Image Properties of Various ML-Based Reconstructions of Very Noisy HRRT Data

Simon Stute, Johan Nuyts, Katrien Van Slambrouck, Mérence Sibomana, Floris van Velden, Ronald Boellaard, and Claude Comtat

***Abstract*–The use of iterative image reconstruction algorithms with resolution modeling allows for reduced partial volume effect without noise increase. However, it is now recognized that EM-ML type algorithms are biased in very low counts images, in particular for cold regions. Alternative ML algorithms that allow for negative image voxels have been proposed to reduce the bias: NEG-ML of Nuyts *et al* and AB-ML of Byrne.**

The aim of this study is to evaluate the NEG-ML and AB-ML algorithms with respect to the EM-ML and 3DRP algorithms for human brain dynamic studies on the high resolution HRRT scanner. As the ground truth is not known, the idea is to distribute the list-mode data into several statistically independent gates of 2ms duration and to compare the summation of the reconstructions of each gate to the reconstruction of the entire list-mode. As both images are produced using the same data, differences are solely due to low count statistics effects. The average number of events per replicate is selected by choosing the number of gates.

A one hour FDG brain PET study was split into different numbers of gates (ranging from 2 to 360). The count statistics of each replica ranged from a 30min to a 10s acquisition, resulting in weakly to extremely noisy data. Two threshold values in the denominators of the NEG-ML equation were used: 1 as originally recommended and 10^{-4} as frequently used for EM-ML. Arbitrary high bound values were used for the AB-ML algorithm. The mean activity concentration was measured in the gray and white matters, based on regions of interest drawn on a MRI of the same subject. As expected, no differences were found for 3DRP between the one hour acquisition and the sum of shorter acquisitions. EM-ML and NEG-ML with the lower threshold lead to an absolute bias ranging from 0.1 to 4.5% for 30min to 10s acquisitions. The bias was reduced to less than 0.5% for AB-ML and NEG-ML with a threshold of 1, for all acquisition durations. Convergence and noise properties were also studied.

The AB-ML and NEG-ML algorithms are almost bias-free alternatives to EM-ML PET reconstruction of data with any noise level. The choice of the denominator threshold of the NEG-ML algorithm seems to play an important role.

I. Introduction

Dynamic positron emission tomography (PET) acquisitions are used to study the radiotracer kinetics inside the body. This consists in dividing the acquisition into several successive frames that are reconstructed to form a series of dynamic images. To accurately measure the radiotracer kinetics, one has typically to reduce the early frames' durations down to 10 seconds, thus producing very noisy data.

The EM-ML algorithm [1] is currently the method of choice for reconstructing emission tomography data. However, it is known to produce significant bias when dealing with very noisy data, mainly due to its positivity constraint in the image domain. Thus the analytical FBP algorithm can still be used in that context to remove bias (no positivity constraint), but produces images with a poorer spatial resolution and potentially more noise. Other interesting alternatives like AB-ML [2] and NEG-ML [3] have been proposed and have been partly studied [4-7]. Though they have received little interest in the past, they gave promising results as they remove the image positivity constraint (allowing negative voxel values) and are iterative, thus allowing for resolution modeling (as opposed to FBP). In this work, we characterized the properties of images reconstructed with such algorithms in the context of very short clinical frames acquired on a patient with the high-resolution HRRT scanner [8]. As the ground truth is not known, we examined the consequences of reconstructing low-statistics data in comparison to high-statistics data sharing the same expectation. An interpretation of these two algorithms is given in [9].

II. Material and Methods

A. PET data

The list-mode data of a 1 hour ^{18}F-FDG brain PET study were used. A post-acquisition gating method was used to split the acquisition into N statistically independent replicates of the acquisition [10]. The gate duration has to be very short (here 2ms) so that the cycle of the N gates can be repeated many times along the acquisition. Each recorded event is attributed to the current running gate. The average counts statistics of each replicate are equal to the counts statistics of the full acquisition divided by N. Also the expectation of any replicate for any N value is the same as the expectation of the whole dataset divided by N, but its noise level is different. The idea is to compare the distribution of the reconstructions of the N replicates to the reconstruction of the entire list-mode (thereafter referred as "static"). As the sum of the N images is produced using the same data as the static image, differences are solely due to low-statistics effects.

All replicates for a given value of N were first rebinned into 3D sinograms with a span of 9 (axial compression) and

Manuscript received November 15, 2011.

S. Stute and C. Comtat are with CEA, Service Hospitalier Frédéric Joliot, 91401 Orsay , France (e-mail: simon.stute@cea.fr).

J. Nuyts and K. Van Slambrouck are with the Department of Nuclear Medicine, K. U. Leuven, B-3000 Leuven, Belgium.

M. Sibomana is with the Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet, Copenhagen University Hospital, Denmark.

F. van Velden and R. Boellaard are with the Department of Nuclear Medicine & PET Research, VU University Medical Center, 1007 MB Amsterdam, The Netherlands.

maximum ring difference of 67. A smooth estimate of randoms was deduced from the whole collection of delayed events [11], and then divided by *N*. The scatter estimate was calculated with the single-scatter simulation method using the whole data set, and then also divided by *N*. In this way, the same random and scatter corrections were applied to all *N* replicates. Attenuation was corrected using a transmission scan, and normalization was also taken into account.

Different numbers of gates *N* were considered, corresponding to different counts statistics: 1 gate (corresponding to the whole acquisition of 60min), 2 gates (each equivalent to a count statistics of a 30min acquisition), 12 gates (5min), 30 gates (2min), 60 gates (1min), 120 gates (30s), 360 gates (10s).

The mean NECs per sinogram bin for the whole acquisition is about 5.2 counts. As any replicates is deduced from the entire list-mode data, the NECs for a given value of *N* is 5.2/*N* (given in table I). This NECs value was obtained inside the head (i.e. considering only sinogram bins where the attenuation correction factor is higher than 1.2). The mean number of prompts per sinogram bin (inside the head) for each value of *N* is given in table I.

B. Reconstruction algorithms

As an unbiased reference, the analytical 3DRP [12] algorithm was used. However, the HRRT sinograms contain gaps. In order to use analytical reconstructions, they have to be filled. In this study, for any number of gates, the gaps in the fully pre-corrected sinograms were estimated using a projection of an EMML reconstructed image using the whole 1 hour dataset. The cut-off was set at the Nyquist frequency, and no apodization window was used.

As the current standard algorithm, the EM-ML algorithm was used. It is described in its additive form by the following equation:

$$\lambda_j^{k+1} = \lambda_j^k + \frac{\lambda_j^k}{\sum_i X_{i,j}} \sum_i X_{i,j} \frac{y_i - \bar{y}_i^k}{\bar{y}_i^k}, \qquad (1)$$

where λ, y and X are respectively the image, the prompt sinogram and the system matrix, and i, j and k the line-of-response, voxel and iteration indices. Also, the Ordinary-Poisson model [13] was used, so that the modeled data $\bar{y}$ and the system matrix X are defined as follows:

$$\bar{y}_i^k = \sum_j X_{i,j} \lambda_j^k + r_i \qquad X_{i,j} = M_i G_{i,j}, \qquad (2)$$

TABLE I
MEAN NECS AND NUMBER OF PROMPTS PER SINOGRAM BIN

N	Duration (s)	NECs/bin	Prompts/bin
1	3600	5.2E+0	1.7E+1
2	1800	2.6E+0	8.4E+0
12	300	4.3E-1	1.4E+0
30	120	1.7E-1	5.6E-1
60	60	8.7E-2	2.8E-1
120	30	4.3E-2	1.4E-1
360	10	1.4E-2	4.7E-2

Only the sinogram bins with an attenuation correction factor higher than 1.2 were considered.

where *r* is the background term containing the non-normalized random plus scatter estimates, *M* is a multiplicative term reflecting the attenuation and normalization effects, and *G* is the projector. A pair of ray-driven / voxel-driven projectors was used for the forward and backward projections, respectively, without considering PSF modeling.

As the first ML alternative to EM, the NEG-ML algorithm as described by Nuyts *et al* [3] was used:

$$\lambda_j^{k+1} = \lambda_j^k + \Lambda_j^k \sum_i X_{i,j} \frac{y_i - \bar{y}_i^k}{\max(\bar{y}_i^k, \psi)}$$
$$\Lambda_j^k = \max\left(\frac{1}{\sum_i \frac{X_{i,j}}{\max(y_i, \psi)} \sum_\xi X_{i,\xi}}, \frac{\lambda_j^k}{\sum_i X_{i,j}} \right). \qquad (3)$$

Here, Λ_j^k represent the step size. The right term of the "max" function corresponds to the EM-ML step size (see equation (1)), the other to the NEG-ML one. The choice of the maximum one allows to accelerate the convergence of the algorithm. For the first iteration, the updates of the image were always performed using the right term (i.e. the EM-ML one). We considered the NEG-ML algorithm using a denominator threshold ψ either fixed at 1 (thereafter denoted as NEG-ML), as recommended in [3], or fixed at 10^{-4} (denoted as NEG-ML$_{low}$), as values lower than 1 are often used in EM-ML implementations [7].

The second ML alternative that was used is the AB-ML algorithm described by Byrne [2] as follows:

$$\lambda_j^{k+1} = \frac{A_j^k b + B_j^k a}{A_j^k + B_j^k}$$
$$A_j^k = \frac{\lambda_j^k - a}{\sum_i X_{i,j}} \sum_i \frac{y_i - \bar{a}_i}{\bar{y}_i^k - \bar{a}_i} \qquad B_j^k = \frac{b - \lambda_j^k}{\sum_i X_{i,j}} \sum_i \frac{\bar{b}_i - y_i}{\bar{b}_i - \bar{y}_i^k}, \qquad (4)$$
$$\bar{a}_i = a \sum_j X_{i,j} \qquad \bar{b}_i = b \sum_j X_{i,j}$$

with the lower *a* and upper *b* image bounds respectively set to sufficiently high negative and positive values in order to avoid any constraint on the algorithm (here, a=-b, set to 5 orders above the maximum voxel intensity for the static N=1 reconstruction). Providing an initial image between these bounds ($a \le \lambda_j^0 \le b, \forall j$), the algorithm ensures that all subsequent images will also be included into this interval.

An ordered subset acceleration scheme with 16 subsets was run for at least 10 iterations for all iterative algorithms, and for all values of *N*. No post-smoothing was applied on the reconstructed images.

C. Evaluation

For each choice of *N*, a standard-deviation image was computed across the *N* replicates as well as a summed image. Comparing the static image (N=1) to the summed images for other values of *N*, allowed to qualitatively visualize the bias imputed by reconstructing low-statistics data. Comparing the standard-deviation images allowed to visualize the noise properties of the different algorithms.

For the quantitative study, regions of interest (ROIs) for the gray and white matter were drawn on a MRI scan of the same patient. After registration, these ROIs were reported on the PET images to compute mean intensity values. For a given ROI, value of N and replicate i_N, such a measurement was denoted $\langle ROI \rangle_{i_N}$.

From these measurements, we first computed for each ROI, the sum through the replicates:

$$\mathrm{Sum}_{ROI,N} = \sum_{i_N=1}^{N} \langle ROI \rangle_{i_N} . \quad (5)$$

Then, for a given ROI, we computed the bias between the static measurement (i.e. N=1) and the sum (5) for a different value of N:

$$\mathrm{Bias}_{ROI,N} = \frac{\mathrm{Sum}_{ROI,N} - \langle ROI \rangle_{static}}{\langle ROI \rangle_{static}} . \quad (6)$$

Finally, we computed the standard-deviation of these measurements, normalized to the static measurement:

$$\mathrm{Stdv}_{ROI,N} = \frac{\sqrt{\frac{1}{N}\sum_{i_N=1}^{N} \left(\mathrm{Sum}_{ROI,N} - N.\langle ROI \rangle_{i_N}\right)^2}}{\langle ROI \rangle_{static}} . \quad (7)$$

III. Results

A. Qualitative evaluation

Figure 1 shows reconstructions for N=1 (top row) and the sum of the reconstructions of all replicates for N=120 (bottom row). The five algorithms are shown, all taken at the 20th iteration (except for 3DRP). The color scale is the same for all images. We first observe that for N=1, all images are quantitatively very similar inside the head. For 3DRP (column A), the images for N=1 and N=120 can not be distinguished, so as for AB-ML (column C). However they present some streak artifacts. For EM-ML (column B) a clear over-estimation of cold regions and under-estimation of hot regions can be seen in the N=120 reconstructions compared to the static N=1 reconstruction, predicting some bias. The same observations can be made for the NEG-ML$_{low}$ algorithm (column E) that is hardly distinguishable from EM-ML. For the NEG-ML algorithm (column D), no substantial difference can be seen, except outside the head.

Figure 2 (top row) shows the reconstructions of only one replicate for N=120 (also for the 20th iteration). The 3DRP, AB-ML and NEG-ML images look much noisier than the EM-ML and NEG-ML$_{low}$ ones (even though the formers

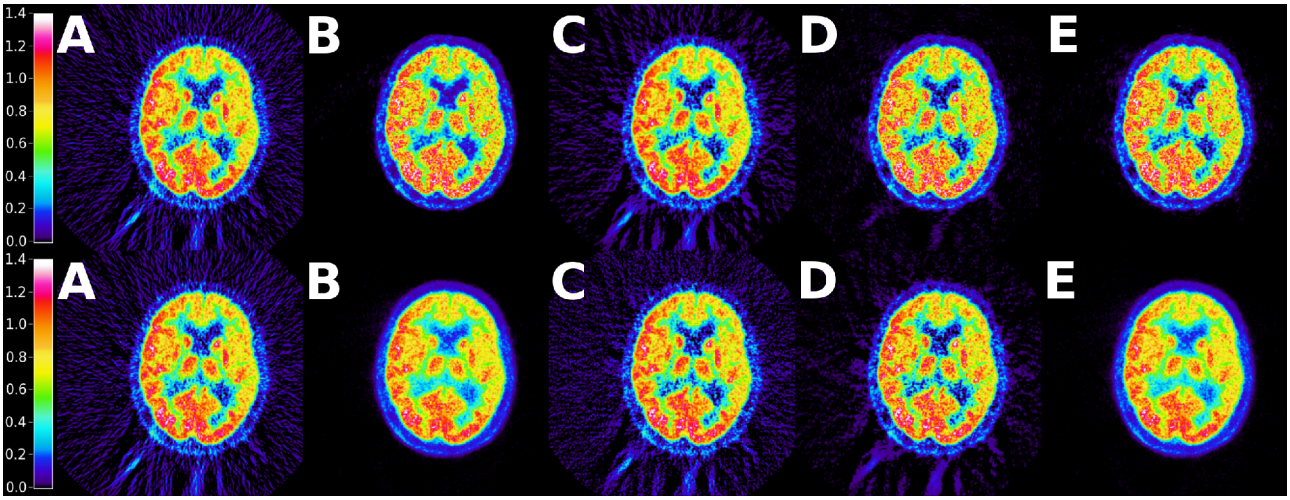

Fig. 1. Reconstructions for N=1 (top row) and the sum of the reconstructions of all replicates for N=120 (bottom row). The five algorithms are shown: A. 3DRP, B. EMML, C. ABML, D. NEGML, E. NEGML$_{low}$. The color scale is the same for all images.

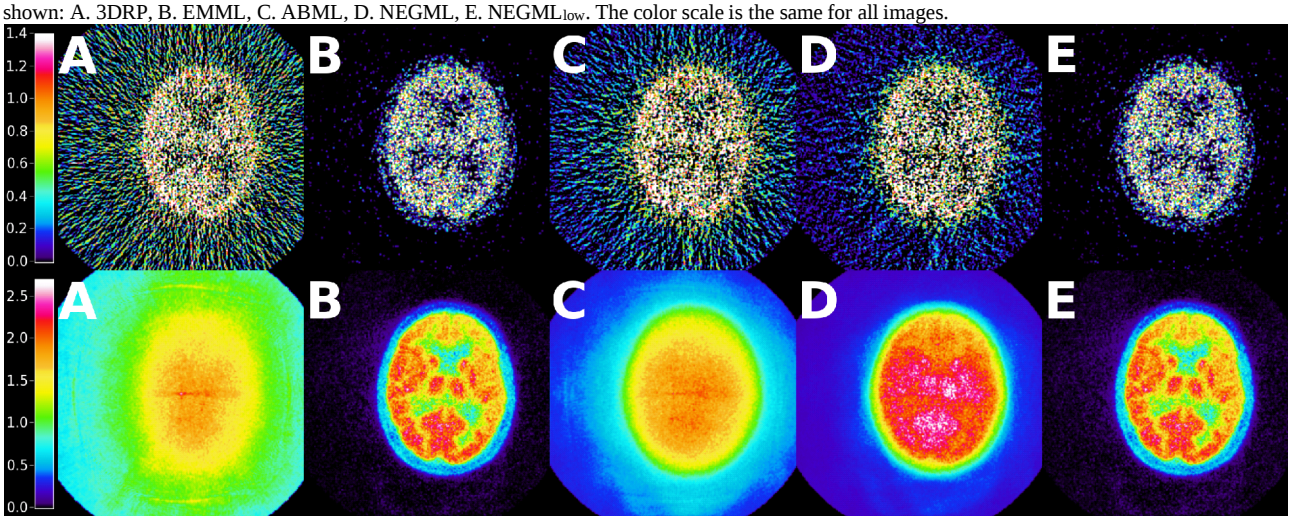

Fig. 2. Reconstructions for N=120 (30s scan duration) of one replicate (top row) and standard-deviation images computed through the N=120 replicates reconstructed independently (bottom row). The five algorithms are shown: A. 3DRP, B. EMML, C. ABML, D. NEGML, E. NEGML$_{low}$. The color scale for top row is the same as in figure 1. The color scale for the bottom row is larger.

exhibit residual activity outside the patient). Also, the EM-ML and NEG-ML images visually seem to get better contrast between hot and cold regions. Keeping in mind that these images would be those actually used and viewed when performing dynamic studies, the EM-ML and NEG-ML$_{low}$ images visually appear to be a better choice, even though the figure 1 predicts the highest bias in its.

Figure 2 (bottom row) shows the standard-deviation images computed through the N=120 replicates reconstructed independently until 20 iterations. These images reflect the noise (i.e. variance) properties of each algorithm. For 3DRP, the noise (here defined as the standard-deviation) inside the head is almost independent of the mean signal, as opposed to EM-ML which is proportional to the mean signal (as already shown by Barrett *et al* [14,15]). AB-ML and NEG-ML are more similar to 3DRP in that the noise is rather independent of the mean signal. However the noise level outside of the head is lower than for 3DRP. Also, the noise level for NEG-ML inside the head is higher than for AB-ML for this same iteration (but maybe due to the slower convergence of AB-ML, see Fig. 3). Finally, as in Fig. 1, the NEG-ML$_{low}$ algorithm is hardly distinguishable from EM-ML.

B. Quantitative evaluation

Figure 3 shows the convergence of the ROI sum values (equation 5) in white and gray matter (top and bottom rows, respectively), for all values of N and for the EM-ML (blue curves), AB-ML (red curves) and NEG-ML (green curves) algorithms. For all algorithms, the global convergence behavior is the same for the white and gray matters, except that they are reversed. For EM-ML, the longer the scan duration (i.e. the more the number of counts), the slower the convergence. The opposite behavior is seen for NEG-ML. For AB-ML, the convergence speed seems to be independent of the scan duration. For both AB-ML and NEG-ML, the convergence speed is globally reduced compared to EM-ML, AB-ML being the slowest. For EM-ML, the ROI value at convergence differs for each scan duration (sign of bias). This is not the case for AB-ML since all curves are superimposed. Also for NEG-ML, the curves for all scan durations seem to converge to the same value (though more iterations should be ran to confirm this statement). Finally, the results for NEG-ML$_{low}$ (not shown) are identical to those for EM-ML.

Figure 4 shows the bias as defined in equation (6) in white and gray matter (top and bottom graphs, respectively) as a function of the scan duration (i.e. values of N). In this figure, the 20th iteration was used except for N=360 for which the 10th iteration was used (no more iterations were ran for this scan duration). The 3DRP shows zero bias for all scan durations, as expected. Under 2min scan duration, the EM-ML and NEG-ML$_{low}$ algorithms show more than 1% bias in both ROIs (up to 4.5% for 10s scan duration). The AB-ML algorithm performs well and is quite stable with respect to the scan duration (except for 10s scan duration, though the bias is less than 0.5%). The NEG-ML algorithm performs well in the white matter, but less in the gray matter (though the bias is around 0.5%). However, the dependance of the NEG-ML bias on the scan duration is unclear.

Figure 5 shows trade-offs between bias and standard-deviation as defined in equation (7), as a function of the number of iterations, for N=120 (i.e. 30s scan duration). The static measurements in (6) and (7) were always performed on the 30th iteration. EM-ML converges to 3% bias. The convergence is slower for AB-ML and NEG-ML (as seen in Fig. 3), but zero bias is reached. At convergence, AB-ML shows more noise than EM-ML in the white matter (cold region), but slightly less in the gray matter (hot region) (probably due to the independence of the noise on the mean signal, see Fig. 2). NEG-ML shows a little increase of noise in both ROIs compared to the other algorithms. For a fixed bias, AB-ML always shows less noise than NEG-ML. As expected, 3DRP is unbiased, but interestingly, it also shows less noise than all iterative algorithms. The results for NEG-ML$_{low}$ (not shown) are identical to those for EM-ML.

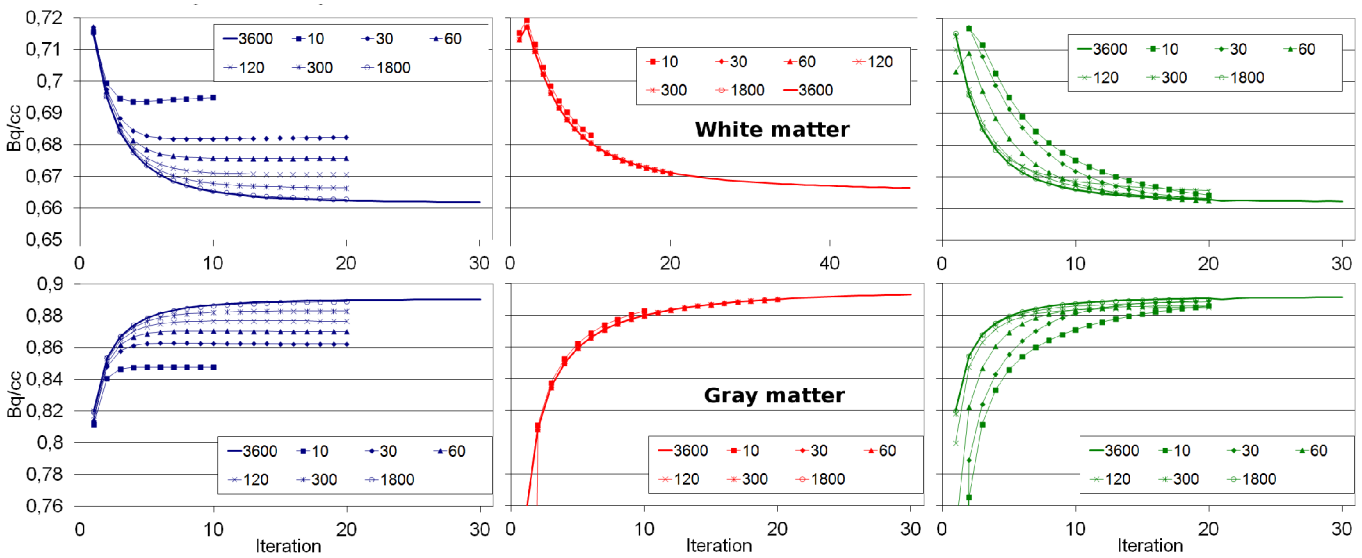

Fig. 3. Convergence of the ROI sum values as defined in (5) for all values of N as a function of the number of iterations. The top row corresponds to the white matter ROI and the bottom row to the gray matter ROI. From left to right, the EM-ML (in blue), AB-ML (in red) and NEG-ML (in green) algorithms are shown.

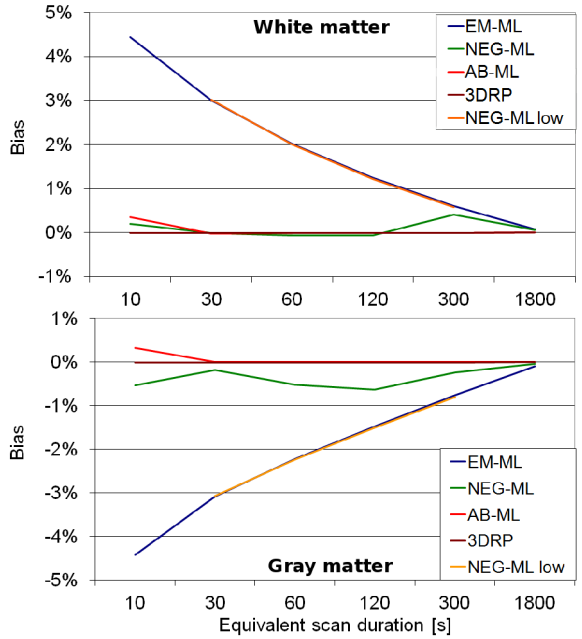

Fig. 4. Bias as defined in equation (6) in white/gray matter (top/bottom graph) as a function of the scan duration (i.e. value of N). Iteration 20 was used except for N=360 (iteration 10).

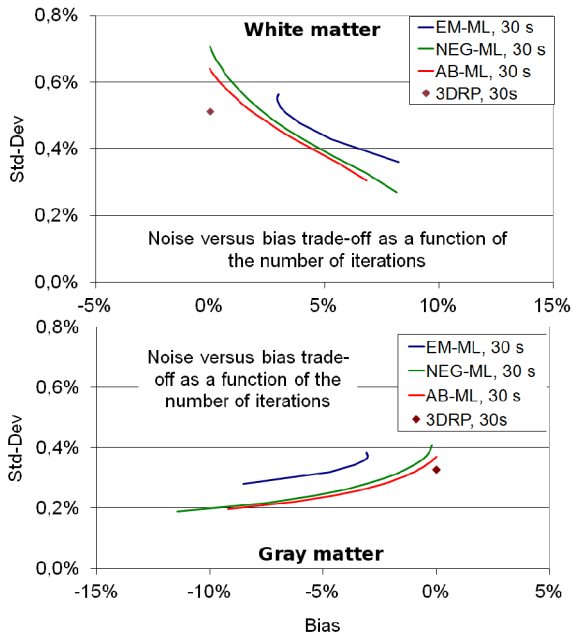

Fig. 5. Trade-off between bias versus standard-deviation as defined in equation (7) as a function of the number of iterations, for N=120 (i.e. 30s equivalent scan duration). Top/bottom graph is for white/gray matter.

IV. Summary

In order to summarize all the results obtained in this study, we extract here a list of the most relevant points:

- EM-ML is biased for low-statistics data.
- AB-ML is almost unbiased.
- NEG-ML highly reduces bias when using ψ=1 (probably almost unbiased if convergence is reached) but is similar to EM-ML when using ψ= 10^{-4} (other values of ψ may have to be considered).
- The convergence speed for AB-ML and NEG-ML is reduced compared to EM-ML, AB-ML being the slowest.
- The convergence speed for AB-ML is independent of the scan duration, thus of the number of counts.
- The noise of AB-ML is similar to 3DRP (almost independent of the mean signal) so as for NEG-ML with a slightly increased noise level at fixed bias.

The AB-ML and NEG-ML algorithms are almost bias-free alternatives to EM-ML PET reconstruction of data with any noise level. Besides, AB-ML has one advantage compared to NEG-ML: the independence of the convergence behavior on the scan duration. Finally, see [9] for a further discussion on the parameters selection for NEG-ML (ψ) and AB-ML (a and b), and an interpretation of each algorithm.

References

[1] L. S. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imaging, vol. MI-1, pp. 113-122, 1982.

[2] C. Byrne, "Iterative algorithms for deblurring and deconvolution with constraints," Inverse Problems, vol. 14, pp. 1455-1467, 1998.

[3] J. Nuyts, S. Stroobants, P. Dupont, S. Vleugels, P. Flamen and L. Mortelmans, "Reducing loss of image quality because of the attenuation artifact in uncorrected PET whole-body images," J. Nucl. Med., vol. 43, pp. 1054-1062, 2002.

[4] K. Erlandsson, D. Visvikis, W. A. Waddington, I. Cullum, P. H. Jarritt and L. S. Polowsky, "Low-statistics reconstruction with AB-EMML," IEEE Nucl. Sci. Symp. Conf. Rec., vol. 2, pp. 249-253, 2000.

[5] L. Grezes-Besset, J. Nuyts, R. Boellard, I. Buvat, C. Michel, C. Pierre, N. Costes and A. Reihlac, "Simulation-based evaluation of NEG-ML iterative reconstruction of low count PET data," IEEE Nucl. Sci. Symp. Conf. Rec., vol. 4, pp. 3009, 2007.

[6] E. Dusch, C. Comtat and R. Trébossen, "Simulation-based evaluation of OSEM reconstruction bias on low activity PET data for the HRRT scanner," IEEE Nucl. Sci. Symp. Conf. Rec., pp. 2770, 2009.

[7] J. Verhaeghe and A J Reader, "AB-OSEM reconstruction for improved Patlak kinetic parameter estimation: a simulation study," Phys. Med. Biol., vol. 55, pp. 6739-6757, 2010.

[8] K. Wienhard et al, "The ECAT HRRT: performance and first clinical application of the new high resolution research tomograph," IEEE Trans. Nucl. Sci., vol. 49, pp. 104-110, 2002.

[9] J. Nuyts, S. Stute, K. Van Slambrouck, F. van Velden, R. Boellaard and C. Comtat, "Maximum-likelihood reconstruction based on a modified Poisson distribution to reduce bias in PET," IEEE Nucl. Sci. Symp. Conf. Proc., NSS-MIC 2011, MIC21.S-288, 2011.

[10] M. D. Walker, M. C. Asselin, P. J. Julyan, M. Feldmann, P. S. Talbot, T. Jones and J. C. Matthews, "Bias in iterative reconstruction of low-statistics PET data: benefits of a resolution model," Phys. Med. Biol., vol. 56, pp. 931-949, 2011.

[11] L. G. Byars, et al, "Variance reduction on randoms from delayed coincidence histograms for the HRRT," IEEE Nucl. Sci. Symp. Conf. Rec., vol. 5, pp. 2622-2626, 2005.

[12] P. E. Kinahan and J. G. Rogers, "Analytic 3D image reconstruction using all detected events," IEEE Trans. Nucl. Sci., vol. 36, pp. 964-968, 1989.

[13] D. G. Politte and D. L. Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography," IEEE Trans. Med. Imaging, vol. 10, pp. 82-89, 1991.

[14] H. H. Barrett, D. W. Wilson and B. M. Tsui, "Noise properties of the EM algorithm: I. Theory," Phys. Med. Biol., vol. 39, pp. 833-846, 1994.

[15] B. M. Tsui, D. W. Wilson and H. H. Barrett, "Noise properties of the EM algorithm: II. Monte Carlo simulations," Phys. Med. Biol., vol. 39, pp. 847-871, 1994.