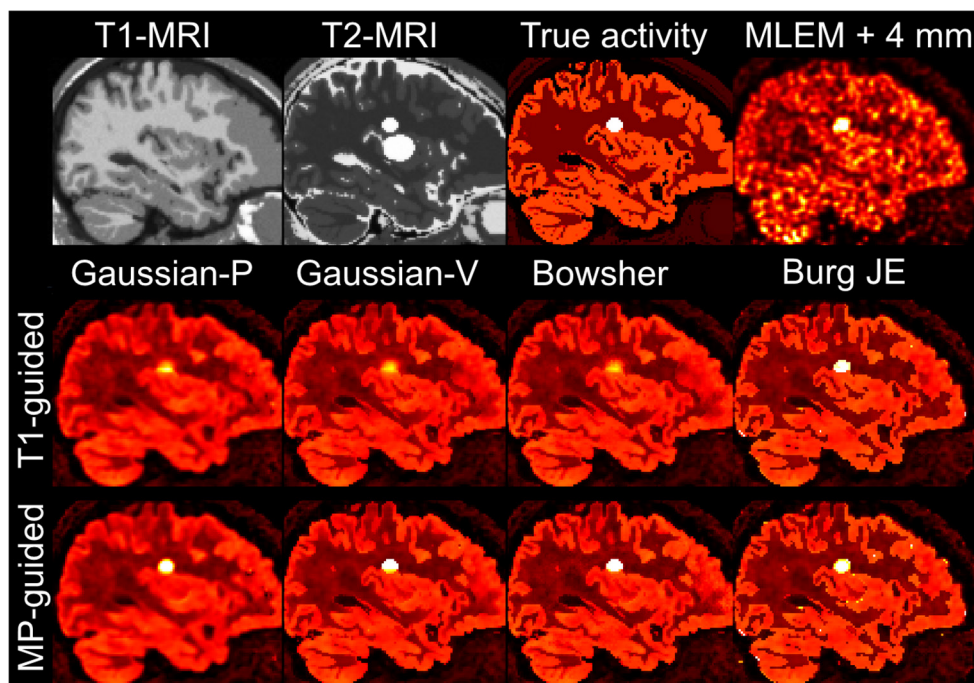


Figure 6. The results of the MAP image reconstructions of the brain phantom using the T1-MR and MP guided anatomical priors. All PET images are shown with the same display window.

Table 3 presents the NRMSE results of the reconstruction methods calculated over GM, WM and tumours of the brain phantom. The results shows that in GM, the Bowsher gives rise to the lowest errors especially with its MP extension. In the WM of the simulated FDG phantom, where there is less uptake, both MP Bowsher and Burg JE prior outperform the other methods, whereas in the tumours, the TV and JE priors achieve the lowest errors. As can be seen, the Bowsher and Gaussian priors lead to the highest NRMSE in the tumours, which is consistent

Table 3. The NRMSE results of the reconstruction methods in the brain phantom using the conventional T1-MR guided anatomical priors and the MP guided anatomical priors (GM: grey matter, WM: white matter). The bold numbers highlight the best two methods in each column.

Methods	GM	WM	Tumours
MLEM	33.63	63.57	25.52
Local Tikhonov	27.49	55.98	26.93
Local TV	31.68	60.31	23.43
Kaipio	22.21	45.72	26.73
Gaussian-P	21.08	44.95	33.14
Gaussian-V	16.48	38.15	33.67
Bowsher	13.17	30.73	35.04
Burg JE	16.30	29.89	24.72
Gaussian-P: MP	20.77	43.57	30.42
Gaussian-V: MP	13.75	32.41	28.99
Bowsher: MP	11.59	24.18	28.70
Burg JE: MP	16.81	24.60	24.70

with findings in figure 4. The ability of the TV prior in preservation of the active tumours should be ascribed to the fact that for the voxels that the magnitude of their local differences is large, the prior assigns lower weights on those differences (see equation (A.2) in appendix A), as a result, the smoothing is suppressed in their neighbourhood. Due to the lower uptake of FDG in WM, the reconstructed images are noisier in this region and hence the NRMSE results of the methods are relatively higher in this region. Figure 7 shows the NRMSE performance of the reconstruction method as a function of iteration in the GM, WM and tumours. The algorithms were further quantitatively compared based on their bias-variance performances using 10 noise realizations of 3D reconstructions. In figure 8, the bias-variance results have been presented for both T1-MR guided and MP-guided priors. The results show that in the grey and WMs the Gaussian-P/V, Bowsher and Burg JE priors present the best bias-variance trade-off compared to the Kaipio and the conventional Tikhonov and TV priors. In these ROIs, the performance of the Bowsher and Burg JE priors is comparable, however, the JE prior results in less bias at the expense of a slightly increased variance due to the presence of high-valued isolated voxels in the images reconstructed when using this prior. As shown, the Gaussian-V method outperforms the Gaussian-P method in terms of bias (as their variance performance is almost the same) but it lags behind the Bowsher method. In the tumour ROI, the Gaussian-V/P and Bowsher methods show the largest bias, as they tend to suppress the PET unique features, while the TV and JE priors achieve the best bias-variance trade-off. The results of MP guided reconstructions show that the bias-variance performance of the MAP reconstruction is slightly improved in terms of bias in WM and GM regions, whereas in the tumour ROI the MP extension of these priors substantially improve their performance, while the performance of the MP-Burg prior nearly remains the same as T1- Burg. Figure 9 shows the impact of MR image quality on the performance of the Kaipio, Bowsher and Burg priors. In this evaluation, the true T1-MR image was smoothed using 3 and 4 mm FWHM Gaussian filters and corrupted by Gaussian noise with standard deviation equal to 1% of largest voxel intensity and the resulting MR images were used for the above-mentioned anatomical priors. In these reconstructions, all the involving parameters were kept constant and the same as before. As could be predicted, the performance of the Kaipio prior is deteriorated by the degradation of MR image quality. With the high-quality and noise-free true MR image, this prior performs closely to the other priors and results in a comparable edge enhancement in the PET image. However, with the low

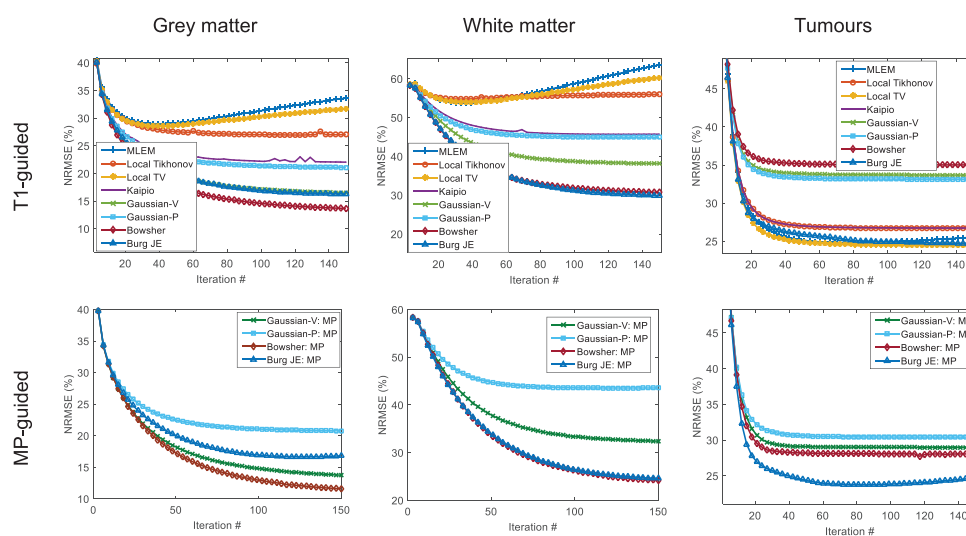


Figure 7. The results of NRMSE for each iteration of the reconstruction methods in the brain phantom.

quality MR images, the reconstructed PET edges are blurred due to the fact that the employed MR normal vectors can no longer distinctly indicate the presence of sharp edges in the MR images and therefore in the PET images. In comparison, the results show the Bowsher and Burg priors are more robust to the quality of the MR images. However, as the boundaries of the MR image get more and more blurred, their corresponding PET boundaries in the images reconstructed by the Bowsher method are also slightly blurred. Whereas in the case of Burg JE method, the sharpness of the edges (or the contrast) remains the same.

3.2. Clinical datasets

3.2.1. Florbetaben dataset. Figure 10 shows the reconstruction results of the [^{18}F]florbetaben dataset using the studied reconstruction methods for a representative sagittal slice. In this experiment, the regularization parameter β was heuristically selected. The β values of 1×10^6 , 1×10^5 , 3×10^5 were set for Tikhonov, TV and Kaipio priors, respectively, while the β value of 2×10^7 was set for Gaussian-P/V, Bowsher and Burg JE methods. As in simulations, the σ_u parameter of the Gaussian-P/V priors were set to 0.08. For the Bowsher prior, the B value was set to 70 as in simulations. The σ_u and σ_v parameters of the JE prior were set to 1 and 10. The results reveal the MLEM reconstruction suffers from a considerable noise and non-uniformity within the WM. The Tikhonov and TV reconstructions can partly reduce noise so that some sulci become more visible. In comparison with these conventional MAP reconstructions, the anatomical MAP methods result in noise reduction and recovery of the anatomical boundaries of the WM given that at late-frame PET scans the florbetaben washes out of GM and accumulates in WM. For this dataset, as can be seen the Kaipio prior gives rise to better reconstruction compared to Gaussian-P method, however, this method leads to a slight underestimation of activity.

In this experiment, all of the reconstructions were terminated after 150 iterations. For higher number of iterations, the results of the Kaipio prior remained the same indicating that the algorithm has reached a stationary point. Consistent with our simulations, the results show

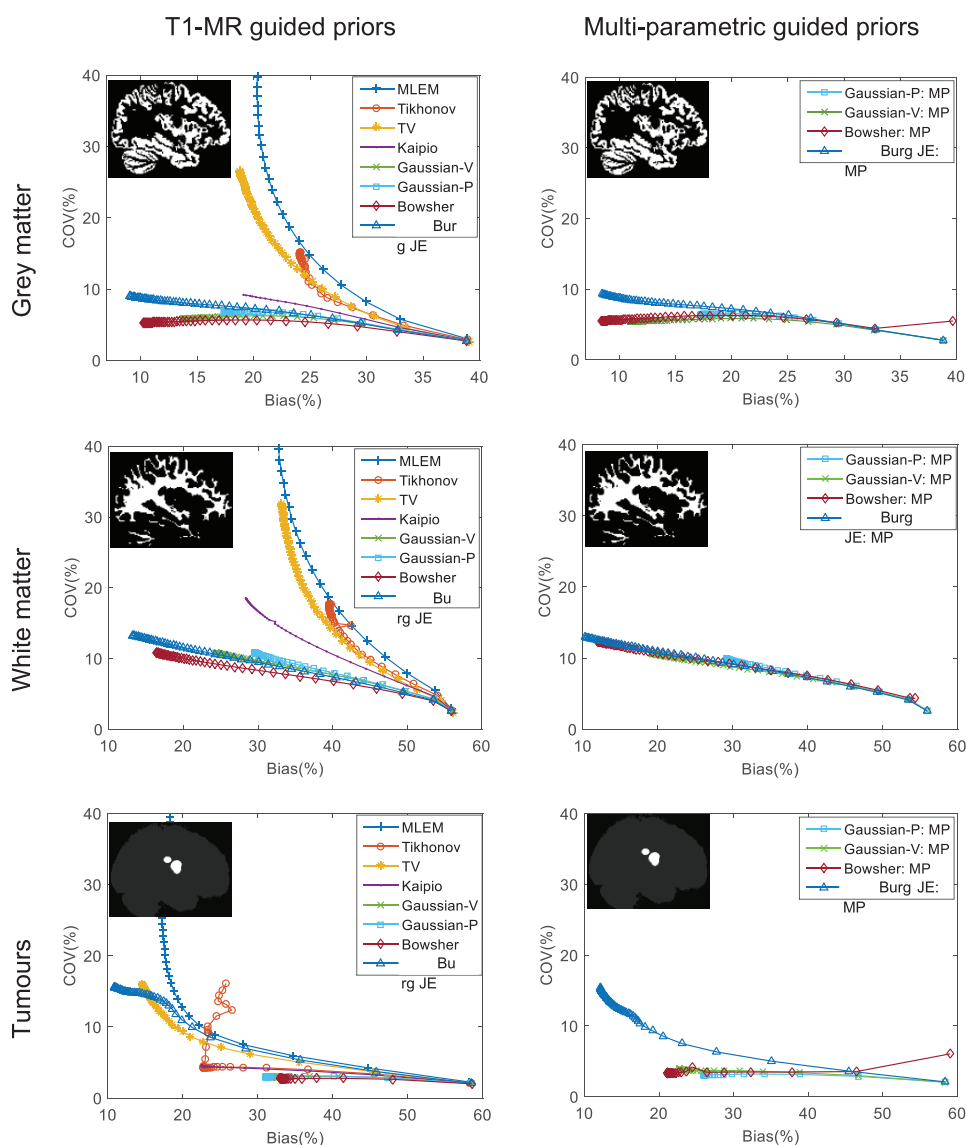


Figure 8. Bias-variance analysis of the reconstruction methods of the brain phantom dataset in GM, WM and tumours. Left and right show the performance of the T1-MR and MP guided priors.

that the Gaussian-P prior is being outperformed by the Gaussian-V prior, which has a very similar performance to the Bowsher prior, of course, depending on the selected parameters. For this dataset, the Bowsher and Burg JE priors demonstrate the best performance in terms of mean activity estimation and the details of the cold GM. The MP extensions of these priors did not reveal substantial differences (results not shown here), since in this dataset the PET tracer is uniformly distributed in the WM and there are not notable mismatches between the PET images and the T1-MR images.

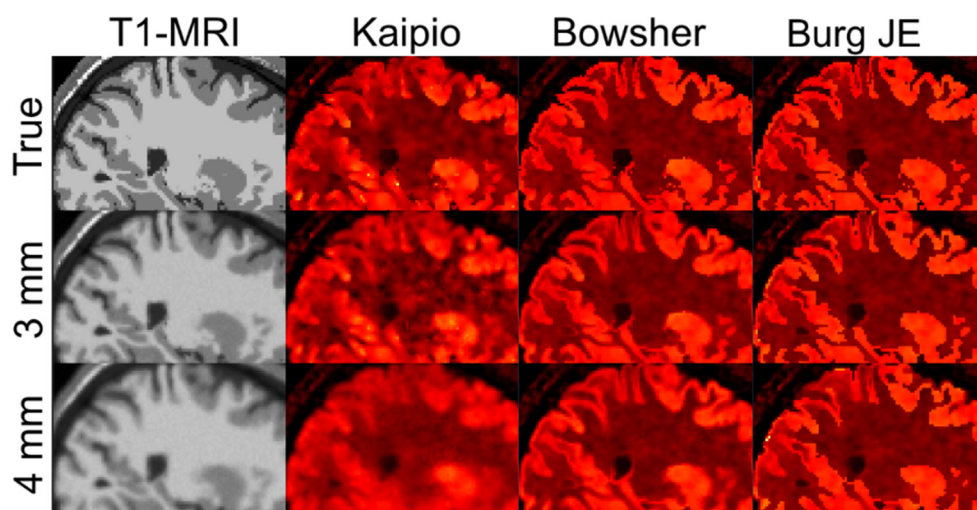


Figure 9. The impact of MR image quality on the performance of the Kaipio, Bowsher and joint Burg entropy anatomical priors. The true MR image was blurred using 3 and 4 mm Gaussian smoothing filter and Gaussian noise was added to the resulting images.

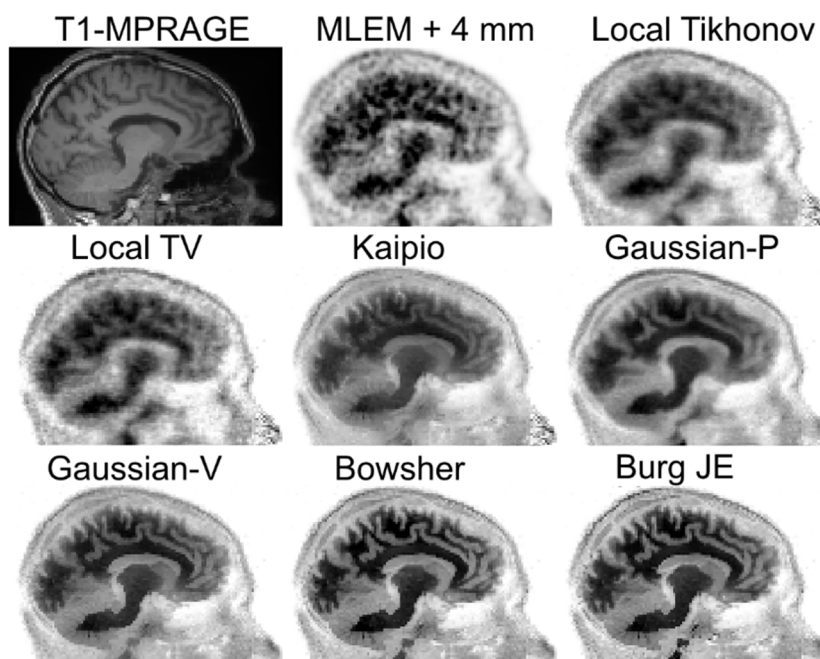


Figure 10. The reconstruction results of the 5 min of the ^{18}F florbetaben dataset using different algorithms. All PET images are shown with the same displaying window.

Figure 11 shows the impact of the shape parameters of the Gaussian-V, Bowsher and Burg priors, as the three best performing priors included in this study, in the anatomical reconstruction of the florbetaben dataset. For the Gaussian-V and JE priors, these shape parameters are the standard deviation of Gaussian kernels used to measure voxel similarities in the MR

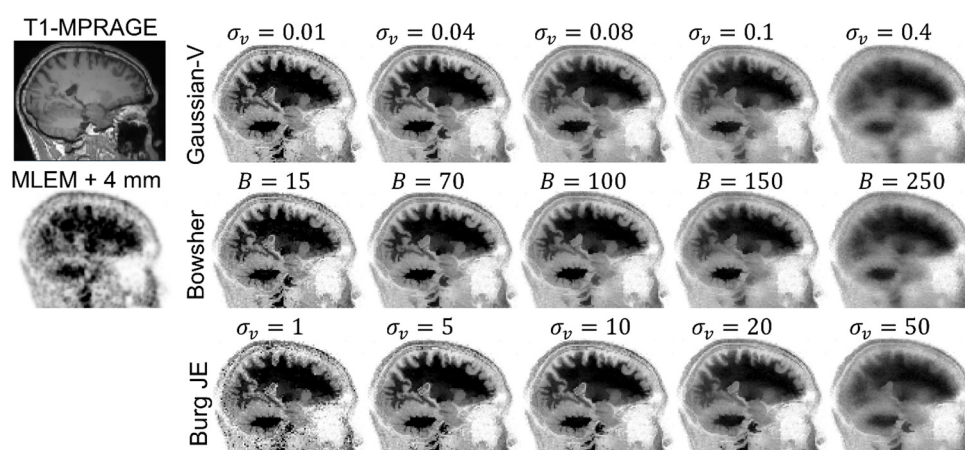


Figure 11. The impact of the shape parameters of the Gaussian-V, Bowsher and Burg JE priors on their reconstruction results.

images, σ_v , while for the Bowsher prior it is the number of most similar neighbours, B . With increasing σ_v , the calculated similarities between a given voxel and its neighbouring voxels is increased, since the Gaussian function would relatively map a given local difference to a higher value. As a result, the similarity kernels become more and more uniform and as can be seen in this figure, the Gaussian-V and Burg priors approach an isotropic Tikhonov prior that suppress both noise and valid edges. Similarly, as the B-value of the Bowsher prior is increased, more and more neighbours are included for the calculation of local differences, as a result, the Bowsher prior also becomes an isotropic Tikhonov prior. It can be seen that for small shape parameters the priors tend to preserve noise and even introduce noise-like artefacts. Note in this experiment, all images were reconstructed with the same regularization parameter, $\beta = 2 \times 10^7$, and shown at the same iteration number $n = 150$. The joint Burg entropy prior has an additional shape parameter defined for the PET images, σ_u , for all JE reconstructions, this parameter was set to 1.

3.2.2. FDG dataset.

3.2.2.1. Partial volume correction (PVC). Figure 12 compares the reconstruction results of the FDG brain dataset using MLEM, conventional MAP and anatomical MAP reconstructions using the T1-MPRAGE MR image. The results demonstrate that the overall resolution and quality of the PET images have been improved by inclusion of anatomical information, especially using the Gaussian-V, Bowsher and JE priors. As a result, the spill-over and partial PVEs visible in the MLEM and Tikhonov and TV reconstructions have been considerably reduced by the anatomical priors. In this experiment, the regularization parameter β of 1×10^4 , 5×10^2 , 1×10^3 was heuristically set for the Tikhonov, TV and Kaipio priors, respectively, while a value of 2×10^5 was set for β for the Gaussian-P/V, Bowsher and Burg JE methods. The σ_u parameter for the Gaussian-P/V priors was set to 0.1. For the Bowsher prior, the B value was set to 70. The σ_u and σ_v parameters of the JE prior were set to 0.1 and 5 respectively.

Figure 13 shows the reconstruction results of the FDG dataset with simulated lesions shown in the T1-MR image. The reconstructions with anatomical priors lead to improved definition of the GM and its separation from the cold WM. Note that the MLEM reconstructions were smoothed using a 3 mm Gaussian kernel compared to the standard 4 mm kernel.

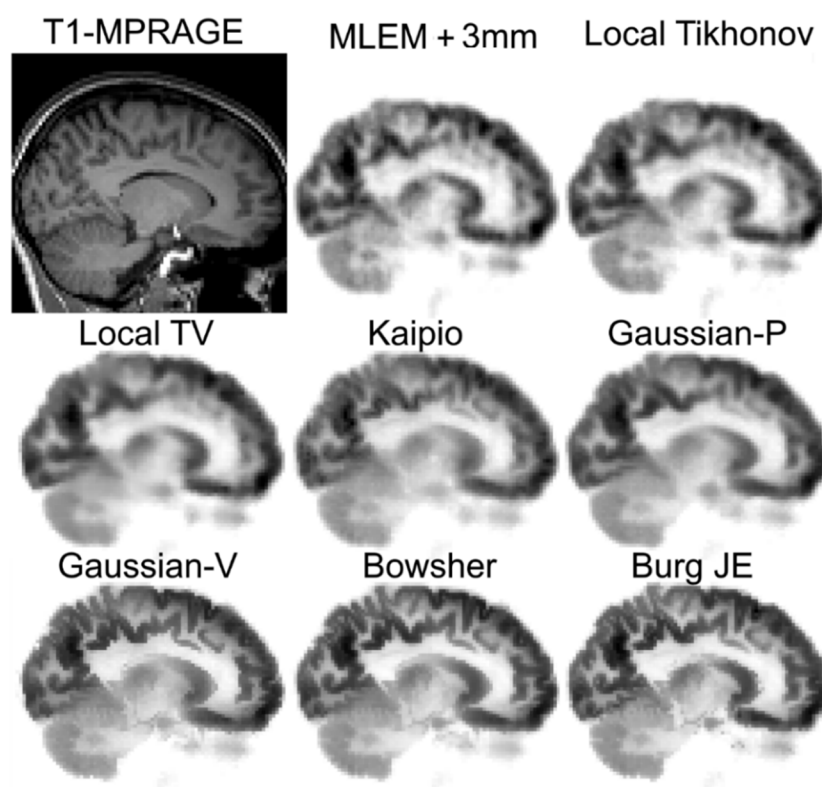


Figure 12. The reconstruction results of the clinical FDG dataset using the different methods.

As a result, there are some residual normalization artefacts in the MLEM images, which are removed in the MAP-EM images as a result of regularization. Figures 13(A) and (B) compare the activity profiles of the reconstructed images along the lines shown on the large and small lesions, respectively. In these profiles, the result of the MLEM reconstruction without post-reconstruction smoothing has also been included. The anatomical priors, Kaipio, Gaussian-P/V and Bowsher notably suppress the activity of the lesions compared to the anato-functional joint Burg entropy prior and conventional priors. These results highlight the value of MP anato-functional priors for improved PET image reconstruction while preserving PET unique lesions.

The MP extension of the Bowsher prior was further evaluated in comparison with the proposed JE prior. Figure 14 compares the PET images reconstructed by MLEM with the cases of using the Bowsher, MP Bowsher and JE priors for two transverse slices through the simulated lesions. In figures 14(A) and (B), the regional mean activity of the tumours at each iteration of the reconstruction algorithms are shown for the large and small tumours, respectively. The images have been shown with the same display range of grey scale intensities in order to highlight their relative performance in preserving the PET unique tumours. As can be seen, the conventional Bowsher prior, relying on the T1-MR image only, has notably reduced the estimated activity of the tumours, especially the smaller one which has been completely suppressed. On the other hand, the MP extension of the Bowsher prior, relying on both the T1-MR image and the PET image reconstructed at each iteration, gives rise to the improved

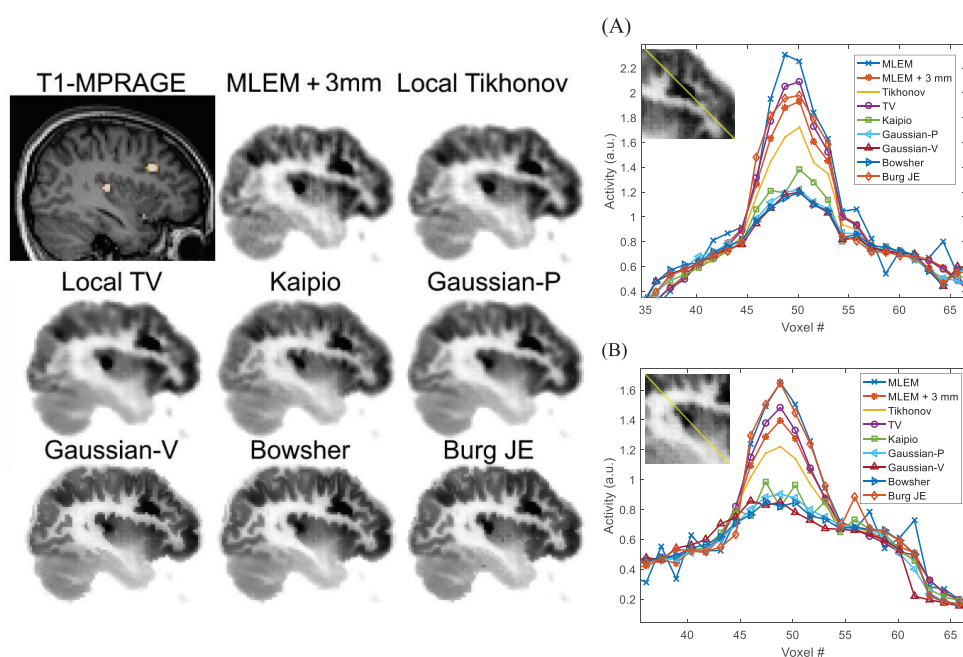


Figure 13. The reconstruction results of the clinical FDG dataset with simulated tumours superimposed on the T1-MR image. The activity profiles of the reconstruction methods has been shown along the indicated lines. The simulated tumours are shown superimposed on the T1-MR image. All PET images are shown with the same grey scale intensity range.

performance of the prior. In comparison, the proposed joint Burg entropy, which by definition relies on both the PET and MR images, is capable of preserving the tumours and at the same time, similar to the Bowsher prior, improving PET image quality and details. In figures 14(A) and (B), the graphs show that the mean activities of the tumours estimated by the Bowsher-MP prior are still far from those estimated by the MLEM algorithm for iterations greater than 70 iterations, where the regional activity has almost converged. However, the results show that compared to post-smoothed MLEM algorithm the JE prior resulted in 5.3% and 9.3% errors in estimation of mean activity of the small and lesion large tumours, respectively, whereas the Bowsher-MP resulted in -25.1% and -24.4% errors in those mean activities, respectively.

3.2.2.2. Improved image quality of low-count PET scans. Apart from PVC of PET images, the MAP image reconstruction using anato-functional priors can be exploited to achieve comparable image quality to that obtained by the conventional MLEM reconstructions but with less data. To demonstrate this new application, the performance of the joint Burg entropy, as the best performing anato-functional prior introduced in this paper, was first evaluated against the MLEM algorithm for reconstruction of the full-statistics FDG dataset for a range of regularization parameters. Figure 15 shows the results for representative transverse and sagittal slices. The MLEM images are shown without post-reconstruction smoothing to reveal the presence of noise and normalization artefacts. Next, the MAP reconstruction with $\beta = 7 \times 10^4$ was selected for comparison with MLEM reconstruction at different count levels. The reason for choosing this β among those considered was in order to obtain images comparable to the MLEM reconstruction of high-count data.

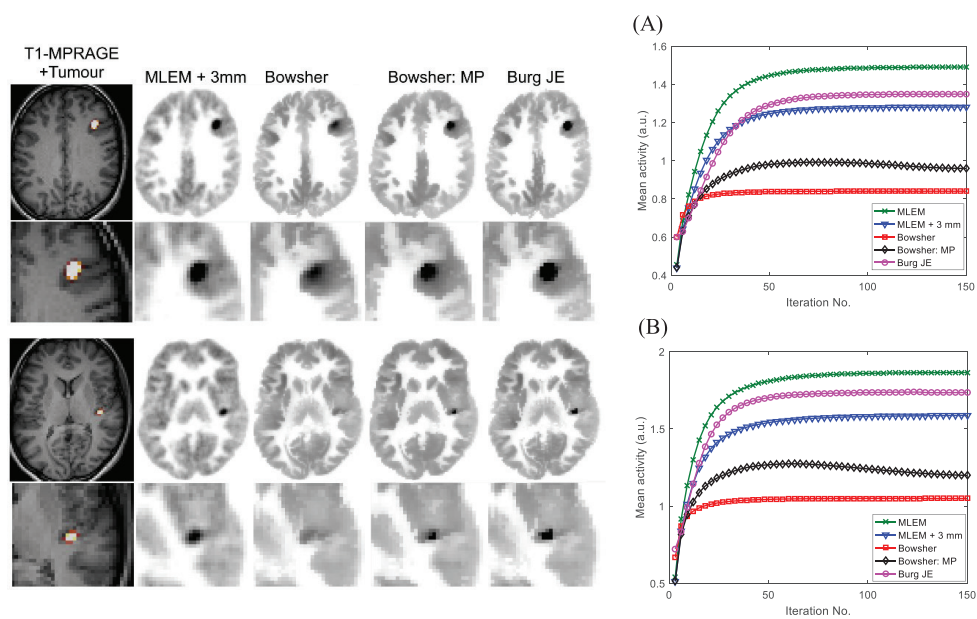


Figure 14. Comparison of anatomical and MP Bowsher priors with the proposed joint Burg entropy prior in the recovery of the PET unique tumours, simulated in the clinical FDG dataset. (A) and (B) The mean activity of the large and small tumours per iteration of the studied reconstruction methods in the top and bottom panels, respectively. All PET images are displayed using the exact same grey scale range of intensities.

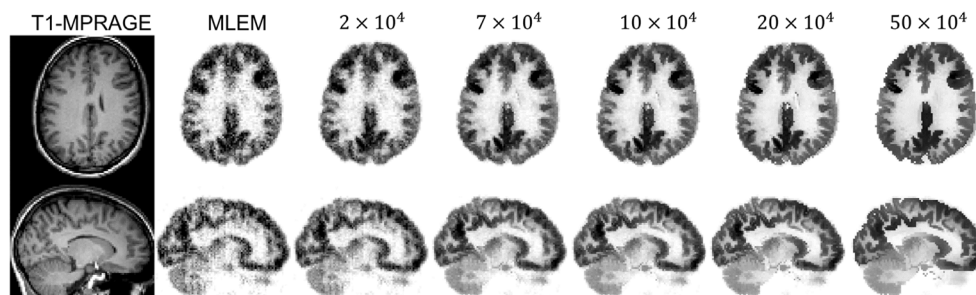


Figure 15. The impact of regularization parameter, β , on the quality of the MAP reconstruction of the FDG dataset using the joint Burg entropy prior in comparison with the MLEM algorithm (which is equivalent to $\beta = 0$).

As described before, the full-statistics sinogram data were randomly down-sampled to achieve datasets with 20%, 8%, 3% and 1% of the data. The datasets were then reconstructed using the MLEM and MAP-EM algorithms. Figure 16 shows the reconstruction results for different percentages of the data. Figure 6 in the supplementary materials compares the images in representative transverse slices. The MLEM algorithm post-smoothed with Gaussian kernels with different FWHMs are also shown. Since the MAP reconstructions with the considered β -value were still suffering from noise, they were post-smoothed using a 2.5 mm Gaussian kernel. The β -values used for the 20%, 8%, 3% and 1% datasets were respectively chosen to be 3.5×10^5 , 8.5×10^5 , 3.5×10^6 and 2×10^7 , which were estimated by scaling

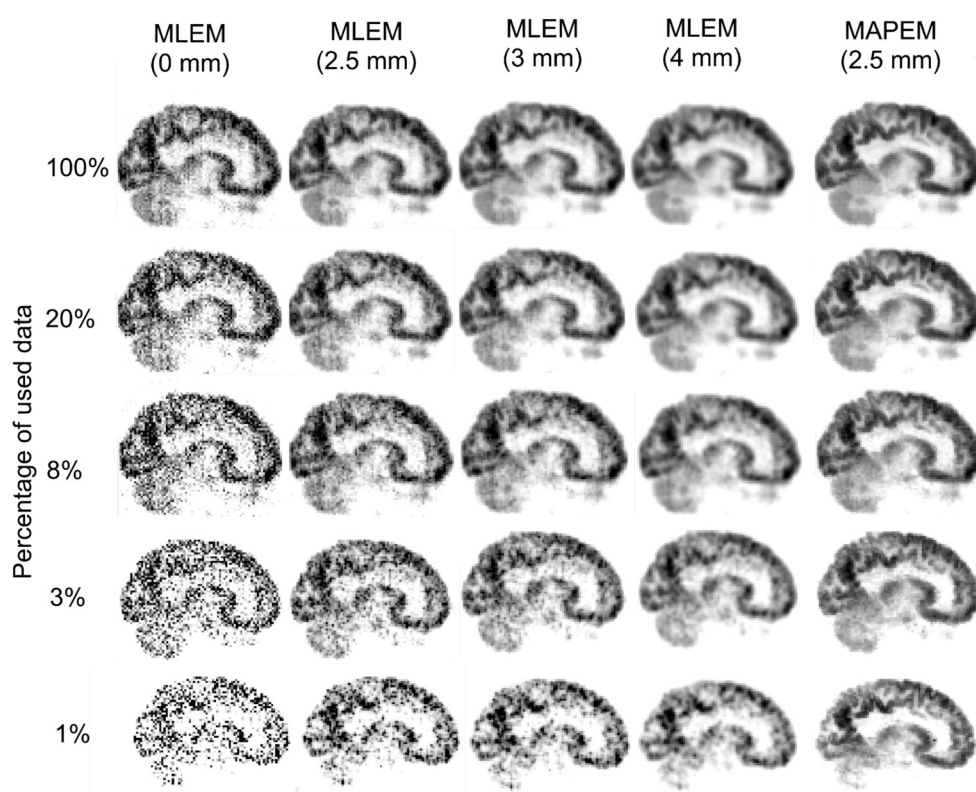


Figure 16. The reconstruction of various percentages of the FDG PET data using the MLEM and MAP-EM algorithms. The MLEM reconstructions have each been post-smoothed using a Gaussian kernel with 0, 2.5, 3 and also 4 mm FWHM. The MAP-EM reconstructions were performed using the proposed joint Burg entropy prior and post-smoothed using a 2.5 mm Gaussian kernel. All PET images are shown with the same grey scale intensity range.

the considered β -value used for the full-statistics data (i.e. $\beta = 7 \times 10^4$) with the ratio of the total counts in the full-statistics to that of the down-sampled sinograms. The reconstructed images were scaled by the same ratio used as a calibration factor to account for differences in total count level. As a result, all images in figure 16 were displayed using the same grey scale intensity range. For a given percentage of used data, for instance 100%, as the Gaussian kernel's FWHM is increased, the noise level of the MLEM images is reduced, however, at the expense of resolution degradation. For the 100% dataset, post-smoothing with a 3 mm Gaussian kernel appears to give the best compromise between noise and resolution, while for other datasets, broader kernels are required to reduce noise. On the other hand, the results of the MAP reconstructions show that the images reconstructed by very low amounts of data are fairly comparable to the MLEM reconstructions with higher count levels. For instance, the MAP-EM with 8% of the data is comparable to, or for some details even better than, the MLEM + 3 mm smoothing when using only 20% of the data (see also supplementary materials figure 6 for this comparison). For very low-count levels, the results show that the MAP reconstruction is capable of the recovery of the details that are lost even with higher count levels, as shown for instance in the results using just 1% of the data.

4. Discussion

4.1. Performance of the anatomical priors

Our realistic 3D simulation results, presented in figure 4 and supplemental materials figure 1, showed that for the conventional MR-guided PET image reconstruction, the Gaussian-P/V and Bowsher methods lead to a substantial suppression of PET unique lesions and boundaries. In comparison, the proposed joint Burg entropy prior not only can well preserve the PET boundaries with matched MR anatomical boundaries but also the PET unique lesions. This promising aspect of the joint priors is brought by the fact that they rely on the joint distribution of MR and PET images, which leads to the exploitation of both PET and MR information in the identification of both common and unique features. In addition, the results presented in figure 6 and supplemental materials figure 3 showed that the proposed MP extension of the Gaussian and Bowsher priors can significantly improve the performance of these priors in the preservation of PET lesions and boundaries. The NRMSE results presented in figure 7 show that the MP anato-functional extension is more beneficial for the Gaussian and Bowsher priors than that of the JE prior. For instance, for tumour ROIs, this extension leads to 14%, 15%, 20% and 10% reduction of errors for the Gaussian-P, Gaussian-V, Bowsher and Burg priors, respectively.

The images reconstructed by the Burg prior show some isolated voxels with high intensity values, especially in its MP extended version. This issue has also been previously reported for the case of the original Shannon JE prior in Somayajula *et al* (2011) and is evident from the results in Tang and Rahmim (2009). As pointed out in Somayajula *et al* (2011), for low count datasets, the prior tends to increase the variance in the distribution of PET image intensities, thus causing isolated high-valued voxels. However, depending on the selection of the regularization parameters and Parzen bandwidths the occurrence of these voxels can be reduced.

The reconstruction of clinical datasets using the anatomical priors also showed the clinical feasibility of these advanced reconstruction methods especially in low-count and short-frame PET scans, where the conventional MLEM reconstruction fails to fully recover the anato-functional details. In the reconstruction of the florbetaben and FDG PET datasets, it was found that the anatomical MAP reconstructions notably compensate for partial volume effect in the estimated uptake of the white and GMs, respectively.

In terms of computational complexity, the MP version of the Gaussian-P prior is the most time consuming prior since the PET component of the similarity kernels needs to be recalculated after each iterative update for a small patch around each voxel in a large neighbourhood. The MP version of the Gaussian-V, Bowsher and the joint Burg entropy priors come second in terms of computational cost, and require the same computations in the recalculation of the PET part of the voxel-based similarity kernels. Finally, the similarity kernels for conventional Bowsher and Gaussian-P/V priors can be precomputed from the MR images and so does not add substantial computational load to the reconstruction.

4.2. Comparison with the prior work

In Vunckx *et al* (2012), three different anatomical priors, including an MR segmentation-based prior, the Bowsher and the Shannon JE priors, were evaluated in brain PET imaging. The authors reported that in comparison with the Bowsher prior, the Shannon JE often converges to a local maximum and its performance highly depends on the section of the many involved parameters (i.e. the number of bins by which the joint probability distribution is integrated, bandwidths of the Parzen windows, weight of the prior). To avoid convergence to

an undesirable local maximum, the author proposed the initialization of the JE reconstruction using an image reconstructed by a few iterations of the Bowsher-MAP method or the gradual increase of the regularization parameter. It was concluded that the Bowsher prior can fairly outperform the Shannon JE prior as it behaves more predictably and its parameters can be much more easily selected, which make this prior practical for clinical routines. Moreover, it does not require the segmentation of MR images, as the segmentation-based priors heavily rely on the accuracy of the segmentation, especially in the case of lesions and abnormalities. In comparison, our results showed that the proposed Burg JE prior has a number of favourable properties. As shown in figure 8, this prior has very similar convergence properties as the Bowsher method in the anatomical regions (i.e. WM and GM). Note all of our reconstructions were initialized using a uniform initial guess. Based on the results in figures 6 and 8, it performs quite similar to or even slightly better than Bowsher prior in WM and GM and more importantly it outperforms the Bowsher and other priors in the reconstruction of PET unique lesions and edges. Compared to the Shannon JE, the proposed Burg JE prior is free of the selection of number of bins and computationally much less expensive, as the prior is degenerated to a weighted non-local Tikhonov prior, where the weights can be precomputed for multi-contrast MR images.

In this work, the performance of the Kaipio prior, as a new class of anatomical priors were also studied. As we elaborated, this prior can also be interpreted as a non-local Tikhonov prior in which the similarity kernels are calculated based on MR normal vectors (see equation (15)). Our results in figure 9 showed that the performance of this prior highly depends on the quality of the MR image as it relies on MR normalized gradient vectors. The normalization of this vectors makes them independent of their magnitude but very sensitive to noise and blurring artefacts. In Ehrhardt *et al* (2016), it was shown that the Kaipio prior can outperform the Bowsher prior in 2D PET image reconstruction for a 3D 3×3 neighbourhood. However, our 3D realistic reconstruction results showed that this prior is far being outperformed by the Bowsher prior for a 3D $7 \times 7 \times 7$ neighbourhood. This could be ascribed to the size of the neighbourhood and the selected number of most similar neighbours. As pointed out in Vunckx *et al* (2012), the Bowsher prior with larger neighbourhoods results in improved results compared to smaller ones. In Vunckx *et al* (2012) and Ehrhardt *et al* (2016) and our study, this prior was evaluated using 4 neighbours in a 3×3 neighbourhood, 80 neighbours in a $5 \times 5 \times 5$ neighbourhood and 70 neighbours in a $7 \times 7 \times 7$ neighbourhood respectively, which can explain the improved performance of the Bowsher in our results. Note that in this study, we did not include the parallel level set anatomical prior proposed in Ehrhardt *et al* (2016), as similar to TV prior, it is not continuously differentiable and its derivation is not as straightforward as the Kaipio or a quadratic prior which can be suitably employed in one-step-plate optimization of the MAP-EM algorithm. In Ehrhardt *et al* (2016), the authors employed a limited-memory BFGS, quasi-Newton type algorithm for the optimization.

In this study, we explicitly emphasized two applications of anato-functional priors for PVC of PET data as well as reducing PET scan time or alternatively reducing the injected dose of radiotracer. For low-dose PET scans, TV regularization has been the primary focus of study in the literature, however as can be seen in Müller *et al* (2011), at such low-statistics the reconstructed images suffer from stair-casing artefacts in spite of improved signal to noise ratio. In this study, a smoothed TV prior was included due to the differentiability requirement of the OSL optimization algorithm, therefore those artefacts were eliminated at the expense of reducing the edge-preserving properties of the TV prior. Nonetheless, in this study, we have highlighted low-dose PET image reconstruction using anato-functional priors, which can find immediate application in dynamic or follow-up PET scans.

4.3. Parameter selection

The performance of the MAP image reconstruction highly depends on the selection of the involving hyper-parameters, which in turn depends on the task of the image reconstruction, that is, lesion detection, tracer quantification, and etc. These parameters in fact control the properties of the prior functions and their impact on the image being reconstructed. In this study, the shape parameters of the studied priors, listed in table 2, were subjectively selected, while for simulation the β parameter was objectively selected with respect to a ground truth. As could be expected, in our experiment, we found that the performance of the MAP reconstruction highly depends on the regularization parameter, β , which varies depending on the type of prior and the object's activity level, scan duration and body weight. With increasing β and thus the strength of the anatomical prior, the PET edges that have matched MR edges are strongly preserved while the PET unique edges are suppressed. Therefore, it is expected that with a lower β , the bias-variance performance of the priors in tumour ROIs can change, however at the expense of increased variance in GM and WM. Nonetheless, with the MP extension of the prior, it is possible to still use high-valued β to reduce noise, yet better preserve PET unique features. Another important parameter that is specific to the non-local prior is the size of the neighbourhood. In Vunckx *et al* (2012), it is been reported that the performance of the Bowsheer prior is improved with increasing the neighbourhood size, which can be attributed to the fact that in larger neighbourhoods the similarity between voxels and the continuity of edges are better captured. However, the suppression of PET unique features is presumably reduced with smaller neighbourhoods. Additionally, most of the anatomical priors included in this study have additional shape parameters that considerably affect the quality of the reconstructed images, as shown in figure 11. These B , σ_v parameters are defined for the MR part of the similarity kernels prior to the PET reconstruction. In this study, we also predefined the σ_u parameter which is related to the PET part of the kernels. Adaptive selection of this parameter has also been explored in Tang and Rahmim (2009) based on the intensity of the PET image estimates at each iteration. Some of these parameters such as neighbourhood size \mathcal{N} , feature vector size, \mathcal{F} , standard deviation of MR Gaussian kernels, σ_v , can be fixed and standardized between patients, however the others such as the regularization parameter, β , and standard deviation of PET Gaussian kernels, σ_u , should be adjusted for each patient based on the count level, particularly β .

4.4. Future work

In this work, we studied the Gaussian similarity kernels as a commonly used radial basis function (RBF). However, other RBFs such as the Bessel function, multiquadrics, thin-plane splines and Matern can also be investigated in order to find the best kernel with least shape parameters Fasshauer and Zhang (2007). The proposed MP regularization method can also be evaluated for post-reconstruction denoising of the PET images (Turkheimer *et al* 2008, Chan *et al* 2014), MR-guided deconvolution of PET images (Yan *et al* 2015) or multi-scale resolution recovery methods (Shidahara *et al* 2009) in order to preserve the PET unique features. Finally, future work includes the assessment of the added value of the MP MAP PET image reconstruction in PVC and improved quantification of tracer uptake in brain imaging of patients suspected of Alzheimer's disease using a large patient population. In Loeb *et al* (2015), the anatomical Bowsheer prior has been successfully applied for low-dose dynamic PET-MR image reconstruction with matched PET and MR lesions in both simulation and clinical datasets. Hence, the improved performance of the proposed MP joint Burg entropy prior remains to be demonstrated in dynamic PET-MR scans where there are mismatched anato-functional features.

5. Conclusion

In this work, several state-of-the-art anatomical priors were studied and extended to MP anato-functional priors in a common framework. Moreover, a modified MP joint Burg entropy prior was introduced. In addition to PVC, application of anato-functional priors to the reconstruction of low-count PET data for achievement of comparable image quality to that obtained by MLEM reconstruction with high-count data was demonstrated. In both our simulation and clinical results, the conventional anatomical priors resulted in the suppression of PET unique features, which was notably reduced by the MP extension of these priors. The results showed that the Tikhonov priors with Gaussian similarity kernels, calculated using voxel-based feature vectors and with the Bowsher similarity kernels and the proposed prior result in the most accurate recovery of PET details. It was also found that the proposed prior is more robust in preserving PET unique features.

Acknowledgments

This work is supported by Engineering and Physical Sciences Research Council (EPSRC) under grant EP/M020142/1. The authors would like to thank Jane MacKewn and Colm McGinnity from King's College London & Guy's and St Thomas' PET Centre for provision of the clinical brain PET-MRI raw data, Radhouene Neji from Siemens Medical Solutions for providing support with VB e7 tools and Sam Ellis for his consultation on down-sampling of the FDG PET raw data. According to EPSRC's policy framework on research data, all simulation and clinical results supporting this study will be openly available at <https://doi.org/10.5281/zenodo.580304>.

Appendix A

The smoothed TV prior is defined as:

$$R(\mathbf{u}) = \sum_j^N \sqrt{\sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b)^2 + \delta^2}. \quad (\text{A.1})$$

Using the chain rule of differentiation of a composite function, the first derivative of this prior is given by:

$$\frac{\partial R}{\partial u_j} = \sum_{b \in \mathcal{N}_j} \frac{\xi_{jb} (u_j - u_b)}{\sqrt{\sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b)^2 + \delta^2}}. \quad (\text{A.2})$$

Appendix B

The partial derivative of the JE prior defined in (13) with respect to u_j is given by:

$$\frac{\partial H}{\partial u_j} = \frac{1}{N} \sum_{b=1}^N \frac{\frac{1}{N} \mathcal{G}(u_j, u_b, \sigma_u) \mathcal{G}(v_j^{(1)}, v_b^{(1)}, \sigma_{v^{(1)}}) \dots \mathcal{G}(v_j^{(P)}, v_b^{(P)}, \sigma_{v^{(P)}})}{p(u_j, v_j^{(1)}, \dots, v_j^{(P)})} \frac{1}{\sigma_u^2} (u_j - u_b). \quad (\text{B.1})$$

The derivative of JE evaluated at j th voxel depends on all voxel intensities simultaneously, which leads to discarding regional information. As suggested in Somayajula *et al* (2011), the above derivative can be approximated by evaluating the summation for the voxels that are in the neighbourhood of the j th voxel, that is:

$$\frac{\partial H}{\partial u_j} = \frac{1}{N} \sum_{b \in \mathcal{N}_j} \frac{1}{\sigma_u^2} \omega_{jb} (u_j - u_b) \tag{B.2}$$

where ω_{jb} are normalized Gaussian weighting coefficients given by:

$$\omega_{jb} = \frac{\mathcal{G}(u_j, u_b, \sigma_u) \mathcal{G}(v_j^{(1)}, v_b^{(1)}, \sigma_{v^{(1)}}) \dots \mathcal{G}(v_j^{(P)}, v_b^{(P)}, \sigma_{v^{(P)}})}{\sum_b \mathcal{G}(u_j, u_b, \sigma_u) \mathcal{G}(v_j^{(1)}, v_b^{(1)}, \sigma_{v^{(1)}}) \dots \mathcal{G}(v_j^{(P)}, v_b^{(P)}, \sigma_{v^{(P)}})} \tag{B.3}$$

Appendix C

The Kaipio prior reformulated in Ehrhardt *et al* (2016) can be defined as following:

$$R(\mathbf{u}) = \sum_j \left\| \tilde{\xi}_j \mathbf{g}_j \right\|_2^2 - \left\langle \tilde{\xi}_j \mathbf{g}_j, \mathbf{n}_j \right\rangle^2 \tag{C.1}$$

where $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ and $\mathbf{x}, \mathbf{y} = \sum_i x_i y_i = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \cos(\theta)$. In this prior, the L_2 norm-square of the weighted gradient vector, $\mathbf{g}_j = [u_j - u_1, \dots, u_j - u_D]^T$, of the image \mathbf{u} at j th voxel is subtracted by the square of the inner product of the normal vectors $\mathbf{n}_j = [n_{j1}, \dots, n_{jD}]^T$ and the vector $\tilde{\xi}_j \mathbf{g}_j$, in which the elements of \mathbf{g}_j are weighted by $\tilde{\xi}_j = \text{diag} \{ \sqrt{\xi_{jb}} \}$, where ξ_{jb} are the proximity weighting coefficients. The (C.1) can be expanded as:

$$R(\mathbf{u}) = \sum_j \sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b)^2 - \left(\sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}} (u_j - u_b) \right)^2 \tag{C.2}$$

By the expansion of the second term in (C.1) and knowing that $\|\mathbf{n}_j\|_2 = 1$, one can also show that this prior can be formulated as in (9), that is:

$$\begin{aligned} R(\mathbf{u}) &= \sum_j \left\| \tilde{\xi}_j \mathbf{g}_j \right\|_2^2 - \left\| \tilde{\xi}_j \mathbf{g}_j \right\|_2^2 \|\mathbf{n}_j\|_2^2 \cos^2(\theta_j) \\ &= \sum_j (1 - \cos^2(\theta_j)) \left\| \tilde{\xi}_j \mathbf{g}_j \right\|_2^2 \\ &= \sum_j \sin^2(\theta_j) \sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b)^2 \end{aligned} \tag{C.3}$$

where θ_j is the angle between the vectors \mathbf{g}_j and \mathbf{n}_j . The derivative of the prior defined in (C.2) is given by:

$$\begin{aligned} \frac{\partial R(\mathbf{u})}{\partial u_j} &= 2 \sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b) - 2 \left(\sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}} \right) \sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}} (u_j - u_b) \\ &= 2 \sum_{b \in \mathcal{N}_j} \xi_{jb} (u_j - u_b) - 2 \sum_{b \in \mathcal{N}_j} \left(\sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}} \right) n_{jb} \sqrt{\xi_{jb}} (u_j - u_b) \\ &= 2 \sum_{b \in \mathcal{N}_j} \xi_{jb} \left(1 - n_{jb} \frac{1}{\sqrt{\xi_{jb}}} \sum_{b \in \mathcal{N}_j} n_{jb} \sqrt{\xi_{jb}} \right) (u_j - u_b) \end{aligned} \tag{C.4}$$

ORCID

Abolfazl Mehranian  <https://orcid.org/0000-0003-4584-4453>

References

- Belzunce M A and Reader A J 2016 Time-invariant component-based normalization for a simultaneous PET-MR scanner *Phys. Med. Biol.* **61** 3554–71
- Bodini B et al 2016 Dynamic imaging of individual remyelination profiles in multiple sclerosis *Ann. Neurol.* **79** 726–38
- Bowsher J E et al 2004 Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors *IEEE Nuclear Science Symp. and Medical Imaging Conf.* pp 2488–92
- Buades A, Coll B and Morel M J 2005 A review of image denoising algorithms, with a new one *Multiscale Model. Simul.* **4** 490–530
- Byrne C L 1993 Iterative image reconstruction algorithms based on cross-entropy minimization *IEEE Trans. Image Proc.* **2** 96–103
- Chan C, Fulton R, Barnett R, Feng D D and Meikle S 2014 Postreconstruction nonlocal means filtering of whole-body PET with an anatomical prior *IEEE Trans. Med. Imaging* **33** 636–50
- Chen Y, Ma J, Feng Q, Luo L, Shi P and Chen W 2008 Nonlocal prior bayesian tomographic reconstruction *J. Math. Imaging Vis.* **30** 133–46
- Cocosco C A, Kollokian V, Kwan R K-S and Evans A C 1997 BrainWeb: online interface to a 3D MRI simulated brain database *NeuroImage* **5** S425
- De Pierro A R 1995 A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography *IEEE Trans. Med. Imaging* **14** 132–7
- Dutta J, Leahy R M and Li Q 2013 Non-local means denoising of dynamic PET images *PLoS One* **8** e81390
- Ehrhardt M, Markiewicz P, Liljeroth M, Barnes A, Kolehmainen V, Duncan J, Pizarro L, Atkinson D, Hutton B and Ourselin S 2016 PET Reconstruction with an anatomical MRI prior using parallel level sets *IEEE Trans. Med. Imaging* **35** 2189–99
- Fasshauer G E and Zhang J G 2007 On choosing ‘optimal’ shape parameters for RBF approximation *Numer. Algorithms* **45** 345–68
- Gindi G, Lee M, Rangarajan A and Zubal I G 1993 Bayesian reconstruction of functional images using anatomical information as priors *IEEE Trans. Med. Imaging* **12** 670–80
- Green P J 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm *IEEE Trans. Med. Imaging* **9** 84–93
- Huber P J 1981 *Robust Statistics* (New York: Wiley)
- Kaipio J P, Kolehmainen V, Vauhkonen M and Somersalo E 1999 Inverse problems with structural prior information *Inverse Probl.* **15** 713–29
- Kazantsev D, Bousse A, Pedemonte S, Arridge S R, Hutton B F and Ourselin S 2011 Edge preserving bowsher prior with nonlocal weighting for 3D SPECT reconstruction *IEEE Int. Symp. on Biomedical Imaging* pp 1158–61
- Kazantsev D, Lionheart W R, Withers P J and Lee P D 2014 Multimodal image reconstruction using supplementary structural information in total variation regularization *Sens. Imaging* **15** 97
- Lange K 1990 Convergence of EM image reconstruction algorithms with Gibbs smoothing *IEEE Trans. Med. Imaging* **9** 439–46
- Loeb R, Navab N and Ziegler S I 2015 Direct parametric reconstruction using anatomical regularization for simultaneous PET/MRI data *IEEE Trans. Med. Imaging* **34** 2233–47
- Lysaker M, Osher S and Tai X C 2004 Noise removal using smoothed normals and surface fitting *IEEE transactions on image processing: a publication of the IEEE Signal Process. Soc.* **13** 1345–57
- Müller J, Brune C, Sawatzky A, Kösters T, Schäfers K P and Burger M 2011 Reconstruction of short time PET scans using Bregman iterations *IEEE Nuclear Science Symp. and Medical Imaging Conf.* pp 2383–5
- Nguyen V G and Lee S J 2013 Incorporating anatomical side information into PET reconstruction using nonlocal regularization *IEEE Trans. Image Process.* **22** 3961–73
- Noll G 1999 *Radiation Detection and Measurement* (New York: Wiley)

- Nuyts J 2007 The use of mutual information and joint entropy for anatomical priors in emission tomography *IEEE Nuclear Science Symp. and Medical Imaging Conf.* pp 4149–54
- Pluim J P, Maintz J B and Viergever M A 2003 Mutual-information-based registration of medical images: a survey *IEEE Trans. Med. Imaging* **22** 986–1004
- Rahmim A, Qi J and Sossi V 2013 Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls *Med. Phys.* **40** 064301
- Reader A J, Ally S, Bakatselos F, Manavaki R, Walledge R J, Jeavons A P, Julyan P J, Zhao S, Hastings D L and Zweit J 2002 One-pass list-mode EM algorithm for high-resolution 3D PET image reconstruction into large arrays *IEEE Trans. Nucl. Sci.* **49** 693–9
- Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68
- Shidahara M, Tsoumpas C, Hammers A, Boussion N, Visvikis D, Suhara T, Kanno I and Turkheimer F E 2009 Functional and structural synergy for resolution recovery and partial volume correction in brain PET *Neuroimage* **44** 340–8
- Somayajula S, Panagiotou C, Rangarajan A, Li Q, Arridge S R and Leahy R M 2011 PET image reconstruction using information theoretic anatomical priors *IEEE Trans. Med. Imaging* **30** 537–49
- Tang J and Rahmim A 2009 Bayesian PET image reconstruction incorporating anato-functional joint entropy *Phys. Med. Biol.* **54** 7063–75
- Tomasi C and Manduchi R 1998 Bilateral filtering for gray and color images *6th Int. Conf. on Computer Vision (ICCV)* pp 839–46
- Turkheimer F E, Boussion N, Anderson A N, Pavese N, Piccini P and Visvikis D 2008 PET image denoising using a synergistic multiresolution analysis of structural (MRI/CT) and functional datasets *J. Nucl. Med.* **49** 657–66
- Vunckx K and Nuyts J 2010 Heuristic modification of an anatomical Markov prior improves its performance *IEEE Nuclear Science Symp. and Medical Imaging Conf.* pp 3262–6
- Vunckx K, Atre A, Baete K, Reilhac A, Deroose C M, Van Laere K and Nuyts J 2012 Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging *IEEE Trans. Med. Imaging* **31** 599–612
- Wang G and Qi J 2012 Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization *IEEE Trans. Med. Imaging* **31** 2194–204
- Watson C C 2000 New, faster, image-based scatter correction for 3D PET *IEEE Trans. Nucl. Sci.* **47** 1587–94
- Yan J, Lim J C and Townsend D W 2015 MRI-guided brain PET image filtering and partial volume correction *Phys. Med. Biol.* **60** 961–76
- Yu D F and Fessler J A 2002 Edge-preserving tomographic reconstruction with nonlocal regularization *IEEE Trans. Med. Imaging* **21** 159–73