HHS Public Access

Author manuscript

Phys Med Biol. Author manuscript; available in PMC 2016 April 07.

Published in final edited form as:

Phys Med Biol. 2015 January 7; 60(1): 15–29. doi:10.1088/0031-9155/60/1/15.

Evaluation of Bias and Variance in Low-count OSEM List Mode Reconstruction

Y Jian[1], **B Planeta**[2], and **R E Carson**[1,2]

Y Jian: yiqiang.jian@yale.edu

[1]Department of Biomedical Engineering, Yale University, New Haven, CT USA

[2]Department of Diagnostic Radiology, Yale University, New Haven, CT USA

Abstract

Statistical algorithms have been widely used in PET image reconstruction. The maximum likelihood expectation maximization (MLEM) reconstruction has been shown to produce bias in applications where images are reconstructed from a relatively small number of counts. In this study, image bias and variability in low-count OSEM reconstruction are investigated on images reconstructed with MOLAR (motion-compensation OSEM list-mode algorithm for resolution-recovery reconstruction) platform. A human brain ([^{11}C]AFM) and a NEMA phantom are used in the simulation and real experiments respectively, for the HRRT and Biograph mCT. Image reconstructions were repeated with different combination of subsets and iterations. Regions of interest (ROIs) were defined on low-activity and high-activity regions to evaluate the bias and noise at matched effective iteration numbers (iterations x subsets). Minimal negative biases and no positive biases were found at moderate count levels and less than 5% negative bias was found using extremely low levels of counts (0.2 M NEC). At any given count level, other factors, such as subset numbers and frame-based scatter correction may introduce small biases (1–5%) in the reconstructed images. The observed bias was substantially lower than that reported in the literature, perhaps due to the use of point spread function and/or other implementation methods in MOLAR.

1. Introduction

Statistical image reconstruction algorithms have been widely used in PET for the past few decades. The introduction of a statistical model substantially improves the robustness in estimating images from noisy measurements, and thus can result in a better tradeoff between object resolution and image noise over analytical reconstruction methods (Qi and Leahy, 2006). In PET imaging where the data are typically treated as Poisson measurements, a classic solution is to iteratively update the estimates by maximizing the likelihood functions, i.e., iterative MLEM (maximum likelihood expectation maximization) reconstruction (Shepp and Vardi, 1982; Lange and Carson, 1984). However, MLEM and its variant (OSEM, ordered subset EM) (Hudson and Larkin, 1994) have been shown to produce bias in applications where images are reconstructed from a relatively small number of counts, e.g., in short-duration frames in dynamic PET studies. Such a bias would be very problematic for quantitative imaging, since the use of shorter frames, particularly important when trying to derive an image-derived input function (Watabe *et al.*, 2001; de Geus-Oei *et al.*, 2006; Fung

and Carson, 2013), will alter the outcome measures. Also, in applications such as brain imaging where count levels can be lower in subjects with larger body mass (assuming same injected dose), artificial relationships between brain function and body mass index might be found, cause by count-level induced bias.

In a simulated [^{18}F]MPPF brain study on the HR+ (CTI/Siemens Medical Solutions USA, Inc.), overestimation and underestimation were observed in regions of interest (ROIs) with relatively low activity concentration and in ROIs with high activity concentration, respectively (Reilhac *et al.*, 2008). Similarly on the HRRT (CTI/Siemens Medical Solutions, USA), some researchers have reported up to 10–20% bias in some low-activity ROIs reconstructed at 0.2 M noise equivalent counts (NEC) (Johansson *et al.*, 2007; van Velden *et al.*, 2009; Walker *et al.*, 2011). Overestimation was postulated to be due to the intrinsic non-negativity constraint of image voxels in the MLEM update equation. Similar bias was observed with other reconstruction algorithms, such as the shifted Poisson ordered-subsets EM (SP-OSEM) and ordered-subsets weighted least squares (OSWLS) (van Velden *et al.*, 2009).

Several variants of ML estimation have been proposed to overcome this issue by allowing negative voxels in the images. The AB-MLEM algorithm (Byrne, 1998) imposes a lower bound (typically a negative value) and an upper bound on image voxels, and NEG-ML (Nuyts *et al.*, 2002) utilizes a new preconditioner to enable negative voxels. Both algorithms were shown to reduce image bias (Erlandsson *et al.*, 2000; Grezes-Besset *et al.*, 2007), but at the expense of higher standard deviation (Verhaeghe and Reader, 2010) and slower convergence speed (Stute *et al.*, 2011). Additional parameters introduced in these methods must be tuned for particular reconstruction tasks and the robustness of these choices must be investigated. While the inverse relationship between image bias and number of the counts is generally recognized in MLEM reconstruction, it is still not clear whether the positivity constraint is the dominant factor contributing to the low-count bias. One example is the inconsistent bias level found between the sinogram reconstruction (Walker *et al.*, 2011; Reilhac *et al.*, 2008) and list mode reconstruction (MOLAR(Carson *et al.*, 2003)) in comparable experiments carried out with the HRRT. Specifically, bias calculated in ROIs was minimal within a reasonable low-count range for the HRRT reconstruction with MOLAR (Planeta-Wilson *et al.*, 2008; Jian and Carson, 2013).

The primary goal of this study was to further evaluate the low-count bias and variance in MOLAR reconstructions. Simulations of HRRT imaging were performed to quantify bias and noise in brain ROIs over a range of reasonable count levels. In addition, NEMA phantom list-mode data acquired from the Biograph mCT scanner (Siemens Medical Solutions, USA) were resampled into low-count data sets and then used to evaluate bias in ROIs with different contrast, also using MOLAR (Jin *et al.*, 2013).

2. Materials and methods

2.1 List-mode MLEM

The statistical imaging model for PET imaging is written as

$$E[Y_i]=T(L_iA_iN_i\sum_{j=1}^{J}c_{i,j}\lambda_j+R_i+S_i),\quad i=1,\ldots,I \quad (1)$$

where λ_j represents the radioactivity at voxel j, $E[Y]$ represents the expectation of the measurement Y, and for line of response i, L is the decay factor, A is the attenuation factor, N is the normalization factor, R is the random coincidence rate, S is the scatter coincidence rate, T is the frame duration and c is the system matrix. By maximizing the Poisson likelihood function of the measured counts Y using the EM algorithm, one can reconstruct the image λ through a iterative update process (Shepp and Vardi, 1982; Lange and Carson, 1984). MOLAR (the motion-compensation OSEM list-mode algorithm for resolution-recovery reconstruction) has been developed using the Poisson measurement model in Eq 1, and applied to the HRRT (Carson *et al.*, 2003), the Biograph mCT (Jin *et al.*, 2013), and the Focus-220 (Jin *et al.*, 2014). The iterative reconstruction formula of MOLAR, an OP-OSEM algorithm, is written as:

$$\lambda_j^{(n+1)}=\frac{\lambda_j^n}{Q_j}\sum_{k=1}^{K}\frac{c_{k,j}L_kA_kN_k}{T\left(\sum_{j'}c_{k,j'}L_kA_kN_k\lambda_{j'}^{(n)}+R_k+S_k\right)} \quad (2)$$

Q in Eq. 2 is the voxel-dependent sensitivity including effects of normalization, attenuation and motion.

2.2 Data re-sampling

A resampling strategy was used to generate independent low-count sample lists L_I (each with I counts) from high-count data list L_h (with s x I counts by redistributing events in list mode files sequentially to each sample:

$$\mathrm{L}_l[i]=\mathrm{L}_h[l+s\cdot(i-1)],\quad i=1,2\ldots.I \quad (3)$$

where s represents the number of low-count sample lists, l represents the sample index, and i is the index of the events in each list. Since the events were sampled in an interleaved manner, all the samples are statistically equivalent and independent.

3. Experiments

3.1 Human brain simulation and reconstruction

A simulation study was performed to evaluate the regional bias and variance at various low-count levels. A 3D brain image (figure 1) reconstructed from 60–120 min post-injection of a human [^{11}C]AFM (serotonin transporter ligand, (Naganawa *et al.*, 2013)) study was utilized as the simulation source distribution. The original image reflects distinctive tracer uptake among brain ROIs. The simulated list mode data were generated by using the forward-projector of MOLAR, modeling the HRRT scanner and creating raw list-mode data with attenuation and normalization effects (Carson *et al.*, 2003).

As scatter and randoms are important factors that could possibly contribute to bias in images, we created two forms of list mode data: S^{off} and S^{on} data sets. S^{off} is the list mode data set without scatter and random effects in the simulation or correction in the reconstruction. In this experiment, the data containing only true coincidence events were used to assess the low-count bias and variability purely caused by the OSEM algorithm. S^{on} is list mode data set with scatter and random effects in simulation and correction in reconstruction. In this simulation, ~30% scattered events and ~10% random events were added to the simulated data based on the corresponding fractions estimated from the real human dataset reconstruction. During image reconstruction, scatter correction was performed in the same way, regardless of the count level of the resampled data. Specifically, scatter profiles were simulated based on the SSS model (Watson *et al.*, 1997) from intermediate emission images, and then scaled to match the number of measured scattered events (prompt events minus delayed events) outside the attenuation volume. In MOLAR, the scatter profiles are updated in the first few OSEM iterations; this was previously found to be sufficient for the scatter estimate to converge. 500 M true events and 1.5 B prompt events were generated for S^{off} and S^{on}, respectively. In this way, the noise equivalent counts (NEC, (Strother *et al.*, 1990)) were ~500 M in both simulations.

In this study, the simulated list mode data were resampled in four ways, i.e., s = 10, 100, 500, and 2500. NEC levels for these four sets of samples were 50, 5, 1 and 0.2 M. Note that NEC = 0.2 M is lower than the count level in any single frame of human brain reconstruction on the HRRT. This NEC level is referred to as ultra-low counts.

While the default subset number for MOLAR's OSEM reconstruction on the HRRT is 30, we performed a series of image reconstructions from the same data sets by choosing different numbers of subsets (1, 5, 10, 15 and 30). The images for evaluation were obtained from the same effective iteration numbers, which is defined as the product of subsets and iterations. These reconstruction experiments investigated the impact of OSEM subset number upon image bias and noise. The subset parameter is denoted as *Sub* hereafter. For instance, Sub_{30} and Sub_1 represent 30 subsets and 1 subset, respectively.

3.2 NEMA phantom reconstruction

A low-count evaluation was also performed with a NEMA IEC body phantom (Data Spectrum, Hillsborough, NC, USA) scanned on the Biograph mCT scanner (Jakoby *et al.*, 2011). The phantom was filed with 74 MBq of [^{18}F]FDG and had a contrast ratio of 6.2:1 between the hot spheres and background. 600 M prompt events (12% random coincidences and 31% scatter coincidences), were obtained in a 30-min acquisition. The original list-mode data were down-sampled, as described above, by factors of 10, 100, and 1000 to generate three groups of low-count datasets: 10 x 60 M, 100 x 6 M and 1000 x 0.6 M. The corresponding NEC levels were 20 M (180 seconds), 2 M (18 seconds) and 0.2 M (~2 seconds) for these three groups. Image reconstruction was performed with MOLAR adapted for the Biograph mCT (Jin *et al.*, 2013). The attenuation map for attenuation correction was generated from the CT image by translating Hounsfield units to attenuation coefficients. The random coincidence rate was estimated from the product of the singles rates of the two detectors for each LOR. To match the default effective iteration number in MOLAR to that

in the sinogram-based OSEM option provided by the manufacturer's software (3 iterations / 21 subsets), two subset numbers were adopted in reconstruction for Biograph mCT data: 3 subsets and 21 subsets, up to a total of 126 effective iterations.

3.3 Low-count bias and variability evaluation

For each choice of down-sampling factor s, the intensity value in a given ROI n of the l^{th} replicate image was denoted as $R_{s,n,l}$. The mean value of $R_{s,n,l}$ across the replicates is

$$\overline{R}_{s,n}=\frac{1}{M}\sum_{l=1}^{M}R_{s,n,l},\quad M=\min(s,100) \quad (4)$$

In Eq. 4, the number of replicates evaluated (M) was chosen to be no more than 100 for image reconstruction in the low-count groups to reduce computation time, at the cost of higher standard deviation. In this case, the 100 replicates were chosen at random from the full set.

In the [^{11}C]AFM brain simulation study, bias and standard deviation of brain ROIs were evaluated in S^{off} and S^{on} experiments. Three 3D ROIs, cerebellum (45856 voxels with low activity), thalamus (9437 voxels with high activity), and striatum (17627 voxels with high activity), were defined on an MR template (Tzourio-Mazoyer *et al.*, 2002) that was registered to the subject's individual MR scan which was registered to the PET images (Viola and Wells, 1997), as shown in figure 1. These ROIs were then applied to all the reconstructed images and the bias for brain ROI n was calculated at each low-count level s as follows:

$$Bias_{s,n}=\frac{\overline{R}_{s,n}-T_n}{T_n}\cdot 100\% \quad (5)$$

where T_n is the true mean intensity in ROI n, taken from the original image used for simulation (figure 1).

To characterize variability of the ROI means across replicates, the coefficient of variation of $CoV_{s,n}$ of ROI n from count-level sample s was calculated as

$$CoV_{s,n}=\frac{\sqrt{\frac{1}{M-1}\sum_{l=1}^{M}(R_{s,n,l}-\overline{R}_{s,n})^2}}{\overline{R}_{s,n}}\cdot 100\% \quad (6)$$

In the NEMA phantom study (figure 2), bias was calculated on hot and cold ROIs and noise was calculated in the warm background. For the hot and cold spheres, 32- and 38-mm diameter spherical ROIs were centered on the 37-mm hot and 45-mm cold spheres. For the NEMA phantom study, ROI bias for ROI n was calculated on low-count images with respect to a reference image reconstructed from the highest count level (s=1) with 126 iterations / 1 subset.

$$Bias_{s,n}=\frac{\overline{R}_{s,n}-\overline{R}_{1,n}}{\overline{R}_{1,n}}\cdot 100\% \quad (7)$$

To characterize the variability of hot and cold ROIs, the CoV was calculated across replicate images using Eq. 6.

4. Results

4.1 Human brain simulation and reconstruction

In the HRRT simulation study, Bias and CoV in three brain ROIs were first evaluated without scatter and random effects (S^{off}) and then with scatter and random effects (S^{on}).

For the S^{off} experiment, mean reconstructed images at various NEC levels are shown in figure 3. There was no visual indication of large bias in these images. Visually, the 0.2 M mean image appears noisier, since only 100 replicates were used in that calculation. The quantitative results of the S^{off} experiment are shown in figure 4. At a fixed number of effective iterations, both ROI bias and variance increased as the count levels decreased, as shown in figure 4(a). Unlike previous work and similar to our previous HRRT study, the bias values in three ROIs were very small, only in the range of −0.5% to −5%, even in reconstructions of 0.2 M NEC. The largest bias was −4.2% in striatum at 0.2 M NEC. Note that the bias was always negative, regardless of the activity level of the ROIs. The ultra-low-count reconstruction (0.2 M NEC) was repeated with 60 effective iterations, but different subset numbers. It is shown in figure 4(b) that both bias and variance in ROIs progressively decreased as subset number decreased. The striatum showed the largest underestimation among the three ROIs, i.e., bias and CoV were −4.2% and 2.9% respectively in the Sub_{30} reconstruction, while the corresponding values were reduced to −2.9% and 1.9% in the Sub_1 reconstruction.

In the S^{on} simulation experiment, somewhat larger bias was found in the same comparisons, as shown in figure 5 compared to figure 4. For example, bias values in the three ROIs are between −5% and −8% in the reconstruction from 0.2 M NEC, as shown in figure 5(a). This larger effect may be due to biased estimation of scatter in low-count reconstructions; this may be produced by the bias in the image, since the scatter component is estimated from the image values. The variability of the ROI estimates is also 1–4% higher in the S^{on} simulation than that in the S^{off} simulation. Other than the larger bias and variability, the dependency of ROI bias and variability on counts and subsets in the S^{on} experiment is consistent with that in the S^{off} experiment. As with the S^{off} case, bias and variability were higher with larger subset numbers in the S^{on} reconstructions, as shown in figure 5(b). The 30-subset reconstructions introduced an additional 5–10% negative bias compared with the 1-subset reconstructions.

We confirmed that the additional negative bias found at 60 effective iterations was caused by the large numbers of subsets rather than differences in convergence speed as a function of subsets. In figure 6, ROI bias as a function of effective iterations is shown for Sub_{30} and Sub_1 reconstructions. In both S^{off} (figure 6(a)) and S^{on} (figure 6(b)) cases, the convergence

curves of mean ROI values reached plateaus by 120 effective iterations, with small differences between 60 and 120 iterations (between −2% and 2%) This indicated that, under approximate convergence conditions, there was 1–5% higher ROI underestimation in Sub_{30} reconstructions than in Sub_1 reconstructions.

Other researchers have reported positive bias in low-activity regions and negative bias in high-activity regions in some comparable low-count studies (Reilhac *et al.*, 2008; Walker *et al.*, 2011). However, since negative bias was found in high- and low-activity ROIs of this study regardless of their absolute activities, we hypothesized that the total activity in the reconstructed images decreased in low-count OSEM reconstruction. In figure 7, image counts in the transaxial planes within the brain area are shown for a few groups of low-count reconstructions. In the reconstructed 3D volume, bias was larger at both ends of the axial FOV than in the middle. This is presumably due to the lower NEC levels in the outer slices than the middle slices due to the axial sensitivity profile of the system. In addition, by comparing the axial image bias in the S^{off} experiment (figure 7(a)) with that in the S^{on} experiment (figure 7(b)), we confirmed that larger bias was found when scatter and random effects were included, consistent with the results seen in figure 5 compared with figure 4.

4.2 NEMA phantom reconstructions

In the NEMA phantom experiment on the Biograph mCT, bias was evaluated in cold and hot ROIs and the variability of bias was calculated across replicates. In figure 8, bias is plotted against CoV for reconstructions at three NEC levels (0.2 M and 2 M and 20 M) as a function of iteration. It was found that both ROI bias and variability increased in magnitude as NEC decreased, consistent with the findings in the HRRT simulation. The values of bias were again very small. For example, 1–2% negative bias was found at 2 M NEC, and 2–3% negative bias was found at 0.2 M NEC after 21subsets / 6 iterations. As iterations proceeded, negative bias tended to increase slightly in the hot ROI and decrease in the cold ROI.

Since a reduction of reconstructed image activity was seen in both hot and cold ROIs, the bias of total image counts was compared between reconstructions with 3 subsets and reconstructions with 21 subsets, as shown in figure 9. Lower-NEC reconstructions showed global negative bias. At a given NEC level, image bias was also related to subset numbers, i.e., larger numbers of subsets produced higher bias. For example, at the NEC level of 0.2 M, bias in total image counts was −4.2% with 3 subsets and −5.9% with 21 subsets after 126 effective iterations. However, the subset-dependent effect only applies to low-count reconstructions. i.e., the reduction in total image counts was < 1% regardless of subsets in reconstructions from 20 M NEC.

5. Discussion

Low-count bias in PET images was presumed to be caused by the positivity constraint in OP-OSEM reconstruction, and has been a concern in the literature. However, our simulation and phantom work replicated a preliminary study (Planeta-Wilson *et al.*, 2008) and showed that there were minor levels of bias associated with MOLAR OP-OSEM reconstructions, much smaller than those found by others (see below). In our experiments, only negative bias was observed in all the regions regardless of relative activity concentration, whereas negative

and positive biases were found in high- and low-activity ROIs, respectively, by other researchers (Johansson *et al.*, 2007; Reilhac *et al.*, 2008; Walker *et al.*, 2011).

In the first part of the simulation study without scatter and random effects, bias evaluated as the mean activity in several brain ROIs was less than −5%, even when the NEC was extremely low (0.2 M). We hypothesized that the minimal image bias associated with MOLAR, compared with other reconstruction algorithms, was due to some aspects of its implementation. Such differences included the use of chronological order of events to define subsets, as opposed to projection-angle order in sinogram-based reconstruction. Another difference could be the larger number of LORs in list mode format than in compressed sinogram format. However, based on additional analyses, neither of these factors had a significant impact on the results (data not shown). We did find that larger subset numbers tended to introduce additional bias in low-count reconstructions. This is presumably due to the fact that there are insufficient events in each subset involved in updating the images. Consider the OSEM update equation at a given voxel *j*

$$\lambda_j^{m+1,n}=\frac{\lambda_j^{m,n}}{Q_j}B^{m+1}\left[\frac{1}{F\left[\lambda^{m,n}\right]}\right] \quad (8)$$

where m is the subset number, n is the iteration number, $B^{m+1}[\bullet]$ is the back projection operator for the next subset of events, $F[\bullet]$ is the forward projection operator, Q_j is the sensitivity factor for voxel j, and λ_j is the estimated value at voxel j. In Eq. 8, the update estimate $\lambda_j^{m+1,n}$ is calculated as the product of the previous estimate $\lambda_j^{m,n}$ and the backprojected values of subset (M+1) of the events. If very few events in this subset interact with voxel j, the voxel will drop to a small value after this update. In the extreme case, if no events interacted with voxel j, it is set to zero, producing a "hole" that results in negative bias in the voxel. Subsequent iterations may lead to increases in neighboring voxels to compensate for the holes, however, it appears that this effect is not sufficient to fully compensate for the underestimation, causing net negative bias in the image.

Larger bias was reported on some comparable studies performed on the HRRT with the sinogram-based reconstruction method (Walker *et al.*, 2011). Some implementation features are similar between the sinogram-based algorithms and MOLAR. Thus, comparison between the results in (Walker *et al.*, 2011) with those in the S^{off} experiment will be meaningful. First, it was shown in (Walker *et al.*, 2011) that resolution modeling, which is also included in MOLAR, reduces image bias in low-count reconstructions. With a PSF, backprojection of each event "touches" more voxels, thus, "holes" are less likely to be created. Second, a modified scatter correction method in (Walker *et al.*, 2011) was used to estimate scatter from high-count data and re-scale it to fit the low-count reconstructions, which ideally is similar to the S^{off} experiments in terms of eliminating bias due to the scatter correction. Even so, lower bias was observed in MOLAR than in (Walker *et al.*, 2011). For example, in the S^{off} experiment, bias in three ROIs was −5% or smaller in the reconstruction from 0.2 M true events. However, calculated from [^{11}C]DASB images reconstructions (also a serotonin transporter tracer like [^{11}C]AFM) at the NEC level of 0.2 M, bias in caudate (high activity), ventricular region (low activity), and cerebellum (low activity) were −17%,

+19% and +4% in (Walker *et al.*, 2011). Similarly, in a Hoffman phantom experiment performed by Velden *et al.* 2009, −10% and +30% biases were found in the grey matter and white mater respectively at a NEC level of 2 M. These values were significantly higher than the ROI biases observed in the S^{on} experiment (< 8% in figure 5). It was noted that positive bias was found to be associated with regions with relatively low activity in (Reilhac *et al.*, 2008; Walker *et al.*, 2011; van Velden *et al.*, 2009), while in our results, the ROI bias was always negative. One of the possible reasons is that the resolution model is implemented in MOLAR as a wide-band projection operator (Carson *et al.*, 2003), making MOLAR less susceptible to "hole" effects than the sinogram versions that use a ray-tracing projector and convolution-based resolution model (Walker *et al.*, 2011). Further investigation is needed to understand the source of this difference. The low-bias property of MOLAR suggests that it is an effective image reconstruction platform for quantification accuracy in dynamic imaging using short time frames, and may be useful with clinical and preclinical systems(Jin *et al.*, 2013; Jin *et al.*, 2014).

In addition to overall activity underestimation at low counts, the results in this study were consistent with previous studies by showing some levels of "activity migration" from hot to cold ROIs, i.e., high-activity regions were more negatively biased than low-activity regions. ROI bias depends not only on the numbers of events measured from the region itself, but also the activity distribution within the whole object. In figures 4 and 5, the low-activity ROI (cerebellum) showed less bias than high-activity ROIs (thalamus and striatum) at any given low-count level. Thus, the positivity constraint may still play a role in the bias by causing a smaller negative bias in low-activity regions.

Comparably small low-count bias and variability patterns were found in NEMA phantom reconstructions with MOLAR on the Biograph mCT scanner. The Biograph mCT version of MOLAR has time-of-flight (TOF) capability (Jin *et al.*, 2013), where backprojection of events is restricted to small regions in the image. Intuitively, this might produce an adverse effect in low-count reconstructions based on the aforementioned "hole" theory. However, the poorer resolution of the Biograph mCT means that each event interacts with a wider band of voxels. It will be worthwhile to compare the Biograph mCT version of MOLAR with the manufacturer's reconstruction algorithms to confirm our hypothesis.

Scatter and randoms caused an additional level of bias in the reconstructed images. One possible reason could be inaccuracies in the implementation of scatter correction method, producing minor overestimation of the scatter fraction and underestimation of the images. More work is needed to validate the scatter correction in the simulation experiment. In addition, our experiment compared bias between S^{on} and S^{off} reconstructions at the same NEC level, assuming the effective counts are identical in the two cases. While NEC has been used extensively to compare count levels with different random and scatter fractions, it is unknown whether it is fair to compare low-count bias solely based on NEC levels, especially when using a non-linear algorithm such as OSEM.

Image variability showed a similar dependency to subsets as bias, i.e., variability decreased as the subset number decreased. Our comparison was based on the assumption that images reconstructed under the same effective iterations (iterations x subsets) have the same level of

convergence. However, convergence can be complicated in low-count reconstructions due to the fact that there are too few events in each subset used for image updates. Therefore, larger variability in higher-subset reconstructions might be a consequence of slightly faster convergence toward a biased value at the matched effective iteration number.

Since the number of subsets affected bias and variability in low-count reconstructions, to reduce these effects, a sophisticated subset selection method may be valuable, especially in list mode reconstruction in which the computation time is proportional to the number of events, not the number of subsets. In other words, the increased computation time by using less subsets in low-count frames might be affordable. For example, a relevant count-regulated OSEM was proposed in (Vaissier *et al.*, 2013) for low-count SPECT reconstruction. This algorithm selects the number of subsets based on local count information, i.e., high-count regions were updated with more subsets to accelerate reconstruction, and low-count regions were updated with less subsets to reduce image bias. This method could be adapted to PET reconstruction. On the other hand, as bias is partially attributed to the current scatter correction method, it is worthwhile to characterize the source of this error and design an improved low-count scatter correction strategy in MOLAR.

7. Conclusion

In this study, image bias and variability in low-count OSEM reconstruction was investigated in simulation and phantom studies. Negative image bias was found at extremely low levels of counts in image reconstruction. A few factors found to contribute to the bias of the reconstructed images include larger subset numbers and frame-based scatter correction. Compared to sinogram reconstructions, we only observed small bias in the reconstructions at reasonably low count levels, with the difference compared to other methods presumably due to subtle implementation details in MOLAR. Nevertheless, improving iterative algorithms to control and correct for low-count bias is necessary to maintain quantification consistency in dynamic PET studies.

Acknowledgments

The authors thank Zhongdong Sun for programming support and the staff of the Yale-PET Center for the studies that formed the basis of this work. The authors acknowledge the discussion and advice of Dr. Margaret E. Daube-Witherspoon. This work was supported by Grant No. R01NS058360 from the National Institute of Neurological Disorders and Stroke.

References

Byrne C. Iterative algorithms for deblurring and deconvolution with constraints. Inverse Problems. 1998; 14:1455–67.

Carson RE, et al. Design of a motion-compensation OSEM list-mode algorithm for resolution-recovery reconstruction for the HRRT. Proc IEEE Nucl Sci Symp Med Imag Conf. 2003:3281–5.

de Geus-Oei L-F, et al. Comparison of image-derived and arterial input functions for estimating the rate of glucose metabolism in therapy-monitoring 18F-FDG PET studies. J Nucl Med. 2006; 47:945–9. [PubMed: 16741303]

Erlandsson, K., et al. Low-statistics reconstruction with AB-EMML. Nuclear Science Symposium Conference Record; 2000; IEEE; 2000. p. 249-53.

Fung EK, Carson RE. Cerebral blood flow with [15O] water PET studies using an image-derived input function and MR-defined carotid centerlines. Phys Med Biol. 2013; 58:1903–9. [PubMed: 23442733]

Grezes-Besset, L., et al. Simulation-based evaluation of NEG-ML iterative reconstruction of low count PET data. Nuclear Science Symposium Conference Record, 2007. NSS'07; IEEE; 2007. p. 3009-14.

Hudson HM, Larkin RS. Accelerated Image-Reconstruction Using Ordered Subsets of Projection Data. IEEE transactions on medical imaging. 1994; 13:601–9. [PubMed: 18218538]

Jakoby BW, et al. Physical and clinical performance of the mCT time-of-flight PET/CT scanner. Phys Med Biol. 2011; 56:2375–89. [PubMed: 21427485]

Jian, Y.; Carson, RE. Effect of subsets on bias and variance in low-count iterative PET reconstruction. Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC); 2013; IEEE; 2013. p. 1-4.

Jin X, et al. List-mode reconstruction for the Biograph mCT with physics modeling and event-by-event motion correction. Phys Med Biol. 2013; 58:5567–91. [PubMed: 23892635]

Jin X, et al. Evaluation of Frame-Based and Event-by-Event Motion-Correction Methods for Awake Monkey Brain PET Imaging. J Nucl Med. 2014; 55:287–93. [PubMed: 24434295]

Johansson, J., et al. Quantitative brain imaging using the new, fast iterative histogram-mode reconstruction for the HRRT PET scanner. Nuclear Science Symposium Conference Record, 2007. NSS'07; IEEE; 2007. p. 3463-7.

Lange K, Carson R. EM reconstruction algorithms for emission and transmission tomography. J Comput Assist Tomogr. 1984; 8:306–16. [PubMed: 6608535]

Naganawa M, et al. Tracer kinetic modeling of [(11)C]AFM, a new PET imaging agent for the serotonin transporter. J Cereb Blood Flow Metab. 2013; 33:1886–96. [PubMed: 23921898]

Nuyts J, et al. Reducing loss of image quality because of the attenuation artifact in uncorrected PET whole-body images. J Nucl Med. 2002; 43:1054–62. [PubMed: 12163632]

Planeta-Wilson, B., et al. Quantitative accuracy of HRRT list-mode reconstructions: effect of low statistics. Nuclear Science Symposium Conference Record, 2008. NSS'08; IEEE; 2008. p. 5121-4.

Qi J, Leahy RM. Iterative reconstruction techniques in emission computed tomography. Phys Med Biol. 2006; 51:R541–78. [PubMed: 16861768]

Reilhac A, et al. Simulation-based evaluation of OSEM iterative reconstruction methods in dynamic brain PET studies. Neuroimage. 2008; 39:359–68. [PubMed: 17920931]

Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. IEEE transactions on medical imaging. 1982; 1:113–22. [PubMed: 18238264]

Strother S, et al. Measuring PET scanner sensitivity: relating countrates to image signal-to-noise ratios using noise equivalents counts. IEEE Trans Nucl Sci. 1990; 37:783–8.

Stute, S., et al. Image properties of various ML-based reconstructions of very noisy HRRT data. Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC); 2011; IEEE; 2011. p. 4311-5.

Tzourio-Mazoyer N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002; 15:273–89. [PubMed: 11771995]

Vaissier P, et al. Fast Count-Regulated OSEM Reconstruction with Adaptive Resolution Recovery. IEEE transactions on medical imaging. 2013; 32:2250–61.

van Velden FH, et al. Accuracy of 3-dimensional reconstruction algorithms for the high-resolution research tomograph. J Nucl Med. 2009; 50:72–80. [PubMed: 19091902]

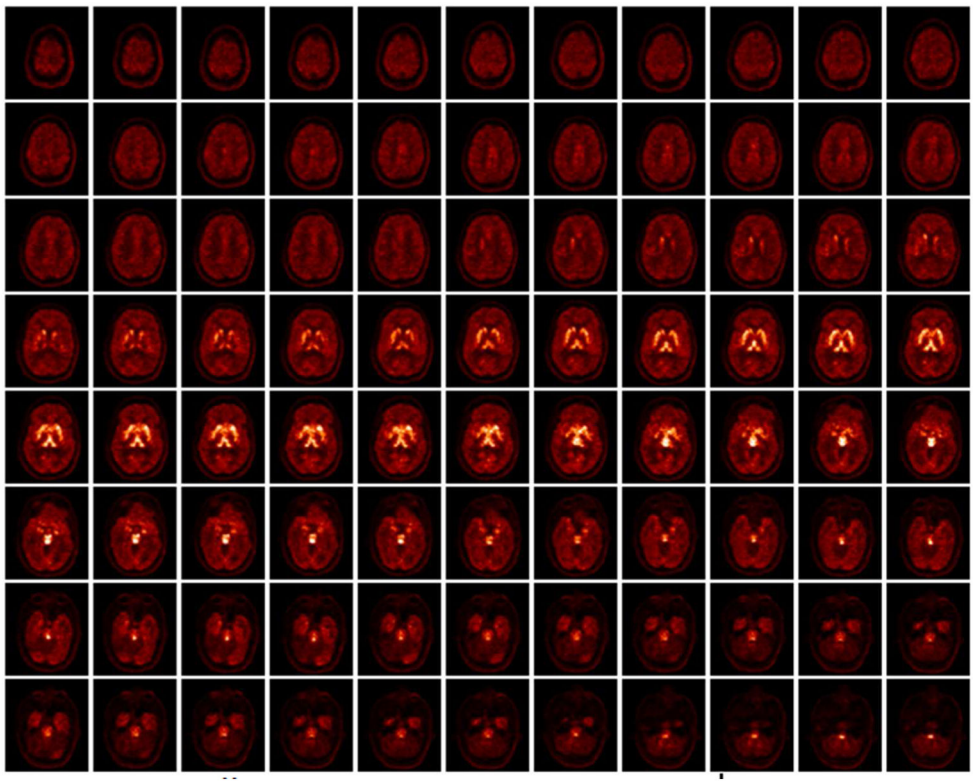
Verhaeghe J, Reader AJ. AB-OSEM reconstruction for improved Patlak kinetic parameter estimation: a simulation study. Phys Med Biol. 2010; 55:6739. [PubMed: 21030748]

Viola P, Wells WM. Alignment by maximization of mutual information. International Journal of Computer Vision. 1997; 24:137–54.

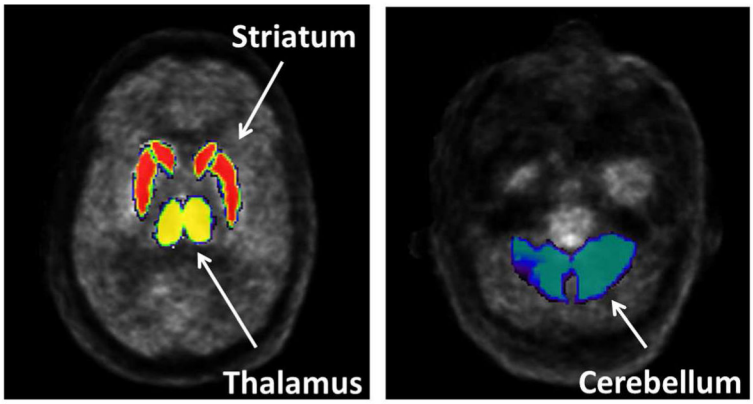
Walker MD, et al. Bias in iterative reconstruction of low-statistics PET data: benefits of a resolution model. Phys Med Biol. 2011; 56:931–49. [PubMed: 21248391]

Watabe H, et al. Noninvasive estimation of the aorta input function for measurement of tumor blood flow with [/sup 15/O] water. IEEE Trans Med Imag. 2001; 20:164–74.

Watson CC, et al. Evaluation of simulation-based scatter correction for 3-D PET cardiac imaging. Ieee T Nucl Sci. 1997; 44:90–7.



(a)



(b)

Figure 1.

Human [^{11}C]AFM brain image used for simulation. (a)Slices across the whole brain. (b)Two Slices of the image with three drawn ROIs.

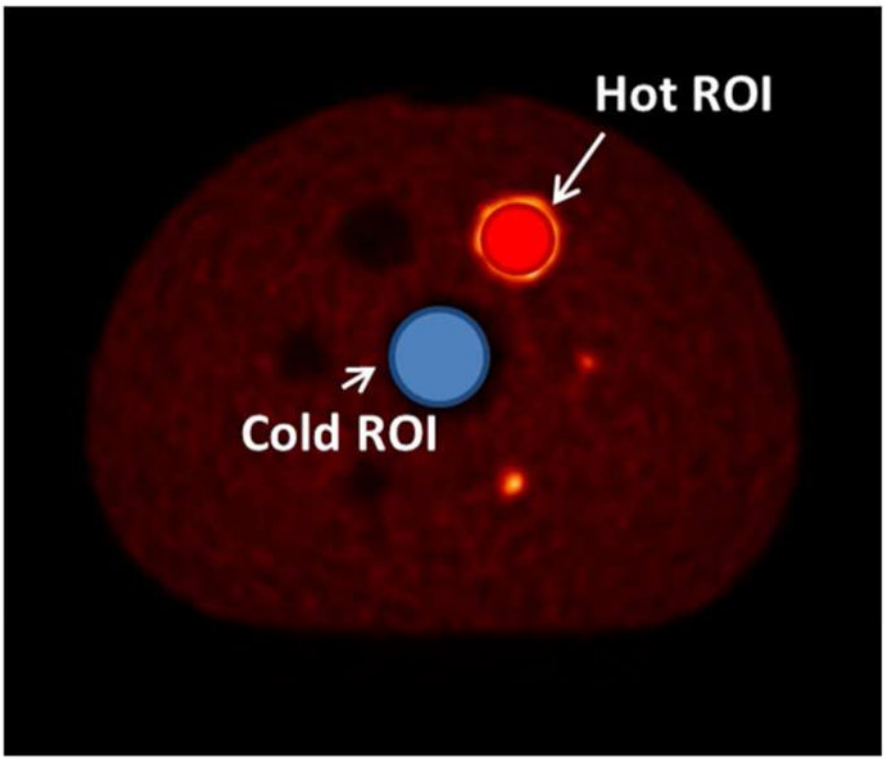


Figure 2.
Hot and cold ROIs on the NEMA phantom image.

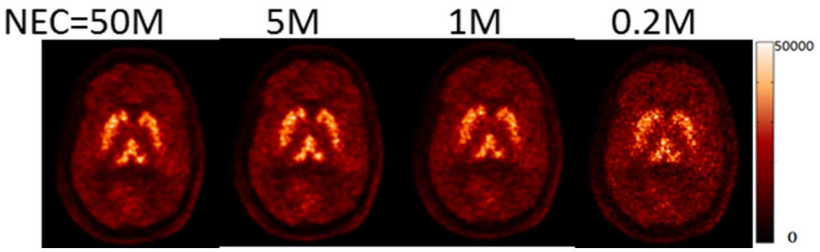


Figure 3.

Mean [^{11}C]AFM brain images in the simulation study (S^{off} experiment without scatter and random effects) reconstructed from various NEC levels (from left to right: 50 M, 5 M, 2 M and 0.2 M) after 2 iterations/ 30 subsets. Images are displayed on a fixed scale.

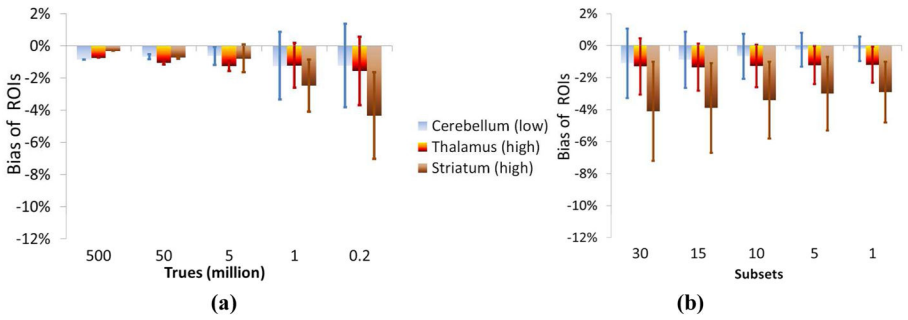


Figure 4.

Brain ROI bias and variability in the HRRT brain simulation study without scatter and random effects. (a) ROI bias and variability (error bars) are shown in reconstructions from various NEC levels. All images for evaluation were obtained from 30 subsets / 2 iterations. (b) ROI bias and variability are shown in reconstructions with various subset numbers. All the images for evaluation are reconstructed with ultra-low-count levels (0.2 M NEC) from 60 effective iterations.

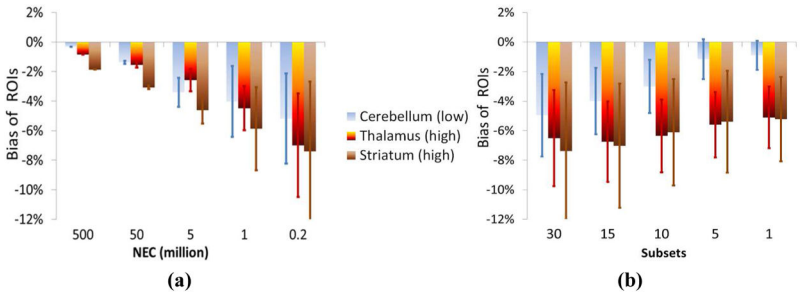


Figure 5.

Brain ROI bias and variability in HRRT brain simulation study with scatter and random effects. (a) ROI bias and variability (error bars) are shown in reconstructions from various NEC levels. All images for evaluation were obtained from 30 subsets/ 2 iterations. (b) ROI Bias and variability are shown in reconstructions with various subset numbers. All the images for evaluation are reconstructed with ultra-low-count data set (0.2 M NEC) from 60 effective iterations.

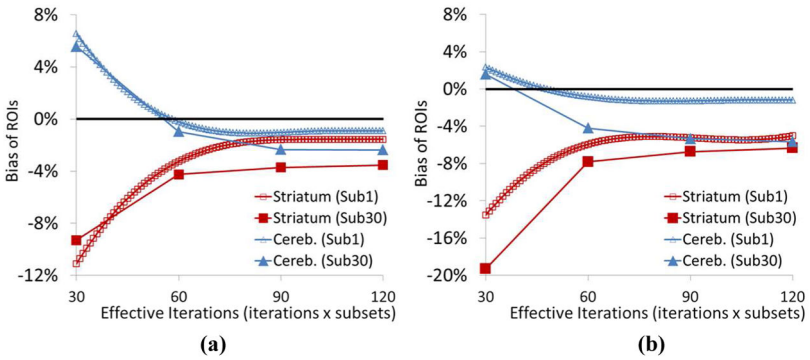


Figure 6.

Convergence of ROI means in simulated ultra-low count (0.2 M NEC) study. Images were reconstructed with 30 (Sub30) and 1 (Sub1) OSEM subsets. % Bias was calculated from Eq. 5. (a) Scatter and random effects not included. (b) Scatter and random effects included.

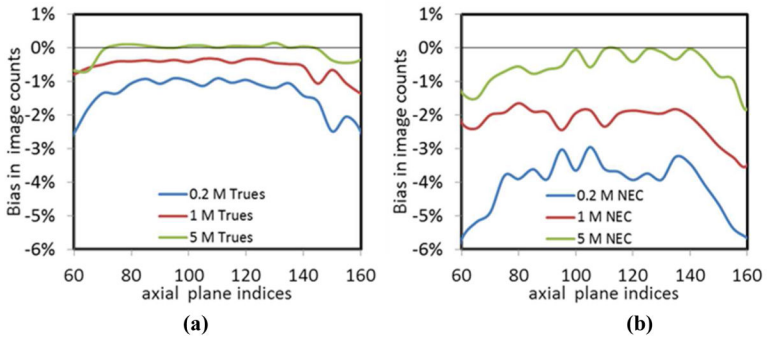


Figure 7.

The axial bias distribution of [^{11}C]AFM brain images reconstructed from simulated data at different count levels with 30 subsets. Image counts in each transverse plane were calculated within the head region, from slice 60 to 160 (of 207). The % bias was calculated between the total reconstructed activity in the low-count images averaged from replicates and the true image. (a) Scatter and random effects not included. (b) Scatter and random effects included.

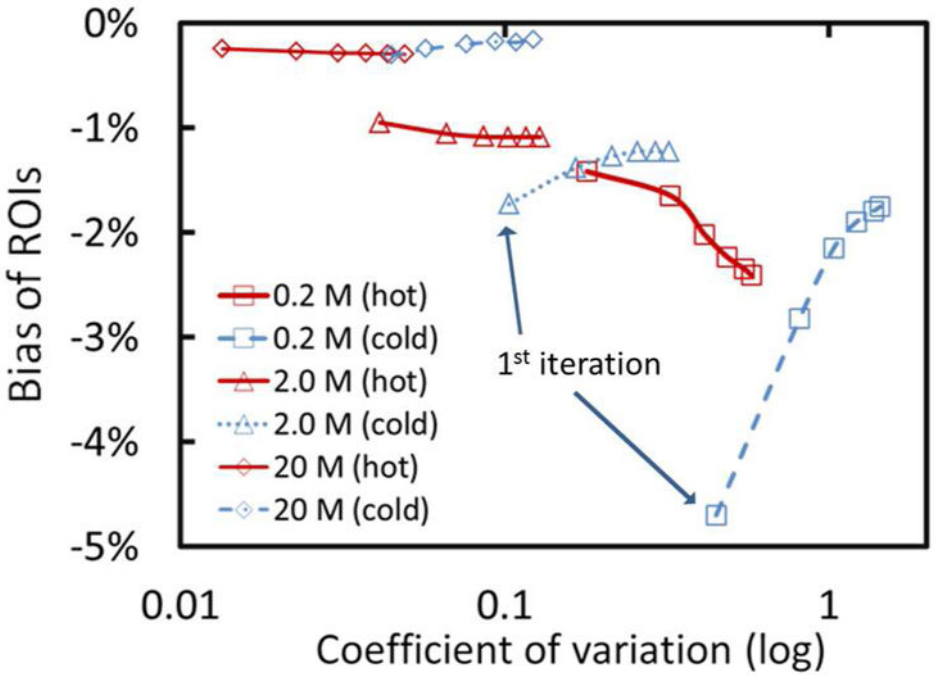


Figure 8.

Bias and CoV of the NEMA phantom images reconstructed from three NEC levels (0.2 M, 2.0 M and 20 M). ROI bias was calculated for hot and cold regions with respect to the values obtained from images reconstructed from 200 M NEC with 1 subset / 126 iterations. Coefficient of variation was calculated across replicates. Each curve in the figure represents the values from 1st to 6th iterations (21 subsets).

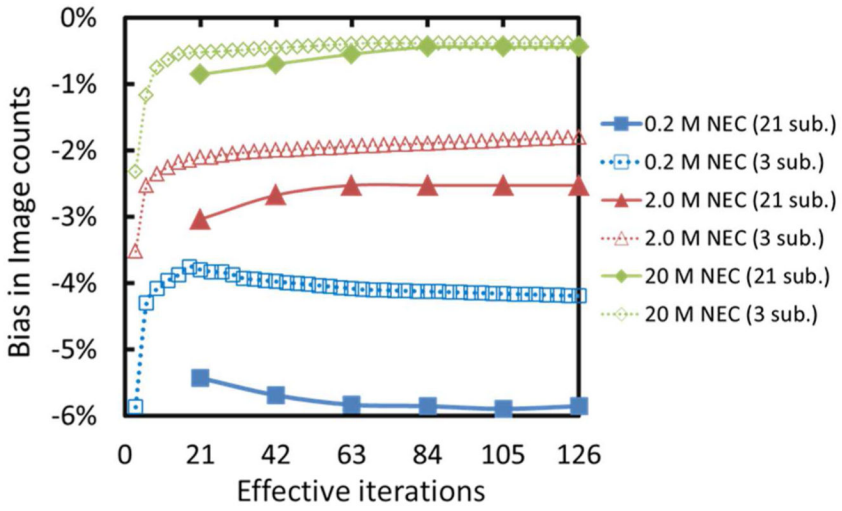


Figure 9.

Bias of total reconstructed image activity as a function of iterations in NEMA phantom reconstructions. Images were reconstructed from three NEC levels (0.2 M, 2.0 M and 20 M) with two subset schemes (3 subsets and 21 subsets). Bias was calculated with respect to the images reconstructed from 200 M NEC with 1 subset / 126 iterations.