ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 39 (2008) 359–368

Simulation-based evaluation of OSEM iterative reconstruction methods in dynamic brain PET studies

Anthonin Reilhac,[a,*] Sandrine Tomeï,[a] Irène Buvat,[b] Christian Michel,[c] Frank Keheren,[c] and Nicolas Costes[a,d]

[a] *CERMEP, 59 Boulevard Pinel, F-69667 Bron, France*
[b] *U678 Inserm, CHU Pitié Salpêtrière, Paris, France*
[c] *Siemens Medical Solutions, Molecular Imaging, Knoxville, TN 37932, USA*
[d] *U821, INSERM, Lyon, France*

Received 19 February 2007; revised 16 May 2007; accepted 20 July 2007
Available online 9 August 2007

The reconstruction of dynamic PET data is usually performed using filtered backprojection algorithms (FBP). This method is fast, robust, linear and yields reliable quantitative results. However, the use of FBP for low count data, such as dynamic PET data, generally results in poor visual image quality, exhibiting high noise, disturbing streak artifacts and low contrast. These signal-to-noise ratio and contrast in the reconstructed images may alter the quantification of physiological indexes, such as the regional Binding Potential (BP) obtained from kinetic modeling. Iterative reconstruction methods are often presented as viable alternatives to FBP reconstruction. In this study, we investigated the characteristics of the UW-OSEM and the ANW-OSEM iterative reconstruction methods in the context of ligand–receptor PET studies with low counts. The assessment was conducted using replicates of simulated [^{18}F]MPPF acquisitions. The quantitative accuracy obtained with the iterative and analytical methods was compared. The results show that analytical methods are more robust to the low count data than iterative methods, and therefore enable a better estimate of the regional activity values and binding potential. The positivity constraint in MLEM-based algorithms leads to overestimations of the activity in regions with low activity concentration, typically the cerebellum. This overestimation results in significant bias in BP estimates with iterative reconstruction methods. The bias is confirmed from the reconstruction of real PET data.

Keywords: PET; Image reconstruction; OSEM; Simulation

Introduction

The radiotracer 2′-methoxyphenyl-(*N*-2′-pyridinyl)-*p*-18F-fluoro-benzamidoethylpiperazine ([^{18}F]MPPF) is a specific serotonin 5-HT$_{1A}$ antagonist PET tracer recently characterized, modeled, and used for clinical research to explore abnormalities in the serotoninergic system (Merlet et al., 2004a,b). The *in vivo* exploration of the 5-HT$_{1A}$ receptors with [^{18}F]MPPF PET imaging is of great interest as those receptors are involved in numerous neurological and psychiatric disorders. Similar to other ligand–receptor PET studies, the acquisition protocol consists in the collection of coincidence events into multiple short time frames over a total scanning duration ranging from 60 to 90 min. Whereas iterative reconstruction methods are available with commercial scanners, Filtered Back Projection (FBP3D) (Kinahan and Rogers, 1989) is usually preferred for the reconstruction of dynamic brain PET scans. Indeed, this reconstruction method is fast, robust, linear and yields reliable quantitative results. Iterative reconstruction methods are often based on the Maximum-Likelihood Expectation–Maximization (MLEM) algorithm (Shepp and Vardi, 1982). Accelerated MLEM-based reconstruction algorithms such as Unweighted Ordered Subsets Expectation–Maximization (UW-OSEM) (Hudson and Larkin, 1994) and more specifically Attenuation-Weighted OSEM (AW-OSEM), Attenuation Normalization-Weighted OSEM (ANW-OSEM) (Michel et al., 1998) and Ordinary Poisson OSEM (OP-OSEM) (Yavuz and Fessler, 1996) include some data correction within the iteration process and better account for the nature of the noise. They are often presented as viable alternatives to FBP reconstruction.

Many comparison studies showed that iterative reconstruction outperforms FBP in terms of image quality, signal-to-noise ratio, resolution and contrast (Boucharebet al., 2005; Riddell et al., 2001; Gutman et al., 2003; Wang et al., 1998; Boellaard et al., 2001), and improves lesion detection (Lartizien et al., 2003). They highlighted that the characteristics of the reconstructed images are bound to the chosen number of iterations and to the source distribution (Gutman et al., 2003; Wang et al., 1998). Consequently, for a specific PET protocol, the number of iterations must be carefully selected so as to achieve reliable quantitative results while limiting the noise amplification. Most comparison studies

* Corresponding author. Fax: +33 334 72 68 86 10.
E-mail address: anthonin.reilhac@cermep.fr (A. Reilhac).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.07.038

were conducted using [^{18}F]FDG static scans containing relatively high numbers of detected events as in the context of tumor detection. Consequently, little is known about the characteristics of iterative reconstruction techniques of PET data in low count rate situations.

In this study, we evaluate the characteristics of the UW-OSEM and ANW-OSEM reconstruction methods operating on 3D data and of UW-OSEM and AW-OSEM on 2D data after Fourier rebinning (FORE) (Defrise et al., 1997), using simulated ligand–receptor [^{18}F]MPPF PET studies. In this application, the accuracy and variability of the activity levels and kinetic parameter estimates are of a higher concern than the visual quality of the reconstructed images. We compared the performance of the two iterative methods with performance obtained with analytical methods, i.e. FORE+FBP2D, FBP3D and the Direct Inversion Fourier Transform method (FORE+DIFT) (Matej and Bajla, 1990). The motivation for conducting this study was to answer the practical question, "*Is OSEM reconstruction a viable alternative* to *FBP reconstruction in quantitative PET studies in the case of low count data?*" It is somewhat similar to few previously published studies such as Michel et al. (1999) for benzodiazepine brain PET studies, Morimoto et al. (2006) for [^{11}C]raclopride and [^{11}C]DASB, Bélanger et al. (2004) for [^{11}C]WAY, and Koch et al. (2005) for dopamine SPECT studies. In our context, they provided inconclusive results. The originality of our methodology compared to previously published studies stems from the use of multiple replicates of realistic simulated [^{18}F]MPPF PET scans to accurately assess the impact of noise on the performance on the reconstructed methods. This choice was motivated by the fact that there only exists approximate local variance estimators for nonlinear reconstruction algorithms (Barrett et al., 1994; Fessler, 1996; Kadrmas et al., 1999; Qi and Leahy, 2000; Soares et al., 2000; Wang and Gindi, 1997; Wilson et al., 1994; Buvat, 2002), and that experimentally acquired replicates using phantoms or gated acquisitions as proposed by Riddell et al. (2001) do not provide the required flexibility to model very realistic dynamic acquisitions.

Material and methods

The strategy we used to assess the impact of low count data on the accuracy of the different reconstruction methods consists of comparing the quantitative measurements obtained from the reconstruction of the scans with a statistic varying from low (normal) to high (11 times the normal level). Replications of the same PET acquisition protocol (transmission+dynamic emission) were performed.

Replicated simulations

We conducted the experiment using multiple realizations (N_{real}) of the same PET acquisition from the same brain phantom and time activity curves (TACs). A total of 11 realizations of the PET acquisitions were simulated using PET–SORTEO (Reilhac et al., 2004, 2005) configured for the Ecat Exact HR$^+$ scanner operating in 3D mode (span=9, maximum ring difference=22, energy window=350–650 keV) (Brix et al., 1997). This simulator generates realistic data given a numerical phantom description, the scanner geometry and its physical characteristics. The considered brain phantom included 32 anatomical structures and cerebral regions and was derived from a T1-weighted MRI of an adult subject without pathology, acquired at the McConnell Brain Imaging Centre of Montreal. The TACs, describing the concentration variation across time of [^{18}F]MPPF within each brain region, were derived from actual human PET data which were fully corrected (scattered and random event contaminations, tissue and scanner material attenuation, system dead-time, radioisotope decay) as described in Reilhac et al. (2006). These input reference TACs were also compensated for partial volume effects as explained in the description and validation of the database (Reilhac et al., 2006). Each replicate consisted of simulating a 10-min transmission scan and a dynamic emission scan. The emission protocol consisted of collecting prompt events (unscattered, scattered and random events) over a 60-min period onto 35 time frames: 15 of 20 s, 15 of 120 s and 5 of 300 s. Each prompt sinogram i was then corrected for all the effects (attenuation using the transmission scan, random and scatter, dead-time, normalization, arc and decay) using correction programs from the e7tools package.[1] Randoms were simulated in a delayed coincidence window. The corrected scans are denoted S_i^{P} hereafter. In addition, true data scans (denoted S_i^{T} hereafter), containing unscattered events only (no scattered nor random events), were separated from each prompt replicate i during the simulation process and were only corrected for normalization, dead-time, attenuation, arc and decay. Using these true unscattered scans, we got closer to the ideal conditions of use of the MLEM-based algorithms as they contain no negative values and consequently satisfy the positivity constraints imposed by the MLEM-based algorithm on the sinogram bins. Negative values introduced by the scatter and random corrections alter the statistical Poisson nature of the data which is the main hypothesis this algorithm is based on. In addition to generating low count data replicates (S_i^{P} and S_i^{T}), we derived mean sinograms of the corrected prompts ($S_{\text{avg},n}^{\text{P}}$) and the corrected trues ($S_{\text{avg},n}^{\text{T}}$) with higher statistics by averaging from $n=2$ to $n=11$ (N_{real}) low count scans. The set of low count data replicates (S_i^{P} and S_i^{T}) and the set of mean scans ($S_{\text{avg},n}^{\text{P}}$ and $S_{\text{avg},n}^{\text{T}}$) made it possible to assess the reconstruction methods for a wide range of statistic levels, varying from 1 to 11 times the normal level.

Using the same method and phantom we generated 11 other scans (emission and transmission) from the same TACs except for few regions as the hippocampi in which the theoretical BP was lowered by about 25%. We used the two sets of replicates (group 1=normal, group 2=BP lowered in the hippocampi) to accurately assess the ability of each reconstruction method to accurately characterize the BP decreases.

Data reconstruction

Seven reconstruction methods were tested: (1) FBP3D, ramp filter, 0.5 pixel^{-1} cut-off frequency, (2) FORE+FBP2D with the same characteristics, (3) FORE+DIFT, (4) UW-OSEM3D, (5) FORE+UW-OSEM2D, (6) ANW-OSEM3D and (7) FORE+AW-OSEM2D. All algorithms, except FBP3D, are part of the e7tools package. Neither regularization nor *a posteriori* Gaussian smoothing was applied. All 2D reconstruction methods used FORE to rebin the fully corrected 3D scan into 2D.

Each of these reconstruction methods was applied to the fully corrected data and led to 35 time frames containing 63 axial slices of 128×128 voxels each (2.1×2.1×2.4 mm^3). For each reconstruction method (denoted R), 4 images were usually built from the corrected simulated scans: the reconstructed images of the mean

[1] Provided as a research tool by the manufacturer.

scans with the highest statistics ($n=N_{\text{real}}=11$): I_{H}^{P} and I_{H}^{T}, and the mean images computed from the independent reconstruction of the N_{real} low count replicates (I_{L}^{P} and I_{L}^{T}). Those images were built as follows:

$$I_{\text{H}}^{\{\text{P,T}\}} = \mathcal{R}\left\{S_{\text{avg},N_{\text{real}}}^{\{\text{P,T}\}}\right\} = \mathcal{R}\left\{\sum_{i=1}^{N_{\text{real}}} \frac{S_i^{\{\text{P,T}\}}}{N_{\text{real}}}\right\} \quad (1)$$

$$I_{\text{L}}^{\{\text{P,T}\}} = \sum_{i=1}^{N_{\text{real}}} \frac{\mathcal{R}\{S_i^{\{\text{P,T}\}}\}}{N_{\text{real}}}$$

Where

H = High, L = Low, P = Prompts, T = Trues and $N_{\text{real}} = 11$

Activity quantification and kinetic parameter estimation

We focused on the quantification of the activity and binding indexes of three specific brain regions: the insula, the hippocampus and the prefrontal cortex. These regions are particularly relevant for the neurological and psychiatric disorders resulting from neurotransmission defects. The TACs corresponding to these three regions and to the cerebellum were derived from the reconstructed dynamic volume using Volumes of Interest (VOIs). The Binding Potentials (BPs) were derived for each of the three regions from their measured TAC and using a Simplified Reference Tissue Model (SRTM) (Lammertsma and Hume, 1996) with the cerebellum used as the reference region and without using weighting factors (see Discussion). With this implementation, the BP is derived following a non-linear regression analysis. In addition, the Binding Potential was calculated voxel-wise from the reconstructed images using a similar calculation method (Gunn et al., 1997). The Binding Potential BP characterizes the local quantity of available receptors divided by the tracer affinity.

Determination of the optimal number of iterations and subsets

In a preliminary phase of the study, we determined the number of iterations and subsets enabling the most accurate kinetic parameter estimation. The optimal number of MLEM equivalent updates (iterations × subsets) is object dependent and convergence does not occur at the same iteration for the whole image. The finding of the most appropriate parameters is even more complicated for multi-frame reconstruction before kinetic modeling. In the case of BP determination using a reference region, the accuracy of the activity level measurements across time of both the target and the reference regions impacts on the accuracy of the derived BP. Therefore, we decided to determine the most appropriate reconstruction parameters from the convergence of the binding potential derived from the measured insula, prefrontal cortex and hippocampus TACs.

Each of the eleven realizations (prompts and trues from group 1: S_i^{P} and S_i^{T}) and their corresponding high statistic mean scans ($S_{\text{avg},N_{\text{real}}}^{\text{P}}$ and $S_{\text{avg},N_{\text{real}}}^{\text{T}}$) were reconstructed using the different iterative reconstruction protocols using 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, 24, 28 and 32 iterations with 16 subsets and using 1, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 56, 64 and 72 iterations with 4 subsets. The comparison of the BP measurements made from the different reconstructed volumes was used to determine the optimal reconstruction parameters (number of iterations and subsets).

Impact of the statistics on quantitative accuracy

For each method, we assessed first, the effect of the low statistics by comparing the quantification accuracy (activity and BP) obtained from the reconstruction of the low count scans (I_{L}^{P} and I_{L}^{T}) with the corresponding results obtained from the reconstruction of the high count scans (I_{H}^{P} and I_{H}^{T}). This part was conducted using the replicates from group 1 only (normal scans). The reference BP values were 1.87, 1.42 and 1.15 for the hippocampus, the insula and the prefrontal cortex regions, respectively.

Using the two sets of replicates containing the binding potential differences in the hippocampi, we also assessed the impact of the reconstruction methods on the measurement of the known difference.

Finally, using the set of mean scans $S_{\text{avg},n}^{\text{P}}$ (with $2 \leq n \leq N_{\text{real}}=11$), we calculated the change in activity and BP estimates (and therefore the associated bias) obtained with the iterative reconstructions as the statistics of the input scan increases from low to high (using data from group 1 only).

Comparison with real [^{18}F]MPPF PET studies

Real [^{18}F]MPPF PET data acquired following the same protocol and with the Ecat Exact HR^{+} scanner were reconstructed using the FORE+DIFT and ANW-OSEM3D algorithms. We computed the difference between the cerebellum activity estimates as measured from the FORE+DIFT and the ANW-OSEM3D volume.

Results

Optimal number of iterations and subsets

The impact of the number of subsets (4 and 16) was estimated by comparing the kinetic parameters BP calculated for the 3 regions using the PET volumes obtained with the different reconstruction methods ({ANW, UW}-OSEM−{2D,3D}−{16,4} subsets). Overall, the results show that beyond 80 MLEM equivalent iterations (iterations × subsets), the number of subsets has no impact on the determination of the kinetic parameter values. Indeed, after 80 iterations, the observed relative differences between BP values measured from the volumes reconstructed with 4 and 16 subsets were usually less than 3%. However, the use of 16 subsets instead of 4 subsets speeded up the reconstruction by a factor of 4. Fig. 1(a) shows the relative difference, as a function of the MLEM iteration number (iteration × subsets), between the BP values for the hippocampus region obtained from volumes reconstructed with FORE+ANW-OSEM2D using 4 and 16 subsets. Fig. 1(b) shows the relative differences between the BP values calculated at consecutive iterations i and $i+1$: $\left(\frac{\text{BP}_{i+1}-\text{BP}_i}{\text{BP}_i} \times 100\right)$, computed from the mean reconstructed images of the 11 low count replicates (I_{L}^{P}), and from the reconstructed image of the mean scan (I_{H}^{P}) exhibiting high statistics. Each point of the curve corresponds to the max difference observed between two consecutive iterations, for the three regions (envelope) and using the tested iterative reconstruction algorithms. Results show

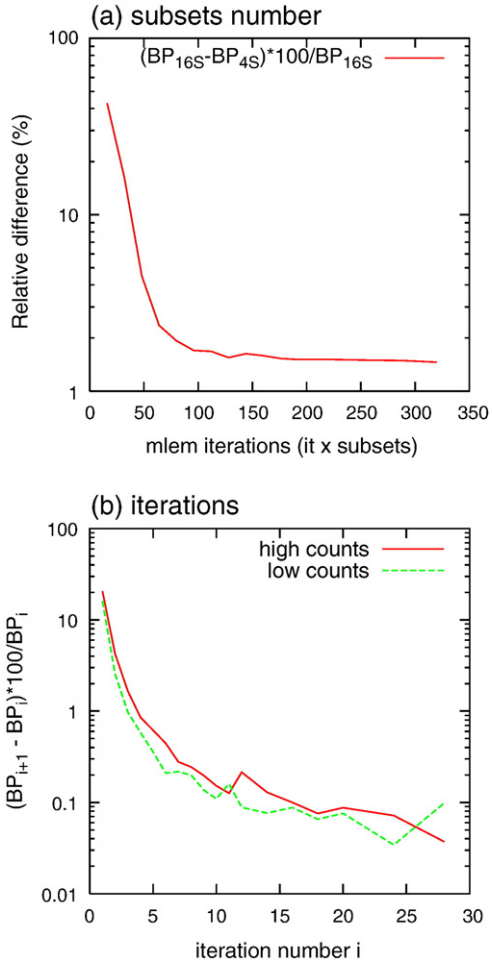

Fig. 1. (a) Relative difference between BP values of the hippocampus region calculated from volumes reconstructed with ANW-OSEM using 16 and 4 subsets. (b) Relative difference between the BP values calculated at consecutive iterations i and $i+1$ from I_L^P (low count data) and I_H^P (high count data).

that 6 iterations with 16 subsets were required to ensure the convergence with the low count data. Beyond 6 iterations, the max difference between the BP calculated at consecutive iterations i was less than 0.2%. No difference was found between the different OSEM variants and between 2D and 3D reconstructions. However, with the high count data, 10 iterations were required to reduce the difference to less than 0.2% between consecutive iterations. This suggests that a higher number of iterations is required when the statistics increases.

Comparison of the reconstruction methods

In this section, all results using iterative reconstruction were obtained using 6 iterations with 16 subsets. This corresponds to an appropriate setting in terms of convergence for our low count data as shown previously. The performance resulting from the reconstruction of the low count data is compared with the one obtained from the reconstruction of the high count scans using the same number of iterations and subsets, even if, for the latter case, 10 iterations would have been more appropriate. Fig. 2 shows the relative difference of the activity levels between the reference TAC assigned to the cerebellum region for the simulation and the TACs measured from the reconstructed images of the low and the high count scans with FORE+DIFT and ANW-OSEM3D. The cerebellum was large enough so that activity measurements are not affected by partial volume effects. This was not the case with the other studied regions which were smaller in size. The graph clearly shows that at high statistics, the cerebellar activity levels measured from the ANW-OSEM and FORE+DIFT volumes are close to the reference values. The other graphs from Fig. 3 show for each reconstruction method, the relative difference across time, between the activity values measured from the high statistic images (I_L^P and I_H^T) and from the mean volume computed from the low count images (I_L^P and I_L^T). The results clearly show that with iterative reconstruction methods, the low statistics in the emission scan yields progressive overestimations of the activity as the statistics decreases (along time). Please also note that discontinuities occur when frame duration changes. At the end of the time course, the difference in estimated activities could be as high as 80% in the case of the cerebellum with UW-OSEM3D reconstruction. The cerebellum was the most affected region, as it is the region with the lowest activity concentration for this tracer. Overall, the results suggest that for low count scans, iterative reconstruction performs better in 2D than in 3D (see Figs. 3(d, e, f) vs. Figs. 3(g, h, i)). This is due to the Fourier rebinning step which rebins the 3D scan into 2D and increases the statistics of the 2D segment. This increase of in counts is maximum at the center of the axial field of view. No major difference was observed between the UW-OSEM and ANW-OSEM methods.

The reconstruction of the true unscattered events only (Figs. 3(f) and (i)) also overestimated activity values, showing bias of the cerebellum activity of about 20% (Fig. 3(f)). This demonstrates that the bias in the cerebellum is also due to the noise in the data and not to the biased estimation of random and scatter. In some cases, the activity was underestimated with iterative reconstruction, for instance within the insula and the prefrontal regions with iterative reconstruction in 2D. It is also true for cases Figs. 3(f) and (i) (reconstruction from isolated unscattered trues). We know it is not due to FORE since analytical reconstructions after FORE do not show any substantial bias. This progressive underestimation at low statistics could be a problem of regional convergence under

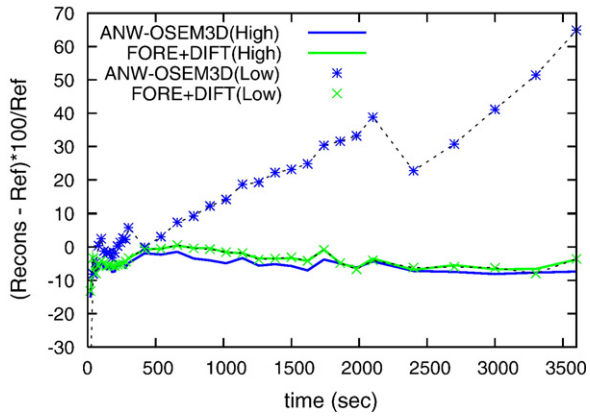

Fig. 2. Relative differences of the activity level between the reference cerebellum TACs used during the simulation and the TACs measured from the reconstructed images of the low and the high count scans with FORE+DIFT and ANW-OSEM3D.

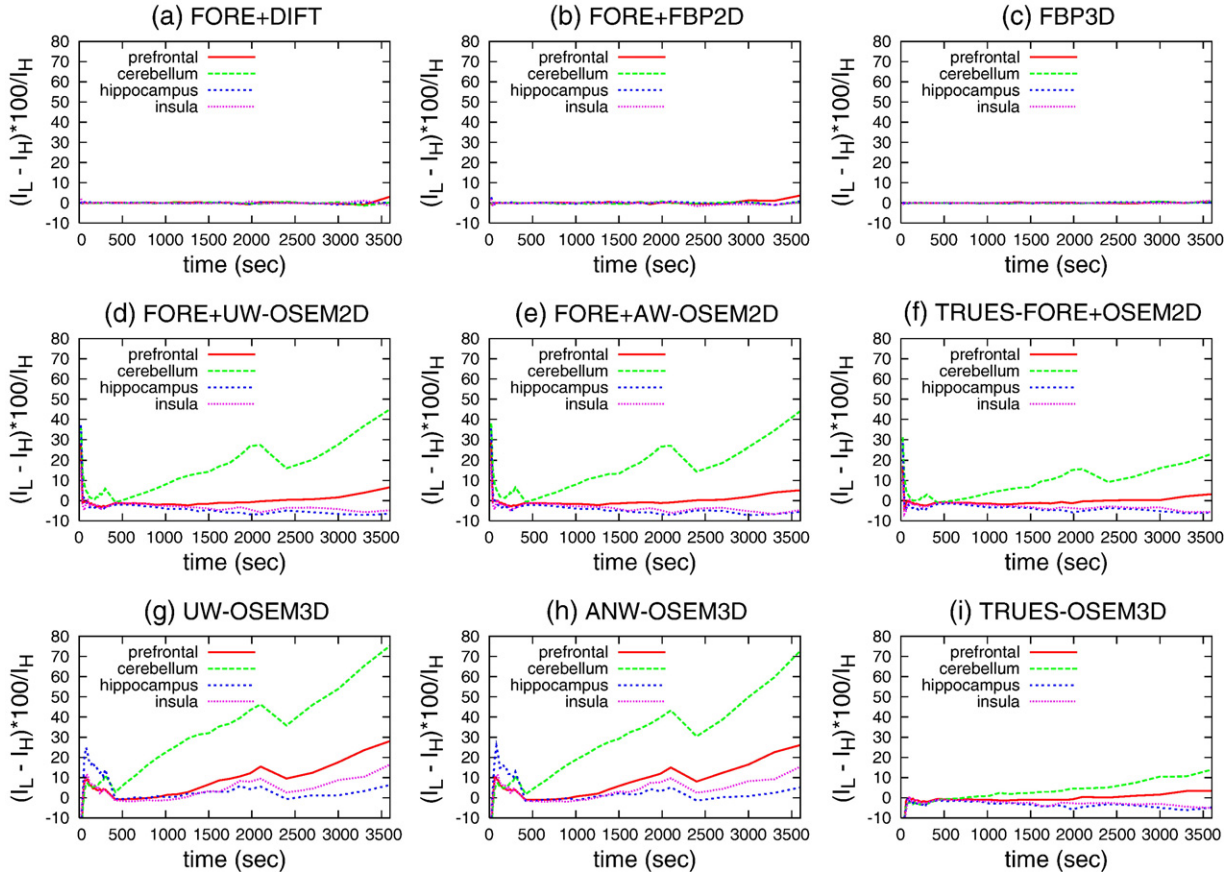

Fig. 3. Relative difference between activity values measured from the mean reconstructed volumes obtained from the low count scans (I_L^P or I_L^T) and the volumes reconstructed from the mean sinogram with high statistics (I_H^P and I_H^T) for hippocampus, insula, prefrontal cortex and for the cerebellum.

noisy conditions. More likely, the positive bias in the cold region may decrease the image values in the other regions in order to have calculated projection close to the measured projections (Ahn and Fessler, 2004). The prefrontal region, which is not a *cold region*, is overestimated. At last, no bias was observable at low statistics with the analytical methods either in 2D after FORE (Figs. 3(a) and (b)) or in 3D (Fig. 3(c)).

Fig. 4 shows the BP parameters calculated for the hippocampus region from the volumes reconstructed with the different methods. For each reconstruction method, two BP values are reported: the BP obtained from the high count scan (BP^H), and the BP obtained from the mean reconstructed volumes using the low count replicates (BP^L). These results show the impact of the bias in activity estimates at low statistics on the BP computation. As expected, no difference was found between the BP estimated from the reconstruction of the low and high count data with the analytical methods. With iterative reconstructions (ANW-OSEM and UW-OSEM), the overestimation of the cerebellum activities yields decreases of about 15% of the estimated BP. When neither randoms nor scatter were present, i.e. the corrected scans did not contain negative values, the iterative reconstruction still underestimated BP. Indeed, in the case of the ANW-OSEM3D reconstruction, a difference of 2.5% was observed between the BP computed from the low and high count images. However, at high statistics, 3D iterative reconstructions led to better accuracy in BP estimates than other methods, enabling BP estimates closer to the true values (determined from the input TACs used for the simulation). Finally, the graph shows that even at high statistics, the BP is underestimated for all the reconstruction methods. This systematic bias is mainly caused by an underestimation of the

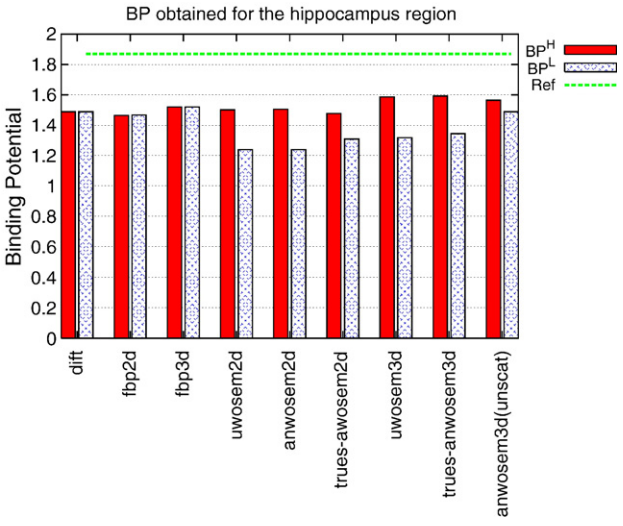

Fig. 4. BP parameters obtained for the hippocampus region from the volumes reconstructed with the different methods. For each reconstruction method, two BP values are reported: the BP obtained from the high count scan (I_H^P), and the BP obtained from the mean reconstructed volumes using the low count replicates (I_L^P).

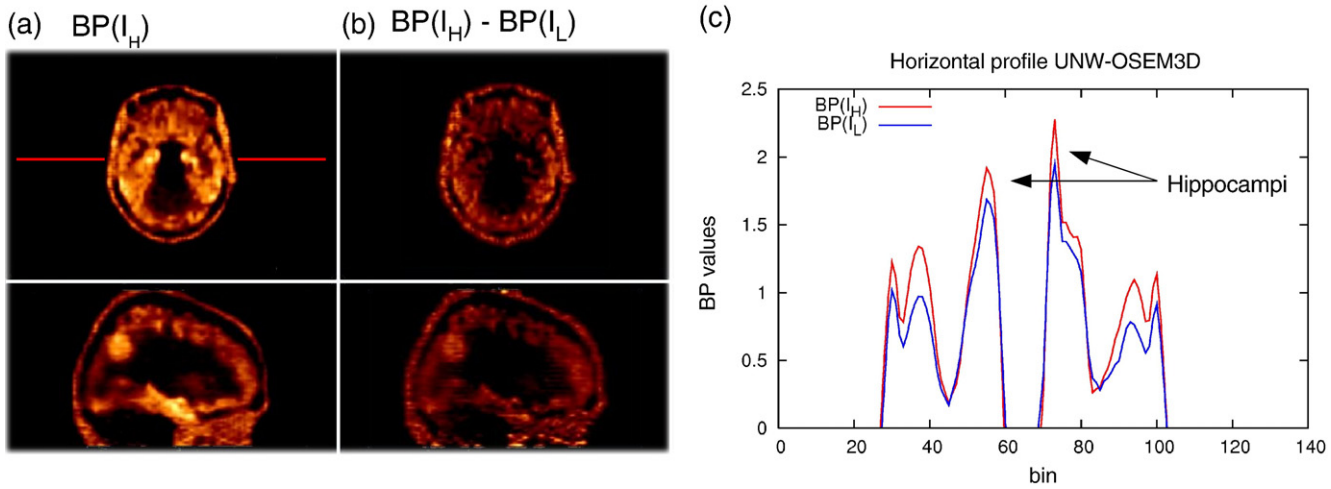

Fig. 5. (a) BP image calculated from the reconstructed volume (UW-OSEM3D) of the high count scan BP(I_H). (b) The voxel to voxel difference between BP(I_H), the BP image calculated from I_H^P and BP(I_L), the BP image calculated from the mean volume I_L^P, obtained from the reconstruction of the low count data replicates. (c) Horizontal profiles extracted from both BP volumes computed from the images containing the low (blue) and the high statistics (red).

activity in the tissue of interest due to partial volume effects. Results for the insula and the prefrontal cortex led to similar findings.

Fig. 5(a) shows the BP image BP(I_H) calculated from the reconstructed volume (UW-OSEM3D) of the high count scan (I_H^P) using the SRTM method applied at each voxel instead of regions of interest. Fig. 5(b) shows the voxel to voxel difference between BP(I_H) and BP(I_L) calculated from the mean volume, obtained from the reconstruction of the low count data replicate using UW-OSEM3D (I_L^P). This image shows the bias at each voxel due to the low statistics of the scans when compared to the use of data with high number of counts. The profiles computed from the images containing the low (blue) and the high statistics (red) are also shown in Fig. 5(c). The corresponding BP images obtained with FORE+FBP2D or FORE+DIFT showed no bias at low statistics.

Detection of the binding potential difference with the reconstruction methods

Fig. 6 shows the relative differences between normal and lowered hippocampal BP. For each reconstruction method, two values are reported: the relative differences between normal and lowered hippocampal BP obtained from the high count scan (denoted hereafter by BP^H), and the relative difference obtained from the mean reconstructed volumes using the low count replicates (denoted hereafter to by BP^L). The standard deviation of the measure is also shown in the case of the low count replicates. The theoretical difference, computed from the TACs employed for the simulation process is also reported. At high statistics, all methods yield a reasonably precise determination of the variation with the exception of FORE+FBP2D. This could well be due to noise in the data reconstructed without statistical model although the difference between FORE+FBP2D and FORE+DIFT should be negligible (except maybe in the high frequency domain). At low counts, FORE+FBP2D yields the best results, but with the highest variability. FORE+DIFT yielded a correct estimate of the difference with a low variability. With iterative methods, the bias generated by the positivity constraint is not observable. With the exception of UW-OSEM, other iterative methods yield comparable or even slightly better detection of BP changes than analytical reconstruction methods. This is in good agreement with Mesina et al. (2003). However, it must be noted that the cerebellum TACs used in the simulation of both group replicates were identical and therefore the bias cancels out when measuring relative changes.

Evolution of activity and BP estimates with the count level

Fig. 7 shows the measured activity of the cerebellum region and the BP of the 3 regions, as the statistics of the input scan increases. Fig. 7(a) shows the TAC of the cerebellum measured from the ANW-OSEM3D reconstruction of the low count scan (x1) and

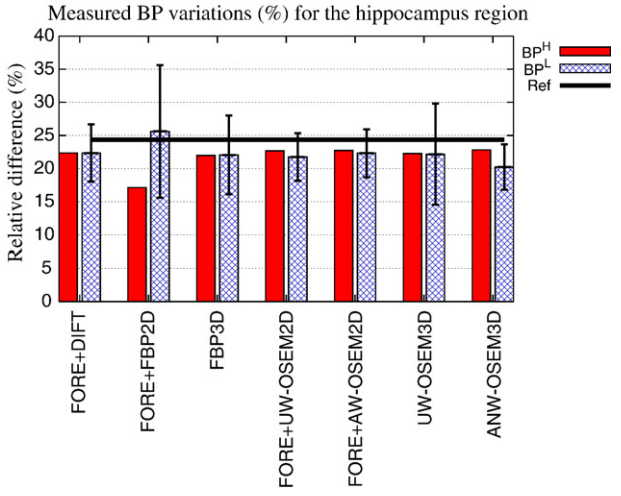

Fig. 6. Relative differences between normal and lowered hippocampal BP. For each reconstruction method, two values are reported: the relative difference obtained from the high count scan (BP^H), and the relative difference obtained from the mean reconstructed volumes using the low count replicates (BP^L). The standard deviations computed from the 11 independent measurements of are also shown for the low count data.

from the reconstruction of the mean sum scans of 2 (×2) and 6 (×6) low count data replicates. The TACs of the cerebellum region used for the simulation (reference) and measured from the FORE+DIFT image are also shown. The TACs obtained from the reconstruction of the mean sum of 6 scans and from the FORE+DIFT reconstruction of a single scan are very close to the reference TAC. The reconstruction of the average sum scans obtained from more than 6 scans did not result in any significant difference. Fig. 7(b) shows the evolution of the BP values computed for the 3 regions as a function of n, the number of low count data replicates used to build the mean scan: $\sum_{i=1}^{n} S_i^P/n$. The BP values significantly increase up to $n=6$. Overall, the results shown in Fig. 7 suggest that below $n=6$, the statistics of the input scan is too low for unbiased ANW-OSEM3D reconstruction of the cerebellum. Note that all frames before $t \leq 600$ s are unbiased.

Comparison with real data

Fig. 8 shows the relative difference between the mean frame activities measured from the FORE+DIFT and ANW-OSEM3D volumes as a function of the number of decays per ml (determined from the DIFT images), for two actual scans (P1 and P2). Results obtained with the AW-OSEM2D and the ANW-OSEM3D reconstruction of a single simulated scan are also shown. Results show that under 250,000 decays per ml, the cerebellum activities measured from the FORE+DIFT and ANW-OSEM3D volumes differed by more than 5%. This value is in good agreement with the lowest limit values of 200,000 predicted with the FORE+OSEM2D reconstruction of the simulated PET scans. The higher noise present in actual studies due the out of field of view activity could explain this higher value.

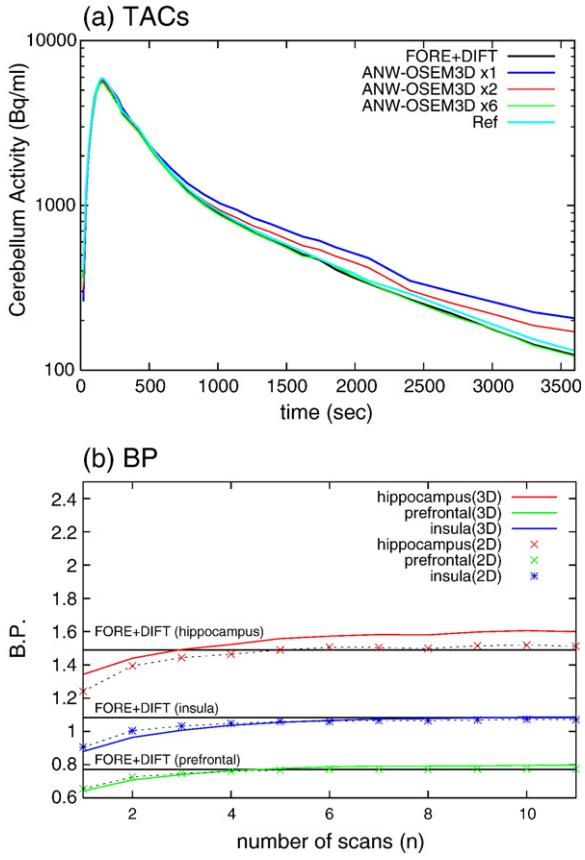

Fig. 7. Evolution of the estimates of the cerebellum activity (a) and the BP (b) as the statistics of the input scan increases. Graph (a) shows the TAC of the cerebellum measured from the ANW-OSEM3D reconstructions of the low count scan (×1) and of the mean sum scans of 2 (×2) and 6 (×6) low count data replicates. The true cerebellum TAC employed in the simulation (Reference) and the TACs measured from the volume reconstructed using the FORE+DIFT algorithm are also shown. Graph (b) shows the change in the estimated BP values as a function of n, the number of low count data replicates used to build the mean scan: $\sum_{i=1}^{n} S_i^P/n$.

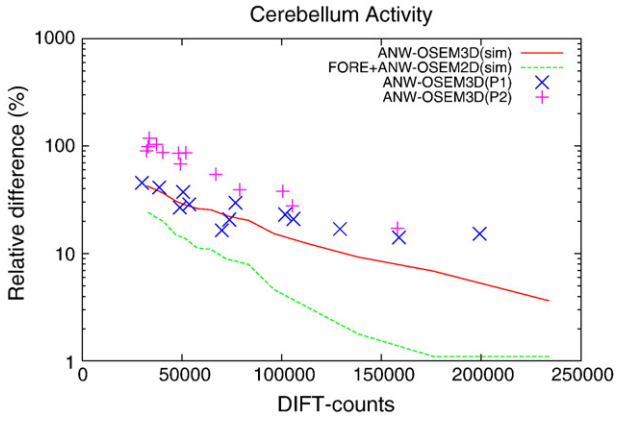

Fig. 8. Relative difference between the mean frame cerebellum activities measured from the FORE+DIFT and ANW-OSEM3D volumes of real [^{18}F] MPPF PET data as a function of the number of decays per ml (determined from the FORE+DIFT images). Results obtained from the FORE+AW-OSEM2D and ANW-OSEM3D reconstructions of a single simulated scan are also shown.

Discussion

In this study, using multiple data replicates, we investigated the performances of various iterative reconstruction schemes in 3D and in 2D after FORE, in case of low count data. Different conclusions can be drawn from the presented results.

Convergence

First, the convergence study showed that the optimal number of iterations depends on the statistics of the input scan. The higher the statistics, the higher the number of iterations is to be used. This is consistent with the study of Erlandsson et al. (2000) wherein the number of iterations required for convergence was assessed in different noise conditions. We found, using the low count replicates, that beyond 6 iterations and 16 subsets, no more quantification change was observable. However, using the high count data, a minimum of 10 iterations was necessary to reach convergence. In the latter case, the activity and BP estimates were more accurate. Six and ten iterations, in conjunction with 16 subsets, correspond to 96 and 160 MLEM updates during the reconstruction. Published values are usually under 100 equivalent MLEM iterations. However, this is consistent with the convergence results reported for cold regions in Erlandsson et al. (2000) using OSEM algorithms. Morimoto et al. (2006) used 6 iterations with 16 subsets as recommended by the manufacturer. Bélanger et al. (2004) found in a similar study with the same scanner that 6 iterations with 32 subsets (192 MLEM

iterations) were required to ensure a 99% recovery of the parameter of interest. However, the protocol and methodology were different.

Algorithm performance

The methodology proposed here using multiple replicates allowed us to accurately assess the impact of the statistics on 7 reconstruction methods among which 3 were analytical methods and 4 were iterative methods. Analytical methods (FORE+FBP2D, FORE+DIFT and FBP3D) yielded identical results whatever the statistics of the scans is. This evidenced the nice property of linearity with the analytical reconstruction methods. The high noise and poor visual image quality do not affect the activity estimates and therefore the derived binding potential. At low statistics, FBP3D performed the best. However, this result should be taken with caution as its implementation was coming from another package (3DRP; Kinahan and Rogers, 1989). FORE+DIFT was reliable as well and generated images with lower noise than FORE+FBP2D (Fig. 6), probably due to the implementation of the gridding (interpolation in Fourier Domain). The performance obtained with iterative reconstruction methods strongly depended on the statistics. Our experiments showed that the input scans should present a level of statistics 6 times higher than in normal MPPF scans to enable unbiased reconstructions using ANW-OSEM3D. This statistics could not be achieved with our scanner. In a similar study using actual data, Bélanger et al. (2004) demonstrated for the [^{11}C]-WAY ligand–receptor tracer, and with the same scanner, that a minimum of 50 Bq/ml (non-decay-corrected) were needed for unbiased FORE+OSEM2D reconstructions (error less than 5%) with frame duration of 10 min. This roughly corresponds to 30,000 decays per ml. Here, the requirement is estimated at roughly 110,000 and 200,000 decays per ml for unbiased 2D and 3D OSEM reconstructions, respectively. Those limits, estimated for our acquisition protocol and scanner, correspond to 180 Bq/ml and 333 Bq/ml (non decay-corrected) for 2D and 3D, respectively, and using frame duration of 10 min. In addition, Bélanger et al. (2004) showed that the model-based scatter correction overestimated scatter at activity under 25 Bq/ml, leading to bias for both analytic and iterative reconstructions. This problem is related to the initial image reconstruction used to estimate the scatter distribution: a streaky FBP image may be unbiased but provides a noisy scatter estimate too. This problem has been solved in e7tools: it uses an iterative reconstruction at very low resolution. A different scatter estimation strategy using analytical reconstruction but combining the statistics of adjacent frames may well be a better approach in the case of dynamic studies. This approach has indeed been proposed by UBC group (Cheng et al., 2007). Here, we did not include the correction of the raw data as a possible bias. Indeed, low-count replicates were fully corrected before reconstructions and the corrected high-counts replicates were obtained from averaging the corrected low-counts replicates. Consequently, the difference in performance observed between analytical and iterative reconstructions in case of low and high statistics came only from the reconstruction part itself. However, using those replicates we found no impact of the statistics on the scatter correction accuracy (results not shown). This is in agreement with Bélanger et al. (2004) as the theoretical cerebellum activity employed in our experiments were always above the limit under which the scatter correction is biased. Bélanger et al. (2004) explained the observed reconstruction bias observed at low count rates, in the case of no scatter correction, by truncation of negative sinogram values created by random correction. At low count activity, these random counts mainly originate from the not perfectly shielded ^{68}Ge rod sources (between 700 and 400 cps). Our simulation model accounted for a constant background random contamination of 700 cps due to the transmission sources. At our activity levels, this corresponded to a random fraction of 11% during the last frame. However, we measured the impact of the rod source contaminations. Indeed, we recomputed the last frame of each of the 11 realizations using the same protocol but without the rod source contamination and we did not measure any noticeable effect (results not shown). The cerebellum activity measured on the last frame of the ANW-OSEM3D reconstructed volumes decreased from 203 Bq/ml with rod source contamination (random fraction 11.2%) to 190 Bq/ml without rod source contaminations (random fraction=0.9%), whereas the value obtained with FORE+DIFT was 123 Bq/ml (see Fig. 7(a)). This study showed that quantification bias remains when the sinogram contains no negative counts (True unscattered events only). This positive bias is due to the image domain non-negativity constraint (OSEM-based algorithms prevent the use of update image containing negative values). This phenomenon was already observed in previously published results (Mesina et al., 2003; Boellaard et al., 2001; Riddell et al., 2001). The iterative reconstruction of the true unscattered events allowed us to assess the magnitude of the positivity bias induced by the non-negativity constraint of the image voxel values only. Indeed, during the iterative reconstruction of positive sinograms, the voxel from low activity regions, with values above the true value can be locally balanced by null voxel values only but not by negative values.

Relevance of the simulated study

The study was conducted using a single functional model (set of TACs) from a database (Reilhac et al., 2006). However, we ascertained ourselves that the selected model was representative of the other data from the simulated database. In other measurements (not described here) and using the whole simulated database, we found systematic underestimations of about 20% of the BP values in the hippocampi regions, when the data were reconstructed using ANW-OSEM instead of FORE+DIFT. In Reilhac et al. (2006), the realism of this simulated database was successfully validated against real [^{18}F] MPPF PET data. Fig. 9 shows the mean time activity curves and the standard deviation for the cerebellum computed from the simulated and real [^{18}F]MPPF PET data. The graph clearly shows the realism of the simulated data regarding the activity level in the cerebellum which is the most critical point with iterative reconstruction.

Protocol acquisition and analysis optimization

In this study, we followed a protocol consisting in the acquisition of 35 time frames over a 60-min time period. The duration of the late frames was 5 min. This study showed that this framing scheme led to insufficient detected counts for unbiased MLEM-based reconstructions. Two practical solutions have been investigated to address this (results not shown):

- Lengthening the frame durations while reducing the number of frames should increase the statistic within each frame. This

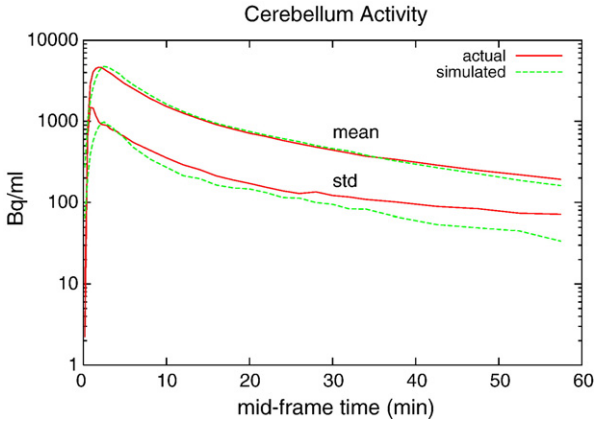

Fig. 9. Mean time activity curves and standard deviation computed for the cerebellum from the simulated and real [^{18}F]MPPF PET data.

would be at the cost of a lower time sampling which may lead to bias in BP computation. To answer this question, we generated the same replicates using a 20-time frame protocol: 10 of 30 s, 2 of 120 s, 5 of 240 s, 1 of 360 s, 1 of 600 s and 1 of 900 s, for a total scan duration of 60 min. The raw data were reconstructed with ANW-OSEM2D and processed following the same methodology as described in the method. The new framing scheme resulted in significant decreases of the bias. The measured cerebellar activities were overestimated by 10%. The incidences on the bias in the hippocampus BP estimates dropped to below 9% (the BP was underestimated by 18% with the 35-frame protocol). However, this new time sampling altered the theoretical BP, computed from the TACs used in the simulation and the time sequence, by more than 10% in some cortical regions. Consequently, it does not seem acceptable to use longer frame since bias in BP appears in other cortical regions.

- Restricting the BP computation for a total time period shorter than the nominal scan duration by excluding the late frames containing too low detected events. We computed the BP from the ANW-OSEM2D PET images using the following time period: 0–40 min, 0–45 min, 0–50 min and 0–55 min. No significant improvement was found. This was expected since low activity cerebellum TAC is only unbiased for the first 10 min.

Finally, the presented BP results were computed without using weighting factors. However, the use of weighting factors $\left(\frac{(\text{frame duration})^2}{\text{counts in frame}}\right)$ did not reduce the bias.

Other MLEM schemes

The observed bias in the low count region of an image reconstructed with low count data is inherent to the MLEM-based algorithm. The use of an Ordinary Poisson scheme (comparing measured and estimated prompt) minimizes that bias (since there is no pre-correction) but does not eliminate it, except at high statistics. Although the HR$^+$ acquires prompt and separate delayed coincidences, the OP-OSEM scheme was not implemented since it requires a projector in line of response space and the present implementation is using parallel projections (i.e. after arc correction). Because the truncation of negative values results mainly in a positive bias, several methods which allow negative values in the sinogram have been recently proposed (Ahn and Fessler, 2004; Li and Leahy, 2006). However, they maintain the non-negativity constraint in the image domain. It would be worthwhile to investigate if methods such as AB-EMML (Byrne, 1998) which removes the non-negativity constraint in image space would help in decreasing the bias of cold regions. Indeed it was observed in a SPECT study (Erlandsson et al., 2000) that, using AB-OSEM (accelerated version of AB-EMML using subsets) with a negative lower bound, the noise in the reconstructed images was much more uniformly distributed than that in the OSEM images. These authors also indicated that the bias observed in the OSEM values would be reduced. The NEG-ML algorithm (Nuyts et al., 2002), developed in the context of whole-body reconstruction of PET data without attenuation correction, is an alternative to AB-OSEM: it also allows negative values in the image and reduces inconsistency in the data, but bounds do not have to be provided. Preliminary results in 2D show that it is indeed the case (Nuyts, 2007). Since both algorithms are not implemented in the clinical processing suite, studying their performance was outside the scope of the present work.

Conclusion

We investigated the performance of the UW-OSEM and ANW-OSEM iterative reconstruction methods of data containing low statistics using multiple simulated replicates. The results showed that the studied iterative reconstruction methods are biased at low statistics, especially in the lower part of the image dynamic range and for cold regions. The bias comes from the lack of detected events rendering the data prone to negative values after corrections along with the zero-thresholding in the sinogram space and the non-negativity constraint in the image space. This study using simulated [^{18}F]MPPF PET data showed that for our scanner, the input scans should include 6 times more counts than normal to enable unbiased reconstructions with ANW-OSEM. At higher count rates, 3D iterative methods perform slightly better than the analytical methods that were considered in this study. In the case of low count data, analytical methods were found to be more robust than iterative methods and allow a better activity estimate and therefore a better kinetic parameter estimate. The reconstruction of the prompt data (OP-OSEM) with NEG-ML may reduce the bias.

Acknowledgments

The authors would like to thank the following people for discussions and assistance in this study:

- Fabrice Bellet and Christophe Pera (CREATIS, Lyon, France) for providing us the computational resources.
- Carole Lartizien (CREATIS, Lyon, France) and Andrew Reader (School of Chemical Engineering and Analytical Science, University of Manchester, UK) for helpful discussions.

References

Ahn, S., Fessler, J.A., 2004. Emission image reconstruction for randoms-precorrected PET allowing negative sinogram values. IEEE Trans. Med. Imag. 23 (5), 591–601 (May).

Barrett, H.H., Wilson, D.W., Tsui, B.M., 1994. Noise properties of the EM algorithm: I. Theory. Phys. Med. Biol. 39 (5), 833–846 (May).

Bélanger, M.J., Mann, J.J., Parsey, R.V., 2004. OS-EM and FBP reconstructions at low count rates: effect on 3D PET studies of [^{11}C] WAY-100635. NeuroImage 21 (1), 244–250 (Jan.).

Boellaard, R., van Lingen, A., Lammertsma, A.A., 2001. Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: quantitative characteristics and effects on kinetic modeling. J. Nucl. Med. 42 (5), 808–817 (May).

Bouchareb, Y., Thielemans, K., Spinks, T., Rimoldi, O., Camici, P.G., 2005. Comparison of analytic and iterative reconstruction methods for quantitative cardiac PET studies in 3D using Oxygen-15 water scans. IEEE Nucl. Sci. Symp. Conf. Rec. 4, 2120–2123.

Brix, G., Zaers, J., Adam, L.E., Bellemann, M.E., Ostertag, H., Trojan, H., Haberkorn, U., Doll, J., Oberdorfer, F., Lorenz, W.J., 1997. Performance evaluation of a whole-body PET scanner using the NEMA protocol. National Electrical Manufacturers Association. J. Nucl. Med. 38 (10), 1614–1623 (Oct.).

Buvat, I., 2002. A non-parametric bootstrap approach for analysing the statistical properties of SPECT and PET images. Phys. Med. Biol. 47 (10), 1761–1775 (May).

Byrne, C., 1998. Iterative algorithms for deblurring and deconvolution with constraints. Inverse Problems 14, 1455–1467.

Cheng, J.-C.K., Rahmim, A., Blinder, S., Camborde, M.-L., Raywood, K., Sossi, V., 2007. A scatter-corrected list-mode reconstruction and a practical scatter/random approximation technique for dynamic PET imaging. Phys. Med. Biol. 52 (8), 2089–2106 (Apr.).

Defrise, M., Kinahan, P.E., Townsend, D.W., Michel, C., Sibomana, M., Newport, D.F., 1997. Exact and approximate rebinning algorithms for 3-D PET data. IEEE Trans. Med. Imag. 16 (2), 145–158 (Apr.).

Erlandsson, K., Visvikis, D.W.A., Waddington, I.C., Jarrit, P.H., Pilowsky, L.S., 2000. Low-statistics reconstruction with AB-EMML. IEEE Nucl. Sci. Symp. Conf. Rec. 3, 15/249–15/253 (Oct.).

Fessler, J.A., 1996. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. IEEE Trans. Image Process. 5, 493–506.

Gunn, R.N., Lammertsma, A.A., Hume, S.P., Cunningham, V.J., 1997. Parametric imaging of ligand–receptor binding in PET using a simplified reference region model. NeuroImage 6 (4), 279–287 (Nov.).

Gutman, F., Gardin, I., Delahaye, N., Rakotonirina, H., Hitzel, A., Manrique, A., Guludec, D.L., Véra, P., 2003. Optimisation of the OS-EM algorithm and comparison with FBP for image reconstruction on a dual-head camera: a phantom and a clinical ^{18}F-FDG study. Eur. J. Nucl. Med. Mol. Imaging 30 (11), 1510–1519 (Nov.).

Hudson, H.M., Larkin, R.S., 1994. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans. Med. Imag. 13 (4), 601–609.

Kadrmas, D.J., DiBella, E.V., Huesman, R.H., Gullberg, G.T., 1999. Analytical propagation of errors in dynamic SPECT: estimators, degrading factors, bias and noise. Phys. Med. Biol. 44 (8), 1997–2014 (Aug.).

Kinahan, P.E., Rogers, J.G., 1989. Analytic 3D image reconstruction using all detected events. IEEE Trans. Nucl. Sci. 36 (1), 964–968.

Koch, W., Hamann, C., Welsch, J., Pöpperl, G., Radau, P., Tatsch, E., 2005. Is iterative reconstruction an alternative to filtered backprojection in routine processing of dopamine transporter SPECT studies? J. Nucl. Med. 46 (11), 1804–1811 (Nov.).

Lammertsma, A.A., Hume, S.P., 1996. Simplified reference tissue model for pet receptor studies. NeuroImage 4 (3 Pt. 1), 153–158 (Dec.).

Lartizien, C., Kinahan, P.E., Swensson, R., Comtat, C., Lin, M., Villemagne, V., Trébossen, R., 2003. Evaluating image reconstruction methods for tumor detection in 3-dimensional whole-body PET oncology imaging. J. Nucl. Med. 4 (2), 276–290 (Feb.).

Li, Q., Leahy, R.M., 2006. Statistical modeling and reconstruction of randoms precorrected PET data. IEEE Trans. Med. Imag. 25 (12), 1565–1572 (Dec.).

Matej, S., Bajla, I., 1990. A high speed reconstruction from projections using direct Fourier method with optimal parameters – an experimental analysis. IEEE Trans. Med. Imag. 9, 421–429 (Dec.).

Merlet, I., Ostrowsky, K., Costes, N., Ryvlin, P., Isnard, J., Faillenot, I., Lavenne, F., Dufournel, D., Bars, D.L., Mauguière, F., 2004a. 5-HT$_{1A}$ receptor binding and intracerebral activity in temporal lobe epilepsy: an [^{18}F]MPPF–PET study. Brain 127 (Pt. 4), 900–913 (Apr.).

Merlet, I., Ryvlin, P., Costes, N., Dufournel, D., Isnard, J., Faillenot, I., Ostrowsky, K., Lavenne, F., Bars, D.L., Mauguière, F., 2004b. Statistical parametric mapping of 5-HT$_{1A}$ receptor binding in temporal lobe epilepsy with hippocampal ictal onset on intracranial EEG. NeuroImage 22 (2), 886–896 (Jun.).

Mesina, C.T., Boellaard, R., Jongbloed, G., van der Vaart, A.W., Lammertsma, A.A., 2003. Experimental evaluation of iterative reconstruction versus filtered back projection for 3D [^{15}O]water PET activation studies using statistical parametric mapping analysis. NeuroImage 19 (3), 1170–1179 (Jul.).

Michel, C., Sibomana, M., Bol, A., Bernard, X., Lonneux, M., Defrise, M., Comtat, C., Kinahan, P.E., Townsend, D.W., 1998. Preserving Poisson characteristics of PET data with weighted OSEM reconstruction. IEEE Nucl. Sci. Symp. Conf. Rec. M6-61.

Michel, C., Liu, X., Sanabria, S., Lonneux, M., Sibomana, M., Bol, A., Comtat, C., Kinahan, P.E., Townsend, D.W., Defrise, M., 1999. Weighted schemes applied to 3D-OSEM reconstruction in PET. IEEE Nucl. Sci. Symp. Conf. Rec. 3, 1152–1157.

Morimoto, T., Ito, H., Takano, A., Ikoma, Y., Seki, C., Okauchi, T., Tanimoto, K., Ando, A., Shiraishi, T., Yamaya, T., Suhara, T., 2006. Effects of image reconstruction algorithm on neurotransmission PET studies in humans: comparison between filtered backprojection and ordered subsets expectation maximization. Ann. Nucl. Med. 20 (3), 237–243 (Apr.).

Nuyts, J., 2007. Private communication.

Nuyts, J., Stroobants, S., Dupont, P., Vleugels, S., Flamen, P., Mortelmans, L., 2002. Reducing loss of image quality because of the attenuation artifact in uncorrected PET whole-body images. J. Nucl. Med. 43 (8), 1054–1062 (Aug.).

Qi, J., Leahy, R.M., 2000. Resolution and noise properties of MAP reconstruction for fully 3-D PET. IEEE Trans. Med. Imag. 19 (5), 493–506 (May).

Reilhac, A., Lartizien, C., Costes, N., Sans, S., Comtat, C., Gunn, R.N., Evans, A.C., 2004. PET–SORTEO: a Monte Carlo-based simulator with high count rate capabilities. IEEE Trans. Nucl. Sci. 51 (1), 46–52 (Oct.).

Reilhac, A., Batan, G., Michel, C., Grova, C., Tohka, J., Costes, N., Evans, A.C., 2005. PET–SORTEO: validation and development of database of simulated PET volumes. IEEE Trans. Nucl. Sci. 52 (5), 1321–1328 (Oct.).

Reilhac, A., Evans, A.C., Gimenez, G., Costes, N., 2006. Creation and application of a simulated database of dynamic [^{18}F]MPPF PET acquisitions incorporating inter-individual anatomical and biological variability. IEEE Trans. Med. Imag. 25 (11), 1431–1439 (Nov.).

Riddell, C., Carson, R.E., Carrasquillo, J.A., Libutti, S.K., Danforth, D.N., Whatley, M., Bacharach, S.L., 2001. Noise reduction in oncology FDG PET images by iterative reconstruction: a quantitative assessment. J. Nucl. Med. 42 (9), 1316–1323 (Sep.).

Shepp, L.A., Vardi, Y., 1982. Maximum likelihood reconstruction in positron emission tomography. IEEE Trans. Med. Imag. 1 (2), 113–122.

Soares, E.J., Byrne, C.L., Glick, S.J., 2000. Noise characterization of block-iterative reconstruction algorithms: I. Theory. IEEE Trans. Med. Imag. 19 (4), 261–270 (Apr.).

Wang, W., Gindi, G., 1997. Noise analysis of MAP-EM algorithms for emission tomography. Phys. Med. Biol. 42 (11), 2215–2232 (Nov.).

Wang, C.X., Snyder, W.E., Bilbro, G., Santago, P., 1998. Performance evaluation of filtered backprojection reconstruction and iterative reconstruction methods for PET images. Comput. Biol. Med. 28 (1), 13–24 (Jan.).

Wilson, D.W., Tsui, B.M., Barrett, H.H., 1994. Noise properties of the EM algorithm: II. Monte Carlo simulations. Phys. Med. Biol. 39 (5), 847–871.

Yavuz, M., Fessler, J.A., 1996. Objective functions for tomographic reconstruction for random-precorrected PET scans. IEEE Nucl. Sci. Symp. Conf. Rec. 2, 1067–1071.