



UNIVERSITY OF LEEDS

This is a repository copy of *Assessment of Maximum A Posteriori Image Estimation Algorithms for Reduced Acquisition Time Medical Positron Emission Tomography Data*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/110287/>

Version: Accepted Version

Book Section:

Deidda, D orcid.org/0000-0002-2766-4339, Aykroyd, RG orcid.org/0000-0003-3700-0816 and Tsoumpas, C orcid.org/0000-0002-4971-2477 (2018) Assessment of Maximum A Posteriori Image Estimation Algorithms for Reduced Acquisition Time Medical Positron Emission Tomography Data. In: Oliveira, TA, Kitsos, CP, Oliveira, A and Grilo, L, (eds.) Recent Studies on Risk Analysis and Statistical Modeling. Contributions to Statistics . Springer, Cham , pp. 3-16. ISBN 978-3-319-76605-8

https://doi.org/10.1007/978-3-319-76605-8_1

© Springer International Publishing AG, part of Springer Nature 2018. This is a post-peer-review, pre-copyedit version of a chapter published in Recent Studies on Risk Analysis and Statistical Modeling. The final authenticated version is available online at: https://doi.org/10.1007/978-3-319-76605-8_

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Assessment of maximum a posteriori image estimation algorithms for reduced acquisition time medical positron emission tomography data

Daniel Deidda, Robert G. Aykroyd and Charalampos Tsoumpas

Abstract This study examines the effects of reduced radioactive dosage data collection on positron emissions tomography reconstruction reliability and investigates the efficiency of various reconstruction methods. Also, it investigates properties of the reconstructed images under these circumstances and the limitations of the currently used algorithms. The methods are based on maximum likelihood and maximum a posterior estimation, but no explicit solutions exist and hence iterative schemes are obtained using the expectation-maximisation and one-step-late methods, while greater efficiency is obtained by using an ordered-subset approach. Ten replicate real datasets, from the Hoffman brain phantom collected using a Siemens Biograph mMR scanner, are considered using standard deviation, bias and mean-squared error as quantitative output measures. The variability is very high when low prior parameter values are used but reduces substantially for higher values. However, in contrast, the bias is low for low parameter values and high for high parameter values. For individual reconstructions, low parameter values leads to detail being lost in the noise whereas high values produce unacceptable artefacts at the boundaries between different anatomical regions. Considering the mean-squared error, a balance between bias and variability, still identifies high prior parameter values as giving the best results, but this is in contradiction to visual inspection. These findings demonstrate that when it comes to low counts, variability and bias become significant and are visible in the images, but that improved reconstruction can be achieved by a careful choice of the prior parameter.

Daniel Deidda

Division of Biomedical Imaging, University of Leeds, Leeds, LS2 9JT, UK, e-mail: umdde@leeds.ac.uk

Robert G Aykroyd

Department of Statistics, University of Leeds, Leeds, LS2 9JT, UK, e-mail: r.g.aykroyd@leeds.ac.uk

Charalampos Tsoumpas

Division of Biomedical Imaging, University of Leeds, Leeds, LS2 9JT, UK, e-mail: C.Tsoumpas@leeds.ac.uk

1 Introduction

Positron emission tomography (PET) is a non-invasive imaging technique used in the clinical setting for routine diagnosis, dose delivery and treatment response evaluation. The use of medical imaging technologies is now commonplace in the clinical setting. Positron emission tomography leads the way in the detection of abnormalities such as cancer since, although it has low spatial resolution, it provides unrivalled functional information. As with all radiation-based methods, however, there is a risk of tissue damage which could lead to cancer at a later date. Hence there is a constant demand for decreases in radioactive dosages. Positrons are emitted by a radioactive-tracer, travelling a few millimetres before interacting with electrons. As soon as the interaction takes place a pair of photons, of identical energies, are emitted in opposite directions. The detection of the two photons allows a line-of-response to be defined which is characterised by an angle and the shortest distance between the line and the center of the detector system – see Figure 1(a) where four events are shown. When a large number of lines-of-response are plotted the resulting graph is half of a sine wave – hence the motivation for this type of graph being called a “sinogram”. With complex objects the data will consist of a large number of overlapping sine waves (Fahey, 2002) – see the real data example in Figure 1(b).

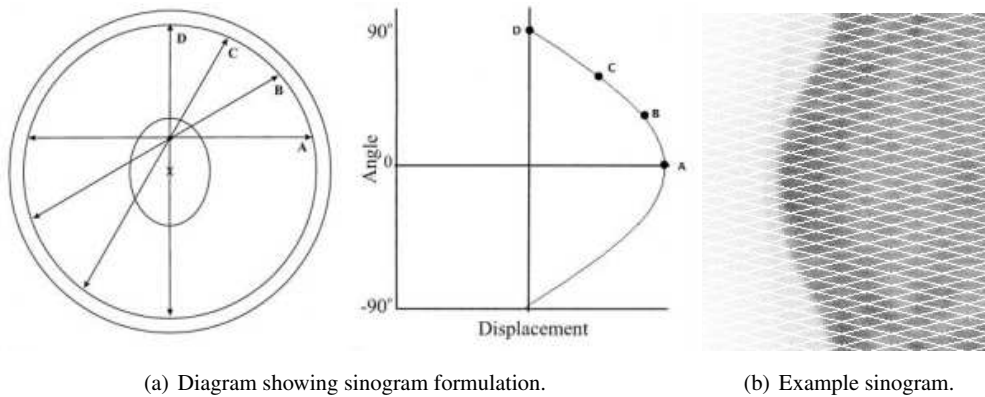


Fig. 1 PET data acquisition: (a) events recorded as a line-of-response, and (b) total counts shown in a sinogram.

If this were a non-random system then the data, Y say, would be a simple linear function of the radioisotope concentrations, Λ say, and coefficients, C say, that is $Y = C\Lambda$. In principle, this can be inverted as $\Lambda = C^{-1}Y$, or the usual linear regression estimate $\Lambda = (C^T C)^{-1} C^T Y$. In a typical data collection system, however, there are about $256 \times 64^2 \approx 16$ million data values and a 3D reconstruction space of $256^3 \approx 16$ million voxels. This means that C is a 16 million by 16 million matrix with high multicollinearity. This is a highly ill-conditioned, and possibly ill-posed, big-data inverse problem which requires careful analysis.

Several iterative algorithms are currently used to estimate the radioactive-tracer distribution based upon the principles of maximum likelihood and maximum a posteriori estimation. These have been widely studied using both simulated and clinical data, but less well for anatomically realistic measured phantom data, as is the theme here. This investigation will consider quantitative voxel-by-voxel output analysis, rather than the more common region-of-interest based approach, in terms of bias, standard deviation and mean

squared error. A further novelty is the use of low count data and hence it will be the first to make the quantitative comparison of algorithm accuracy for such low count phantom data.

This paper is organized as follows. A summary of the basic physical and statistical modelling is given in Section 2 with estimation discussed in Section 3. The data are described in Section 4 along with details of the investigation design. Numerical results are presented in Section 5 with discussion in Section 6.

2 Statistical modelling for PET data

Suppose that the 3D reconstruction space, $R \subset R^3$, is partitioned into small cubic voxels with the true radioactive-tracer concentration in voxel j labelled as λ_j giving the discretised unknown image $\Lambda = \{\lambda_j : j = 1, \dots, M\}$. Note that this partitioning is arbitrary and hence can be tailored to the application or chosen for computational convenience. Let the data sinogram be denoted $Y = \{y_i : i = 1, \dots, N\}$ which depends on the radioactive-tracer concentration through a Poisson model with

$$y_i \sim \text{Poisson} \left(\sum_{j=1}^M c_{ij} \lambda_j \right), \quad i = 1, \dots, N, \quad (1)$$

where c_{ij} represents the, assumed known, probability that an event occurring in voxel j produces a coincidence in the i -th pair of detectors and takes into account attenuation and normalisation corrections. Before moving-on it is worth noting that, for computational efficiency, the full summation over all N voxels is never performed. Instead, for each sinogram element the sum is only taken over voxels with non-negligible contribution, that is with c_{ij} above some threshold. The set of voxels with non-negligible contribution to y_i is denoted I_i and hence the model becomes

$$y_i \sim \text{Poisson} \left(\sum_{j \in I_i} c_{ij} \lambda_j \right), \quad i = 1, \dots, N. \quad (2)$$

Based on this model the corresponding data log-likelihood, $l(Y, \Lambda) = \ln L(Y, \Lambda)$, is given by

$$l(Y, \Lambda) = \sum_{i \in N} \left[y_i \sum_{j \in I_i} \ln(c_{ij} \lambda_j) - \sum_{j \in I_i} (c_{ij} \lambda_j) - \ln(y_i!) \right] \quad (3)$$

and the maximum likelihood estimates of Λ given by

$$\hat{\Lambda} = \max_{\Lambda} l(Y, \Lambda) \quad (4)$$

with the usual asymptotic approximations available for calculating a covariance matrix and construction of confidence intervals etc. However, given that M is typically very large, e.g. $M = 16$ million, any direct solution of (4) is impractical. An alternative approach is to use an iterative method, such as the EM algorithm which is discussed in the next section.

Given that this is an inverse problem, stable solution of the maximum likelihood problem is unlikely, especially when the number of unknowns is large compared to the amount of data. One approach is to incorporate additional information into the maximisation step creating a penalized likelihood approach.

This can equally be posed in a Bayesian setting with the penalty forming a prior distribution which is combined with the above likelihood to produce the posterior distribution and the maximum a posteriori (MAP) solution used as the reconstruction. Consider a prior distribution defined in terms of a Gibbs distribution

$$f(\Lambda) = \frac{1}{Z(\beta)} \exp\{-\beta V(\Lambda)\} \quad (5)$$

where $\beta > 0$ is a prior parameter and $V(\Lambda)$ is chosen so that f is large for values of Λ believed to be likely and small for implausible values. In particular, the following functional forms will be considered: (i) a Gaussian prior model on local differences

$$V_1(\Lambda) = \sum_{\langle j,k \rangle} w_{jk} (\lambda_j - \lambda_k)^2 / \lambda_j, \quad (6)$$

where $\langle j, k \rangle$ denotes all voxel neighbours, $\{w_{jk}\}$ are positive constants that define a weight value for each neighbouring voxel (in general, 1 for first-order interactions between orthogonal nearest neighbours, and $1/\sqrt{2}$ for second-order diagonal interactions), and (ii) a Gaussian prior on local variability

$$V_2(\Lambda) = \sum_j (\lambda_j - \bar{\lambda}_j)^2 / \bar{\lambda}_j, \quad (7)$$

where $\bar{\lambda}_j$ is the median of the values in the neighbourhood of voxel j . This is considered to be a more robust alternative allowing occasional sharp changes, or jumps, between neighbouring radioactive-tracer values. For further details of these prior distributions see, for example, Alenius and Ruotsalainen (2002) and Bettinardi et al. (2002).

The posterior distribution is then produced by combining the likelihood and the prior density using Bayes's Theorem resulting in the following log-posterior function (and ignoring constant terms)

$$p(\Lambda | Y) = \sum_{i \in N} [y_i \sum_{j \in I_i} \ln(c_{ij} \lambda_j) - \sum_{j \in I_i} (c_{ij} \lambda_j) - \beta V(\Lambda)], \quad (8)$$

with resulting definition of the maximum a posterior (MAP) estimate as

$$\hat{\Lambda} = \max_{\Lambda} p(Y, \Lambda). \quad (9)$$

Again, this cannot be solved easily and hence iterative algorithms can be used, one of which, based on the OSL algorithm (Green, 1990), is described in the next section.

3 Maximum likelihood and maximum posterior estimation using an EM algorithm

In general, it is difficult to directly find the maximum of the log-likelihood in (3), recalling that there are millions of data points and millions of unknown parameters, and instead an iterative approach is used which considers single parameter updates one-by-one. The algorithms currently used in the clinical setting for PET are based on the approach originally proposed in Shepp and Vardi (1982), that is the maximum

likelihood expectation maximisation (MLEM) algorithm, see also Dempster et al. (1977) and Lange and Carson (1984), which can be explained and derived as follows based on a “missing data” argument.

Suppose that instead of only Y being recorded, it had been possible to observe where all events originated, then this leaves a simple task of estimating the radioactive-tracer concentration. Hence a “complete dataset”, $X = \{X_{ij} : i = 1, \dots, N, j = 1, \dots, M\}$, is considered where X_{ij} is defined as the number of photon pairs emitted from j and detected at i and is related to y_i by $y_i = \sum_{j \in I_i} X_{ij}$. The complete data log-likelihood is then:

$$l(X, \Lambda) = \sum_{i \in N} \left[X_{ij} \sum_{j \in I_i} \ln(c_{ij} \lambda_j) - \sum_{j \in I_i} (c_{ij} \lambda_j) - \ln(X_{ij}!) \right]. \quad (10)$$

In order to obtain the algorithm updating formula, the following two steps are necessary. In these n is the iteration number and $\widehat{\Lambda}^{(n)}$ is the estimated radioactive-tracer concentration at iteration n .

- E-STEP: During this step the algorithm estimates the conditional expectation of $l(X, \Lambda)$, $E(l(X, \Lambda) | Y, \widehat{\Lambda}^{(n)})$. The expected value for $l(X, \Lambda)$, given the measured data Y and $\widehat{\Lambda}^{(n)}$, is:

$$E(l(X, \Lambda) | Y, \widehat{\Lambda}^{(n)}) = \sum_{i,j} \left[\frac{c_{ij} \widehat{\lambda}_j^{(n)} y_i}{\sum_{k \in I_i} c_{ik} \widehat{\lambda}_k^{(n)}} \ln(c_{ij} \widehat{\lambda}_j^{(n)}) - c_{ij} \widehat{\lambda}_j^{(n)} \right] + \text{Const.} \quad (11)$$

In the first iteration the image can consist of any non-negative solution, to assure the non-negativity constraint (Lange and Carson, 1984), though often an initial homogeneous image is used.

- M-STEP: In this step the algorithm finds the image that maximises the log-likelihood computed in the previous step by considering the partial derivatives:

$$\left. \frac{\partial E(l(X, \Lambda) | Y, \widehat{\Lambda}^{(n)})}{\partial \lambda_j} \right|_{\Lambda = \widehat{\Lambda}^{(n)}} = \sum_{i \in J_j} \left[\frac{c_{ij} \widehat{\lambda}_j^{(n)} y_i}{\sum_k c_{ik} \widehat{\lambda}_k^{(n)}} \lambda_j^{-1} - c_{ij} \right] = 0, \quad (12)$$

where J_j is the set of projections to which voxel j contributes. Dempster et al. (1977) showed that equation (12) is equal to $\partial l(Y, \widehat{\Lambda}^{(n)}) / \partial \lambda_j$. The resulting formula describes the MLEM algorithm:

$$\lambda_j^{(n+1)} = \frac{\widehat{\lambda}_j^{(n)}}{\sum_{i \in J_j} c_{ij}} \sum_{i \in J_j} \frac{c_{ij} y_i}{\sum_{k \in I_i} c_{ik} \widehat{\lambda}_k^{(n)}}, \quad j = 1, \dots, M. \quad (13)$$

The resulting value for $\widehat{\Lambda}^{(n+1)}$ is then used in the E-STEP of the next iteration and the procedure is repeated until convergence is reached.

The MLEM algorithm is demonstrated to be a convergent algorithm and appropriately takes into account the random behaviour of the emission process. Nevertheless, it is computationally demanding and takes many iterations to converge. An accelerated version of MLEM was developed in Hudson and Larkin (1994) using ordered subsets (OS) of the data. The resulting OSEM converges in fewer iterations and is widely used in clinical practice because it is easily implemented and provides good images more quickly.

The OSEM algorithm provides acceleration of convergence, proportional to the number of subsets, by simply processing only the data within a subset at each sub-iteration. The data is organised in ordered

subsets and the MLEM method is applied to each subset in turn. The reconstruction after each sub-iteration becomes the starting point for the following subset. In this way every iteration passes through every subset.

These subsets are usually chosen so that the projections within a subset corresponds to the projections of the image with down-sampled projection angles. The number of subsets has to be a divisor of the number of detector blocks in a ring (the Siemens mMR has 63 blocks per ring with 8×8 detectors per block and consequently, the choice could be one of 3, 9, 21 and 63). Following the same approach as for MLEM, the OSEM algorithm is obtained by substituting the sum over i with the sum over $s \in S_m$ in (13), where S_m is the chosen subset of detector pairs and $m = 1, \dots, M$, where M is the number of subsets:

$$\widehat{\lambda}_j^{(n,m+1)} = \frac{\widehat{\lambda}_j^{(n,m)}}{\sum_{s \in S_m} c_{sj}} \sum_{s \in S_m} \frac{c_{sj} y_s}{\sum_k c_{sk} \widehat{\lambda}_k^{(n,m)}}, \quad j = 1, \dots, M, \quad (14)$$

where $\widehat{\lambda}_j^{(n,m)}$ is the estimate of λ_j at sub-iteration m in the n -th full iteration. The resulting value for $\widehat{\lambda}^{(n+1)}$ is then used in the next iteration and the procedure is repeated until convergence is reached. The final solution then yields the maximum likelihood estimate, $\widehat{\Lambda}$, which will later be referred to as the MLE.

The OSEM method has been proven to converge quickly if the subset balance is respected (Hudson and Larkin, 1994). However, with this method the MLEM noise artefacts are magnified after every iteration. For this reason, in the clinical practice it is often stopped at early iterations. Further to remove the effects of noise in the reconstruction a Gaussian smoothing filter is applied as a post-processing step – later this will be referred to as the MLE+G.

To reduce the effects of noise in reconstruction in the final stages of the maximum likelihood algorithm, smoothing can be introduced through a prior distribution which then leads to the maximum a posterior estimate. As with the log-likelihood, the log-posterior function in (9) cannot be maximised directly and hence an EM-based algorithm is again considered.

The OSMAPOSL algorithm is an extension of the OSEM algorithm which iteratively maximises the posterior distribution

$$l(\lambda | Y) = \sum_{i \in N} [y_i \sum_{j \in I_i} \ln(c_{ij} \lambda_j) - \sum_{j \in I_i} (c_{ij} \lambda_j) - \beta V(\lambda)], \quad (15)$$

to produce the MAP estimate with the following updating equation:

$$\widehat{\lambda}_j^{(m+1)} = \frac{\widehat{\lambda}_j^{(m)}}{\left[\sum_{s \in S_m} c_{sj} - \beta \frac{\partial V(\Lambda)}{\partial \lambda_k} \Big|_{\Lambda = \widehat{\Lambda}^{(m)}} \right]} \sum_{s \in S_m} \frac{c_{sj} y_s}{\sum_k c_{sk} \widehat{\lambda}_k^{(m)}}, \quad j = 1, \dots, M. \quad (16)$$

This uses the same ordered-subset approach and hence has faster convergence than the original one-step late algorithm (Green, 1990). The final solution then yields the maximum a posterior estimate, $\widehat{\Lambda}$, which will later be referred to as the MAP. Further, the choice of prior function will be indicated as either V_1 or V_2 as defined in (6) and (7) respectively, and finally the value of the prior parameter will be given, for example $\beta = 100$, to give a full labelling such as MAP, $V_1, \beta = 100$.

4 Data description and assessment criteria

The data used in this study were acquired with a 3D PET-MR system (Biograph mMR, Siemens Healthcare) by colleagues at the Institute of Nuclear Medicine of University College London Hospital. The Biograph mMR has 8 rings, each one divided in 63 blocks. The detector blocks have 8×8 LSO crystals, each $8 \times 8 \times 20$ mm³ in size. Experiments were carried out using the Hoffman 3D Brain Phantom (Hoffman et al., 1990), which can provide a realistic approximation of the radioactive-tracer distribution found in the normal brain. The phantom consists of a robust plastic cylinder (diameter: 20.8 cm, height: 17.5 cm, fillable volume: ~ 1.2 l) and 19 independent plates within the cylindrical phantom. It was filled with 60 MBq ¹⁸F-fluorodeoxyglucose and the acquisition time was 3600 s giving a total number of events of about 10^9 , which represents a standard for brain acquisitions. The numerical procedures have been developed within the Collaborative Computational Project in Positron Emission Tomography and Magnetic Resonance imaging (CCP-PET-MR) – see www.ccpetmr.ac.uk and make extensive use of STIR (Thielemans et al., 2012) for data correction (attenuation, scatter, normalisation and random matches) and reconstruction. The image size after the reconstruction is $289 \times 289 \times 127$ with voxel size $2.04 \times 2.04 \times 2.03$ mm³.

With real data the activity concentration is unknown and hence a gold standard is defined as a reference, using maximum likelihood estimation from the MLEM algorithm with the full 3600 s phantom data. The reconstructed image, as MLEM is globally convergent and because of the very high level of counts, is sufficient in terms of noise, bias and so on. The image, later denoted Λ^* , used as the “true” image was obtained after 126 iterations (see Figure 2), to ensure convergence. Both “gray matter” and “white matter” voxel values can be distinguished well with substantial detail of the undulating and folded structure.

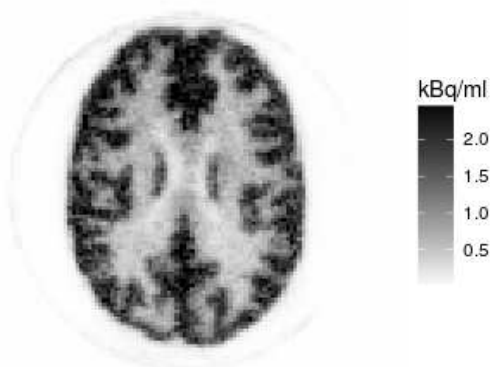


Fig. 2 MLEM image chosen as our high quality reference.

To ensure exactly equal experimental conditions, ten samples with roughly the same number of counts have been created, by sub-sampling from the 3600 s data, so as to mimic different acquisition times and to allow reconstruction assessment. The effective acquisition time was reduced from 3600 s to 36 s in order to simulate low count datasets. To analyse the low-count reconstructed images, different figures of merit have been chosen that are standard deviation (SD), Bias and Root Mean Squared Error (RMSE).

In general, to define a general quantitative analysis design, suppose that K replicate datasets are available, $\{Y^k : k = 1, \dots, K\}$ and that the corresponding result of the algorithm are estimated radioactive-tracer concentration images $\{\hat{\Lambda}^k : k = 1, \dots, K\}$. These can be used to define a mean image, $\bar{\Lambda} = \sum_{k=1}^K \hat{\Lambda}^k / K$. Further, recall that a “gold standard”, Λ^* is available from the 3600 s dataset reconstructed using the MLEM algorithm. The SD, Bias and RMSE are then defined as follows:

$$SD = \sqrt{\frac{1}{K} \sum_{n=1}^K (\hat{\Lambda}^k - \bar{\Lambda})^2}; \quad \text{Bias} = \bar{\Lambda} - \Lambda^*; \quad RMSE = \sqrt{SD^2 + \text{Bias}^2}. \quad (17)$$

The ideal reconstruction algorithm would produce low values for each of these measures indicating high reproducibility from the replicate datasets and lack of overall bias.

5 Experimental results

A preliminary investigation was carried out to choose the number of sub-iterations in the ordered-subset algorithms, OSEM and OSMAP. To do so, two regions of interest (ROI) were chosen as representative of gray and white matter. Figure 3(a) and (b) show the average value within each ROI as a function of sub-iteration from the OSEM reconstruction based on the 3600 s and 36 s datasets respectively. This shows that at the 5-th iteration for 21 subsets (corresponding to 105 sub-iterations), the ROI values in both white and gray matter have stabilized. Although in the clinical setting, with 21 subsets, 63 sub-iteration are often used, here 105 is chosen as more appropriate for our datasets. In the rest of the analysis, 21 subsets and 5 complete iterations (105 sub-iterations) are used without further comment.

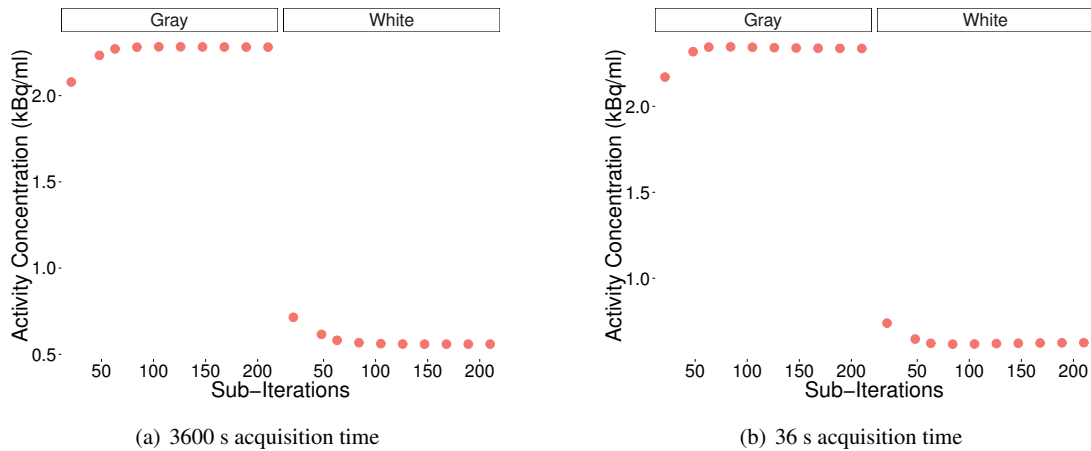


Fig. 3 Convergence of activity concentration values for white and gray matter using 21 subsets with OSEM.

The results of the main statistical investigation with the 36 s dataset over the ten replicate datasets are represented in Figures (4) and (5). These show a single dataset reconstruction (Individual), Mean, SD, Bias and RMSE images for 6 different cases. Figure (4) shows the MLE from the OSEM algorithm, the corresponding image after the Gaussian smoothing post-processing, then the MAP estimate using the

Gray Matter								
	MLE		MAP with V_1			MAP with V_2		
	MLE	MLE+G	$V_1, \beta=10$	$V_1, \beta=100$	$V_1, \beta=1000$	$V_2, \beta=10$	$V_2, \beta=100$	$V_2, \beta=500$
Mean	2.47 ± 0.59	2.39 ± 0.12	2.46 ± 0.51	2.47 ± 0.26	2.39 ± 0.15	2.49 ± 0.47	2.49 ± 0.21	2.42 ± 0.16
SD	1.53 ± 0.44	0.22 ± 0.05	1.34 ± 0.34	0.67 ± 0.13	0.22 ± 0.05	1.23 ± 0.28	0.5 ± 0.1	0.23 ± 0.05
Bias	0.41 ± 0.36	0.22 ± 0.15	0.37 ± 0.31	0.25 ± 0.18	0.22 ± 0.15	0.35 ± 0.3	0.25 ± 0.18	0.23 ± 0.16
RMSE	1.61 ± 0.48	0.33 ± 0.12	1.42 ± 0.37	0.74 ± 0.14	0.33 ± 0.12	1.3 ± 0.31	0.59 ± 0.13	0.34 ± 0.12

White Matter								
	MLE		MAP with V_1			MAP with V_2		
	MLE	MLE+G	$V_1, \beta=10$	$V_1, \beta=100$	$V_1, \beta=1000$	$V_2, \beta=10$	$V_2, \beta=100$	$V_2, \beta=500$
Mean	0.55 ± 0.25	0.61 ± 0.07	0.55 ± 0.23	0.55 ± 0.17	0.57 ± 0.09	0.54 ± 0.18	0.56 ± 0.09	0.62 ± 0.07
SD	0.57 ± 0.35	0.14 ± 0.03	0.54 ± 0.29	0.38 ± 0.13	0.14 ± 0.03	0.41 ± 0.16	0.17 ± 0.04	0.08 ± 0.02
Bias	0.16 ± 0.13	0.14 ± 0.08	0.15 ± 0.12	0.11 ± 0.08	0.1 ± 0.07	0.11 ± 0.08	0.09 ± 0.06	0.13 ± 0.09
RMSE	0.61 ± 0.35	0.20 ± 0.06	0.57 ± 0.3	0.41 ± 0.13	0.17 ± 0.05	0.43 ± 0.16	0.2 ± 0.05	0.16 ± 0.07

Table 1 Summary results, over ten samples, using three circular ROIs for each of gray and white matter.

Gaussian prior on local differences, V_1 , with $\beta = 100$. Figure (5) shows the remaining results for MAP estimates using the Gaussian prior on local differences, V_1 , with $\beta = 1000$ and then MAP estimates using the Gaussian prior on local variability, V_2 , with $\beta = 100$ and finally $\beta = 500$. The values of the prior parameters were chosen to give a range of reconstruction quality.

The SD images summarise the variability of the individual estimates over the ten samples, and show that the MLE and MAP estimates with small prior parameter have SD which is very high. Using a filter for the MLE to produce the MLE+G estimate, and a higher prior parameter for MAP estimation, help to reduce the SD. Overall the Bias is higher in regions with the lowest activity concentration. In addition, too high a prior parameter leads to artefacts around the border between different anatomical features, this is the effect of over-smoothing.

The RMSE images should take into account both Bias and SD, but here they show the same trend as the SD images. That is the variability between the ten samples swamps the Bias. With β higher than 500 and 1000 respectively for Gaussian prior on differences and Gaussian prior on variability, the Bias will become higher and more artefacts will be created. Table 1 shows the mean pixel values for two ROI, representing typical gray and white matter with each made up of three circles. Each number in the table is accompanied by the standard deviation over all the pixels within the ROI for each image in Figure (4) and (5).

6 Discussion

The purpose of this work was to check the feasibility of image reconstruction when a short acquisition time or a low radioactive-tracer dosage is used and to compare the performance of various estimation methods in this situation. The study has used real data to assess how low-count conditions affect image reconstruction

reliability, and in particular has compared different prior assumptions and different prior parameter values in terms of bias, standard deviation and mean squared error. The results show that good estimation can be achieved by a careful choice of the prior parameter. From a global perspective, low-count reconstructions show high noise and bias with all the methods showing the need for improvement. Moreover, it is clear that the convergence rate of OSEM is smaller in regions with lower pixel intensity; in fact early-stopped OSEM images show a systematic bias in regions with lower activity concentration such as white matter and the background. In contrast, MAP methods with the right prior parameter value show better performance as low activity regions have less bias. This is due to the fact that they maximise the posterior distribution introducing prior information to remove the ill-conditioning. The maximum likelihood and MAP estimates with low prior parameter values have very high RMSE while it decreases with higher prior parameter values. Purely in terms of the RMSE, the best estimation occurs with the post-filtered maximum likelihood, and

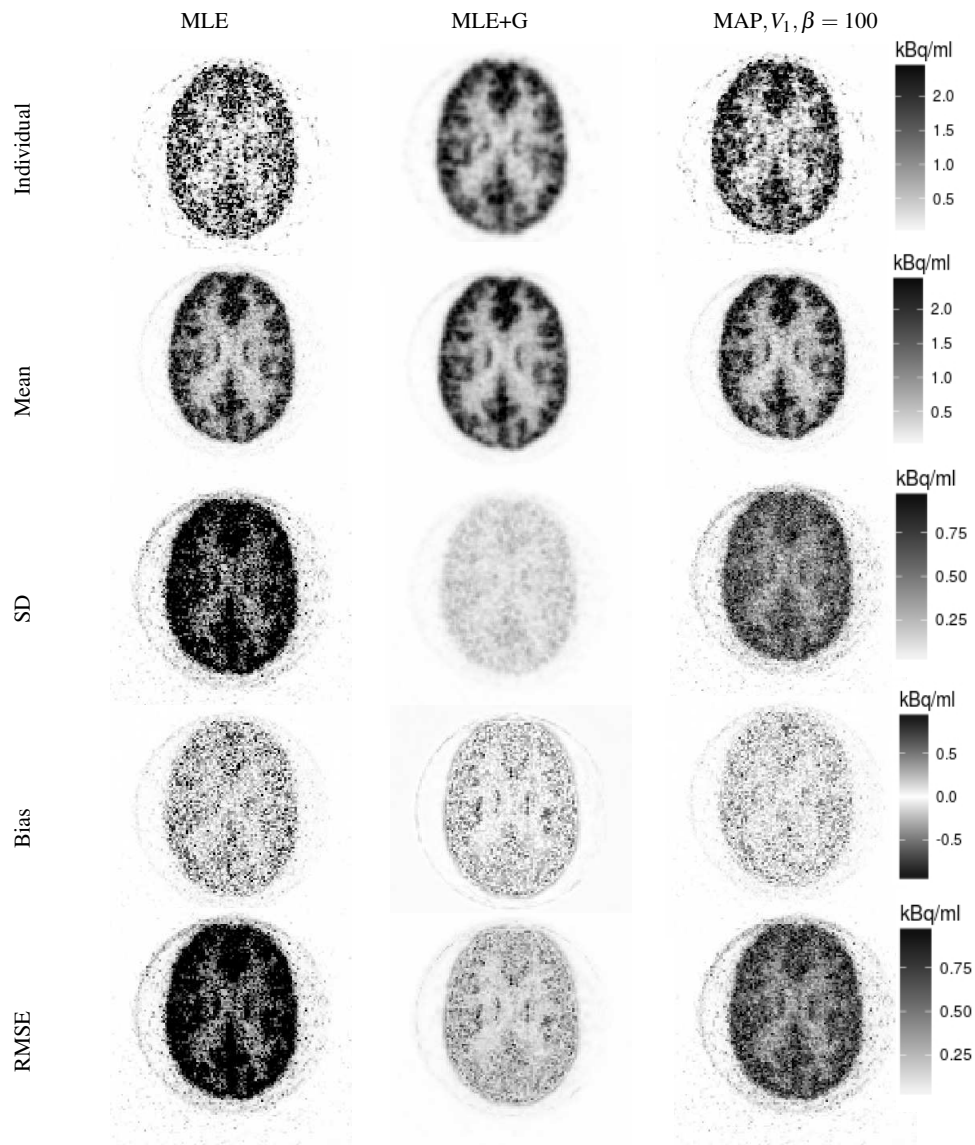


Fig. 4 Transverse view of images reconstructed with 21 subsets at the 105-th sub-iteration: 36 s acquisition.

MAP estimation with the higher values of prior parameter. On visual inspection, however, these estimates are not completely acceptable as there is high bias across the boundary of grey and white matter, and further within a region it is overly smooth. These would make it challenging to recognise small abnormalities, such as cancers. Hence, a global goodness-fit measure which gives greater weight to bias, than does the RMSE, would be more suited to such medical investigations.

A key contribution of this study is to show the difficulty in the choice of a suitable prior parameter with low-count data. Even though the MAP estimates with the highest prior parameter values have low RMSE the images appear to be over-smoothed. In contrast, the maximum likelihood estimate with Gaussian smoothing, which is the preferred method in the clinic, produces results at least as well as any of MAP estimates. Hence, to achieve substantially better results with the MAP estimation methods will need careful choice of prior parameters. Recent studies, such as Ahn et al. (2015) and Karaoglanis et al. (2015), have

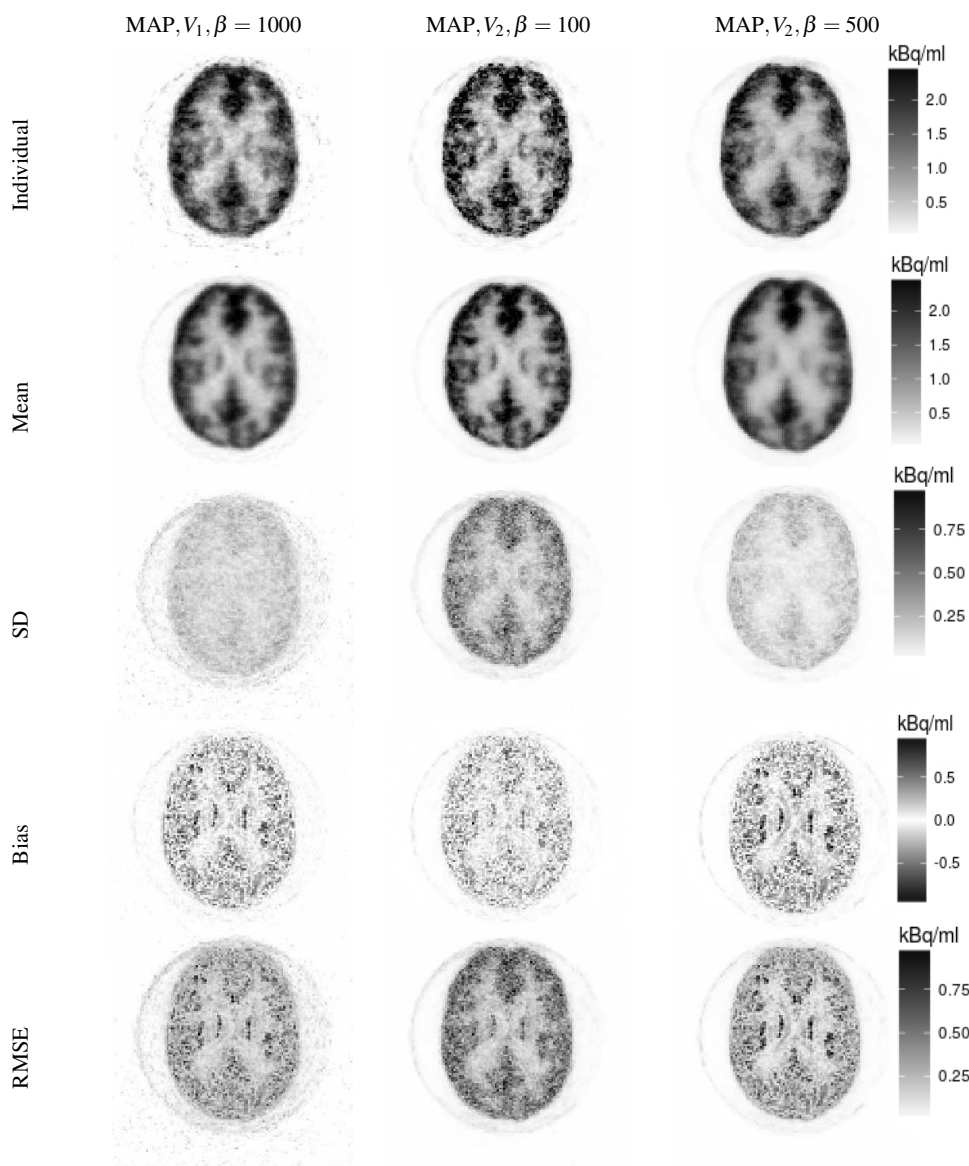


Fig. 5 Transverse view of images reconstructed with 21 subsets at the 105-th sub-iteration: 36 s acquisition.

demonstrated that regularisation can significantly improve quantification and detectability compared to post-filtered maximum likelihood. The results of this investigation confirm what was suspected about low-count data, that is, under this special circumstance reconstruction is more greatly affected by bias and high levels of noise. Under this point of view, our results are in agreement with the results in Walker et al. (2011).

The use of anatomical information from MR will be in the development of new hybrid reconstruction methods. This should help to preserve sharp contrast between adjacent anatomical features and avoid partial volume effect (Vunckx et al., 2012; Novosad and Reader, 2016). This investigation and the procedures for iterative image reconstruction considered will be a useful guide for researchers who wish to study and extend image reconstruction and correction methods for PET data.

Acknowledgements: The authors are grateful to colleagues at University College London for use of the brain phantom data, and to the University of Leeds for financial support of Daniel Deidda through a University of Leeds University Scholarship.

Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

- Ahn, S., S. Ross, E. Asma, J. Miao, X. Jin, L. Cheng, S. Wollenweber, and R. Manjeshwar (2015). Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET. *Phys Med Biol* 60, 5733–5751.
- Alenius, S. and U. Ruotsalainen (2002). Generalization of median root prior reconstruction. *IEEE transactions on medical imaging* 21(11), 1413–1420.
- Bettinardi, V., E. Pagani, M. Gilardi, S. Alenius, K. Thielemans, M. Teras, and F. Fazio (2002). Implementation and evaluation of a 3D one-step late reconstruction algorithm for 3D positron emission tomography brain studies using median root prior. *Eur J Nucl Med Mol Imaging* 29, 7–18.
- Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc (B)* 39, 1–38.
- Fahey, F. (2002). Data acquisition in PET imaging. *J Nucl Med Technol* 30, 39–49.
- Green, P. (1990). Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans Med Imaging* 9, 84–93.
- Hoffman, E., P. Cutler, W. Digby, and J. Mazziotta (1990). 3-D phantom to simulate cerebral blood flow and metabolic images for PET. *IEEE Trans Nucl Sci* 37, 616–620.
- Hudson, H. and R. Larkin (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging* 13, 601–609.

- Karaoglanis, K., I. Polycarpou, N. Efthimiou, and C. Tsoumpas (2015). Appropriately regularized OSEM can improve the reconstructed PET images of data with low count statistics. *Hellenic J Nucl Med* 18, 140–145.
- Lange, K. and R. Carson (1984). EM reconstruction algorithms for emission and transmission tomography. *J Comput Assist Tomogr* 8, 306–316.
- Novosad, P. and A. Reader (2016). MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions. *Physics in medicine and biology* 61, 46244645.
- Shepp, L. and Y. Vardi (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imaging* 1, 113–122.
- Thielemans, K., C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaios, and M. Jacobson (2012). STIR: software for tomographic image reconstruction release 2. *Phys Med Biol* 57, 867–883.
- Vunckx, K., A. Atre, K. Baete, A. Reilhac, C. M. Deroose, K. V. Laere, and J. Nuyts (2012). Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging. *IEEE Trans Med Imaging* 31, 599–612.
- Walker, M., M. Asselin, P. Julyan, M. Feldmann, P. Talbot, T. Jones, and J. Matthews (2011). Bias in iterative reconstruction of low-statistics PET data: benefits of a resolution mode. *Phys Med Biol* 56, 931–949.