Evaluation of Three MRI-based Anatomical Priors for Quantitative PET Brain Imaging

Kathleen Vunckx[1], Ameya Atre[1], Kristof Baete[1], Anthonin Reilhac[2], Christophe M. Deroose[1], Koen Van Laere[1] and Johan Nuyts[1]

***Abstract*—In emission tomography, image reconstruction and therefore also tracer development and diagnosis may benefit from the use of anatomical side information obtained with other imaging modalities in the same subject, as it helps to correct for the partial volume effect. One way to implement this, is to use the anatomical image for defining the a-priori distribution in a maximum-a-posteriori (MAP) reconstruction algorithm. In this contribution, we use the PET-SORTEO Monte Carlo simulator to evaluate the quantitative accuracy reached by three different anatomical priors when reconstructing positron emission tomography (PET) brain images, using volumetric magnetic resonance imaging (MRI) to provide the anatomical information. The priors are: 1) a prior especially developed for FDG PET brain imaging, which relies on a segmentation of the MR-image [1]; 2) the joint entropy-prior [2]; 3) a prior that encourages smoothness within a position dependent neighborhood, computed from the MR-image. The latter prior was recently proposed by our group in [3], and was based on the prior presented by Bowsher *et al.* [4]. The two latter priors do not rely on an explicit segmentation, which makes them more generally applicable than a segmentation-based prior. All three priors produced a compromise between noise and bias that was clearly better than that obtained with postsmoothed maximum likelihood expectation maximization (MLEM) or MAP with a relative difference prior. The performance of the joint entropy prior was slightly worse than that of the other two priors. The performance of the segmentation-based prior is quite sensitive to the accuracy of the segmentation. In contrast to the joint entropy-prior, the Bowsher-prior is easily tuned and does not suffer from convergence problems.**

***Index Terms*—Image reconstruction, anatomical priors, PET, MRI, quantitative accuracy.**

I. Introduction

When the finite spatial resolution of the positron emission tomography (PET) or single photon emission computed tomography (SPECT) system is modeled during maximum likelihood reconstruction, the reconstruction algorithm will automatically apply recovery correction. However, because very high spatial frequencies have been lost during detection, and because high frequencies converge more slowly than low frequencies, this recovery correction is imperfect and produces Gibbs over- and undershoots near edges. An anatomy-based prior can be used to apply smoothing within the tissues but not across the tissue boundaries. This reduces or eliminates the Gibbs over- and undershoots near the tissue boundaries, making the edges steeper and the activity within the different tissues smoother. If there is a strong correspondence between anatomy and tracer uptake (as is often the case, in particular in brain imaging), the resulting PET image will have lower noise and improved quantification. The use of a finite resolution model during reconstruction is essential to ensure that the sharp boundaries promoted by the prior are not in conflict with the smooth boundaries observed in the data.

Many maximum-a-posteriori (MAP) or penalized-likelihood algorithms have been proposed for PET or SPECT reconstruction with anatomical information [1]–[11]. A more complete overview can be found in [8]. Here, we will only briefly discuss the reconstruction algorithms that are studied in this work. Baete *et al.* [1], [10] developed and evaluated a method called A-MAP, which is based on a segmented magnetic resonance (MR) image. Such a segmentation-based method produces excellent results provided that an accurate (and well registered) segmentation is available. We selected this method over other segmentation-based algorithms, because it was originally designed for the application studied in this work. Somayajula *et al.* [11] proposed a prior based on mutual information between the PET and the anatomical image, thus avoiding the need for explicit segmentation. In [2] it was shown that the mutual information based prior may introduce bias, a problem which is avoided by using the joint entropy (JE) instead (i.e. by deleting the marginal entropy terms in the mutual information). Bowsher *et al.* [4] proposed a prior which encourages smoothing over an anatomy-dependent neighborhood, defined by selecting a set of most similar neighbors in the anatomical image. We recently showed that an intuitive modification of this prior leads to a further increase in the quantitative accuracy of the reconstructed image [3]. This prior acts locally, whereas the JE-prior acts globally.

In this paper, three reconstruction algorithms with anatomical priors were compared: the anatomical prior used in A-MAP (which we will call the A-prior) [10], the joint entropy prior [2] and the modified version [3] of the prior originally presented by Bowsher *et al.*. The latter will be referred to as the 'Bowsher'-prior. The priors were evaluated using realistic PET simulations created with the PET-SORTEO Monte Carlo simulator [12], which has been carefully validated [13] and has been used by several other authors. The segmentations from the BrainWeb database at McGill University (Montreal,

[1] Nuclear Medicine, K.U. Leuven, B-3000 Leuven, Belgium, [2] Biospective Inc., Montreal, Quebec, Canada. A. Atre was with [1], but is now with Production Engineering, Machine Design and Automation, K.U. Leuven, B-3000 Leuven, Belgium. A. Reilhac was with [2], but is now with Autralian Nuclear Sciences & Technology Organization, Lucas Heights NSW 2234, Australia.

This work is supported by SBO grant 060819 "Quantiviam" of IWT, by IAP-grant P6/38 of the Belgian Science Policy, and by the Research Foundation - Flanders.

Canada), [14], [15] were used to generate realistic software PET brain phantoms, which encompass ground truth knowledge about both the activity distribution and the segmentation. The performance of the algorithms was assessed using a normal brain phantom and a brain phantom with reduced PET tracer uptake in multiple cortical and subcortical gray matter lesions, which are not visible in the corresponding MR-image from the BrainWeb database. In these simulations, it was assumed that the registration errors between PET and MRI are negligible. No attempt was done to evaluate robustness in the presence of registration errors, in contrast to the work presented in [7], [16]. This assumption is reasonable for PET brain imaging with ^{18}F-FDG or other tracers with a global cortical uptake, as errors below 3 mm were reported for rigid registration of PET/MRI brain images using state-of-the-art registration algorithms [17]. In addition, new hybrid PET/MRI systems that simultaneously acquire functional and anatomical data will be able to provide perfectly registered images [18].

The performance of the anatomical priors is evaluated based on their quantitative accuracy. To avoid overtuning for one specific application, we aim for reconstruction parameter settings that yield a good 'all-round' image that can be used for visual analysis, quantification, and in case of dynamic imaging, for kinetic modeling. To achieve that goal, we propose a figure of merit balancing between accurate and low noise gray matter voxel intensity estimates and accurate region-of-interest quantification. The former is crucial, e.g., to ensure the soundness of automatic voxel-based analyses [19], [20], for discriminating true changes in signal from changes due to morphologic changes (atrophy due to Alzheimer's disease [21], focal cortical dysplasia in patients with epilepsy [22], etc.), and to yield an accurate asymmetry index in patients with epilepsy [22]. The latter is especially relevant for region-based analyses in all neurodegenerative diseases [23], [24], and of utmost importance when normalizing images with respect to a reference region [21]. For lesion detection performance, a well-designed observer study is required. This is considered out of the scope of this work.

This paper is organized as follows. In Section II, the various theoretical aspects of this study are elaborated. First, we explain how realistic software brain phantoms were created (Section II-A). Second, the mechanisms of the three anatomical priors, evaluated in this work, are reviewed (Section II-B). Third, objective figures of merit to evaluate the quantitative accuracy of the different reconstruction images are presented (Section II-C). Fourth, we describe how realistic PET simulations were obtained (Section II-D). Fifth, details about the reconstruction implementation are reported (Section II-E). Sixth, we explain how the algorithms are ranked based on the quantitative accuracy they reach when reconstructing the realistic PET and MR data (Section II-F). Last, information about the performed bias-noise analysis is provided (Section II-G). In Section III, first some example images are shown, obtained with the different reconstruction algorithms (Section III-A). Second, the outcome of the comparative quantitative accuracy study is presented (Sections III-B). Next, the results of the bias-noise analysis of the gray matter voxels, the analysis of the lesion signal-to-noise ratio and the bias-noise analysis of the lesion volumes-of-interest (VOIs) are provided (Sections III-C to III-E). These results are further discussed in Section IV and the main conclusion of this evaluation study is formulated in Section V.

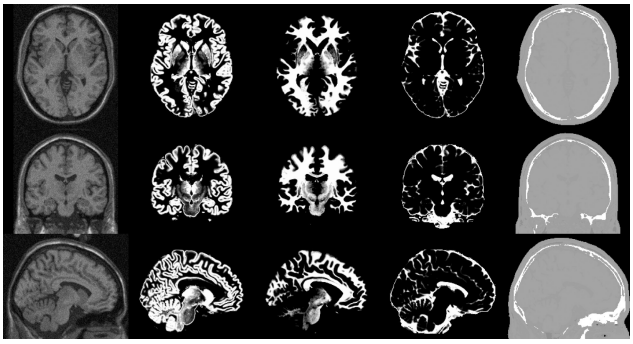

Fig. 1. Orthogonal slices of the phantom. Left column: simulated noisy MR-image. Three central columns: the tissue classes for gray matter, white matter and cerebrospinal fluid. Last column: the PET attenuation image. All images were sampled with $1 \times 1 \times 1$ mm^3 voxels.

II. Methods

A. Brain Images

Figure 1 shows some orthogonal image sections through images from the BrainWeb database, and through the attenuation map that was derived from it. The mean image of 27 high resolution T1-weighted gradient-echo MR-scans of the same volunteer has been segmented in many classes, three of which are shown in the figure (central columns): gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Among the other classes are bone, skin, etc., providing at least one label for every voxel in the image. The BrainWeb also provides simulated MR-scans with and without noise for this software brain phantom. The noisy T1-weighted MR-image, used in this study, is shown in the figure (left). All images were sampled with $1 \times 1 \times 1$ mm^3 voxels. A PET activity distribution was created by first assigning a single label to each voxel, including labels for combinations of GM and WM, and of GM and CSF, and labels for various mixtures of any combination of tissues. Because only discrete levels of activity can be assigned to an input image of PET-SORTEO, a high number of labels was used to retain the correspondence with the tissue classes and the MR-image. Next, taking into account known relative intensities in metabolic or perfusion brain studies using indirect markers such as ^{18}F-FDG or $H_2{}^{15}O$, each tissue class was assigned a clinically realistic activity value: 12500 Bq/ml in GM, 3125 Bq/ml in WM, 0 Bq/ml in air, CSF and bone, and 1000 Bq/ml in all other tissues (i.e. skin, muscle, connective tissue and fat). The resulting PET phantom is shown in Fig. 2 (top). To build the attenuation map, the tissues were assigned the same attenuation coefficients as in PET-SORTEO [12]. The activity and attenuation images were sampled with voxels of $1 \times 1 \times 1$ mm^3 for the generation of accurate sinograms. The images were reconstructed with larger voxels ($2.25 \times 2.25 \times 2.425$ mm^3).

The priors have been evaluated for the normal brain, and for a brain with hypointense gray matter lesions which were

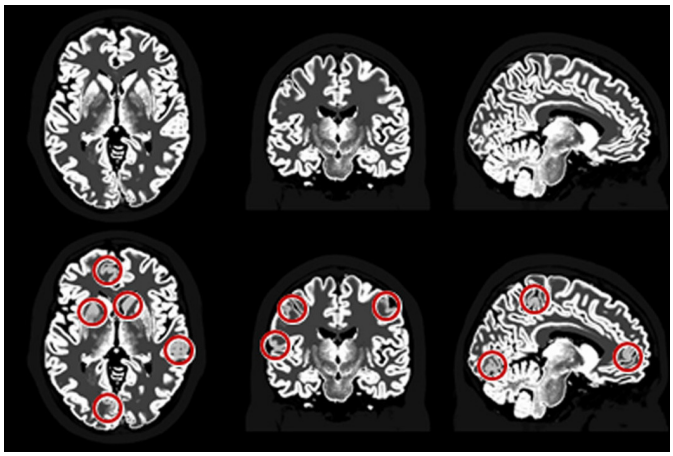

Fig. 2. The simulated brain phantom without and with lesions ($1 \times 1 \times 1$ mm^3 voxels). The lesions are modeled as a 25% reduction of the tracer uptake in the gray matter, which is enclosed in a 20 mm diameter sphere. The lesions visible in the displayed slices were encircled.

invisible to the MR. Twenty lesions were created by distributing 20 spheres with a diameter of 20 mm over the brain, encompassing cortical and subcortical cerebral regions (see Fig. 2). The tracer uptake of the gray matter inside each sphere was reduced by 25%, while the activity of the other classes was unaffected. Consequently, the actual volume of the lesion was different in every lesion (ranging from 1.51 to 3.11 cm^3), and much smaller than the volume of the sphere. This represents a reasonably small to intermediary size for clinical focal lesions e.g. in multifocal vascular disease or cortical dysplastic lesions in epilepsy [25], [26].

B. Anatomical Priors

The *A-prior* was described previously [10]. It assumes that segmentation images of GM, WM and CSF are available. The segmentation images contain values between 0 and 1, denoting the fraction of the voxel that is occupied by that particular tissue. In CSF, a Gaussian intensity prior is applied that favors a reconstructed value of zero. Also in WM, a Gaussian intensity prior is applied. However, for WM, the mean WM intensity is not fixed in advance, but estimated during MAP reconstruction. In GM, a smoothing Markov prior is applied; we used the relative difference prior with $\gamma = 0$ [27], which means that there is no increased tolerance for large edges. γ could be tuned to enhance lesion boundaries, but in practice its ideal value depends heavily on the decreased uptake in the lesions, which is unpredictable for patients, and on the amount of noise. In addition, recent studies indicate that edge tolerance does not improve lesion detection [29], [30]. Therefore, γ was fixed to zero. For voxels containing a mixture of at least 10% GM and either WM or CSF, it is assumed that the WM or CSF concentration is the same as in pure WM or CSF voxels. Consequently, for those voxels, the MAP algorithm only needs to determine the GM value. A similar approach is followed for mixtures with mainly WM, CSF or other tissues. In addition, a smoothing Markov prior is applied to all other tissues. This A-prior was designed for reconstruction of images with normal WM and possible GM lesions. Because the A-prior has a global maximum, the posterior can be optimized with simple gradient ascent algorithms.

The *JE-prior* is obtained by deleting the marginal entropy terms in mutual information, and was found to yield less bias in the reconstructed emission images [2]. In contrast to the A-prior, the joint entropy prior does not need a segmentation, it operates directly on the anatomical image. Because it is computed from the joint histogram, it ignores correlations between neighboring voxels. On the other hand, it has the potential to collect information over large distances, which might be a strength if large organs with uniform tracer uptake are present. It favors piecewise constant images. A problem is that this prior has local maxima: the final reconstruction depends on the initial image and on the optimization algorithm.

The *Bowsher-prior* was originally proposed in [4]. It is a smoothing Markov prior operating on a position dependent neighborhood. It has the following essential parameters:

1) the definition of a neighborhood $\mathbf{N}$,
2) a similarity metric $S(\alpha_j, \alpha_k)$ for the anatomical image,
3) the number of neighbors $B < N$ that must be selected for each voxel j.

Let $\boldsymbol{\lambda}$ be the PET image with voxel values $\lambda_j, j = 1 \ldots J$, and $\boldsymbol{\alpha}$ the registered anatomical image with voxel values $\alpha_j, j = 1 \ldots J$. To compute the prior for λ_j, the neighborhood $\mathbf{N}_j$ of voxel j is scanned, computing the similarity metric $S(\alpha_j, \alpha_k), k \in \mathbf{N}_j$. The B voxels with the highest similarity are selected, the set containing these selected voxels is denoted $\mathbf{B}_j$. The smoothing Markov prior in the PET image is then restricted to $\lambda_k, k \in \mathbf{B}_j$. Like Bowsher *et al.*, we used the absolute intensity difference to compute the similarity and considered neighborhoods of different sizes. Working with a position dependent neighborhood implies that a selected neighbor k of voxel j does not necessarily have voxel j as one of its selected neighbors. With Bowshers rationale in mind, one could argue that this voxel k should influence voxel j, but not vice versa. However, this wish will not be fulfilled in case of a regular implementation of MAP with the Bowsher-prior. Therefore, we recently suggested a modification of the Bowsher-prior [3]. More details on this new prior can be found in the Appendix. In the current study, we used the modified version, because it outperformed the original one in terms of bias-noise characteristics [3]. As for the A-prior, we chose the relative difference prior with $\gamma = 0$ as the smoothing Markov prior [27]. The Bowsher-prior does not rely on a segmentation of the anatomical image and no convergence problems were encountered in practice.

To compute the MAP image, the heuristic but effective gradient ascent algorithm proposed in [27] was used. To accelerate convergence, a scheme with gradually decreasing number of subsets was applied. For the JE-prior, we attempted to avoid convergence to an undesired local maximum by either gradually increasing the weight of the prior, or starting from a fast, low-noise Bowsher reconstruction. The actually used reconstruction parameters can be found in Section II-E.

C. Parameter Tuning

Each of the priors has several parameters, which should be well tuned for a useful performance comparison. For the

tuning, a simple and fast analytical 3D PET simulator was used in order to exclude bias introduced by inconsistencies between the (simulated) measurement and the model prediction used during reconstruction, e.g., spatially variant versus spatially invariant resolution model. The simulated scanner dimensions and parameters were chosen similar to those of the Siemens Ecat HR+ scanner. Its radial sampling distance and axial plane separation are 2.25 mm and 2.425 mm, respectively. This analytical simulator uses a ray tracing projector, which models the resolution by a 3D Gaussian convolution with a full width at half maximum (FWHM) of 4.0 mm in the transaxial planes and a FWHM of 5.45 mm in the axial direction. The choice of these values is explained in the next section. It also models attenuation, but ignores scatters and randoms. Because the reconstructions used the same projector/backprojector, there was an exact match of the resolution between simulation and reconstruction.

The different reconstruction algorithms were studied for two different noise levels, corresponding to a 5-min (relatively low noise) and a 1-min (high-noise) FDG PET-measurement. For each noise level, two noisy PET acquisitions were simulated, one for the normal brain and one for the brain with the 20 lesions. All sinograms were reconstructed with the different algorithms and with different parameters for each prior. To evaluate the results, a figure of merit $\mathrm{FOM}_{\mathrm{GM,N}}$ was computed for the reconstructions of the normal brain as

$$\mathrm{FOM}_{\mathrm{GM,N}} = \sqrt{\frac{1}{n_{\mathrm{GM80}}} \sum_{j \in \mathrm{GM80}} \left(\frac{\lambda_j^{\mathrm{N}} - \lambda_j^{\mathrm{N,true}}}{\mathrm{GM\ activity}} \right)^2} \qquad (1)$$

where GM80 represents all voxels that belong for more than 80% to the gray matter, n_{GM80} is the number of elements in GM80, λ_j^{N} is the reconstructed activity in voxel j for the normal brain and $\lambda_j^{\mathrm{N,true}}$ is the corresponding true activity, "GM activity" is the true activity in GM. The reconstructed and true activity distribution were compared in the PET reconstruction space, i.e. after resampling the true activity image from $1 \times 1 \times 1$ mm^3 voxels to $2.25 \times 2.25 \times 2.425$ mm^3 voxels. Because the GM has a small thickness and the PET voxels are relatively large, only a small fraction of the PET voxels that intersect with the GM will completely fall inside the GM. To increase the number of voxels for our analysis without introducing a large influence of possible bias on the reconstructed WM and CSF, we included all PET voxels that are for at least 80% inside the GM in the analysis. A similar score $\mathrm{FOM}_{\mathrm{GM,L}}$ is computed for the reconstructions of the brain with lesions. However, $\mathrm{FOM}_{\mathrm{GM,N}}$ and $\mathrm{FOM}_{\mathrm{GM,L}}$ are expected to produce very similar values, because the number of lesion voxels is much smaller than the number of normal GM voxels. The FOM of equation (1) was selected, because we want to ensure good quantitative accuracy (with low bias and low noise) in all regions of the GM. This should enable picking up small differences between different regions in the brain, which is important when studying, e.g., Alzheimer's disease [21].

To assess the quality of the 20 reconstructed lesions, also a signal $\mathrm{FOM}_{\mathrm{S}}$ is computed:

$$\mathrm{FOM}_{\mathrm{S}} = \sqrt{\frac{1}{20} \sum_{l=1}^{20} \left(\frac{(\bar{\lambda}_l^{\mathrm{L}} - \bar{\lambda}_l^{\mathrm{N}}) - (\bar{\lambda}_l^{\mathrm{L,true}} - \bar{\lambda}_l^{\mathrm{N,true}})}{\mathrm{GM\ activity}} \right)^2}, \qquad (2)$$

where $\bar{\lambda}_l^{\mathrm{L}}$ represents the mean value of lesion l in the reconstruction of the brain with lesions, and $\bar{\lambda}_l^{\mathrm{N}}$ is the corresponding mean value for the reconstruction of the normal brain. The superscript 'true' denotes the corresponding true values. Again, the analysis was performed in the PET reconstruction space, hence requiring downsampling the binary image defining the lesion voxels. In order to have sufficient voxels for our analysis inside each lesion, all PET voxels that are for at least 80% inside the lesion and consist of at least 80% of GM were included. Accurate lesion quantification is often not required for diagnostic purposes, but is essential for region-of-interest-based inter-patient and inter-population comparisons [28].

These scores were combined in the final figure of merit that was used for optimizing the parameters:

$$\mathrm{FOM} = \sqrt{\frac{\mathrm{FOM}_{\mathrm{GM,N}}^2 + \mathrm{FOM}_{\mathrm{GM,L}}^2}{4} + \frac{\mathrm{FOM}_{\mathrm{S}}^2}{2}} \qquad (3)$$

The choice of assigning an equal weight to the figures of merit for GM and signal was made rather arbitrarily, but usually resulted in favoring images with both good lesion signal and low bias and noise in the GM.

After selection of all the parameters using (3), the algorithms were finally evaluated with realistic simulations.

D. Realistic Simulations

These 5- and 1-min brain PET-scans were also simulated with the Monte Carlo simulator PET-SORTEO, which takes into account all significant physical effects of the PET acquisition [12], [13]. For this study, a realistic model of the system spatial resolution is required, because the mismatch between the true system resolution and the one used during reconstruction is expected to have a direct impact on the accuracy of the reconstructions. Point source measurements were performed in the Siemens Ecat HR+ scanner of the K.U. Leuven, and simulated with PET-SORTEO. The point source was positioned at different distances from the center of the field of view. Fig. 3 shows the result at 15 cm from the center. For eccentric point sources, the sinogram resolution varies with the projection angle. In the graph, zero degrees corresponds to projections along vertical lines. In the central part of the field of view used for brain imaging, the average transaxial resolution was about 4.0 mm FWHM, and the axial resolution was on the average 5.45 mm FWHM. Our experiments confirmed that the mismatch between the true system resolution and the shift-invariant Gaussian model of the (back)projector is similar for a true PET-system and for PET-SORTEO. Therefore, we expect our algorithm evaluation study to have a predictive value for real measurements.

Twelve 30-min ^{18}F-FDG PET-scans of the two previously described brain phantoms (see Fig. 2) were simulated. At the start of each scan, the initial activity distribution assigned to the phantoms was reset. The counts in ten of the resulting

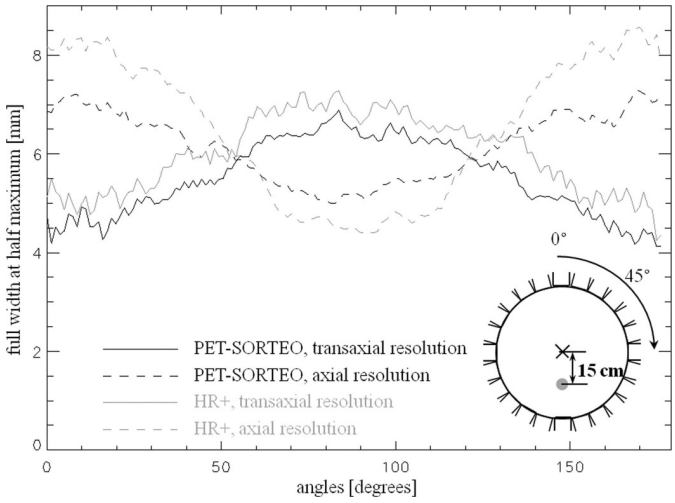

Fig. 3. Sinogram resolution determined with a point source in a real Siemens PET HR+ scanner (gray lines) and in the one simulated by PET-SORTEO (black lines). The point source was positioned at 150 mm from the center of the field of view, as shown in the drawing at the bottom right. This is near the edge of the field of view, where the resolution effects are most dramatic.

sinograms were randomly distributed over 10×6 sinograms. Those of the two remaining 30-min sinograms were randomly distributed over 2×30 sinograms. Because the simulated sinograms are subject to Poisson noise (ensured by PET-SORTEO), the derived sinograms are 60 independent Poisson realizations of a 5-min and a 1-min brain scan, respectively. All newly generated sinograms are subject to the same mean decay of ^{18}F over 30 min. Therefore, the final reconstruction images are corrected for this mean decay to allow comparison with the true activity distribution during image analysis. The simulation of a few long acquisitions rather than of all individual short ones enabled calculation within a realistic time span, because each PET-scan simulation involves running a time-consuming part of the SORTEO code which is independent of the number of disintegrations to be simulated.

In order not to introduce errors due to inaccurate scatter correction, realistic (noisy) scatter sinograms were created and exact scatter models were used during reconstruction. Because PET-SORTEO cannot provide a separate sinogram of the simulated scatter, its scatter modeling was turned off. In addition to the twelve simulated 30-min scans, one 30-min scan was simulated using PET-SORTEO with scatter modeling and combined with a scatter-free simulation to estimate the scatter. The smoothed scatter estimate (used as scatter model) was corrupted with Poisson noise, and added to the scatter-free sinograms. Next, the counts were distributed over the sinograms of the different noise realizations. The exact, rescaled scatter model could then be used for scatter correction during reconstruction. Hence, no additional bias due to inaccurate scatter modeling was introduced. The same attenuation map was used as for the analytical simulations.

E. Reconstruction Implementation Details

In this section, details are provided about the choice of the reference reconstruction algorithms (without use of anatomical information), and about the general and more specific reconstruction parameters selected. Besides the description of the reconstruction voxel size, the iteration scheme and the resolution model, we also explain which parameters were tuned and from which starting image the reconstruction was initiated.

In clinical routine, anatomical information is typically used post-reconstruction by inspection of fused (registered) images, e.g. PET and MRI. Hence, postsmoothed maximum-likelihood expectation-maximization (MLEM) images [32] were computed to enable comparison with the current image quality. The images were reconstructed with a transaxial pixel size of 2.25 mm and a plane separation of 2.425 mm, which correspond to the dimensions of the detectors. To accelerate convergence, ordered subsets (OS) [33] are used with a gradually reduced OS iteration scheme: 6×36, 2×16, 2×1 (global iterations $\times$ number of subsets). During reconstruction, the system resolution was modeled as a 3D Gaussian blurring with a FWHM of 4.0 mm transaxially and 5.45 mm axially. A 3D isotropic Gaussian smoothing filter with a full width at half maximum (FWHM) varying between 2 and 10 mm, in steps of 1 mm, was applied after reconstruction.

This same iteration scheme and resolution model was used for all different reconstruction algorithms. The total amount of iterations, including those required to compute a starting image, was the same for all algorithms.

As MLEM-reconstruction with Gaussian postsmoothing yields a somewhat different point spread function than regularization during MAP-reconstruction, comparison between MAP-algorithms that do and do not include anatomical information might seem more fair than a comparison with postsmoothed MLEM. Therefore, the projection data were also reconstructed using a MAP-algorithm with the relative difference (RD) prior, started from a fast MLEM reconstruction image (1 iteration over 36 subsets). The weight of the prior was tuned based on the FOM of equation (3) and γ was fixed to zero.

A-MAP was started from the same fast MLEM reconstruction image as MAP with the RD-prior. The A-MAP parameters were tuned for reconstruction using the exact GM, WM and CSF segmentation of the MR-image. The parameters to be tuned were the weights of the priors applied to GM, WM, CSF and other tissues. These weights were gradually increased to speed-up convergence.

In order to avoid convergence to an undesired local maximum, the JE-reconstruction was initiated with a homogeneous starting image, and the weight of the JE-prior was gradually increased at every full iteration over 36 subsets. As an alternative, the JE-reconstruction was also started from a fast (1×36) Bowsher reconstruction, and an appropriate constant JE-weight was selected. In addition, the combined use of the JE- and Bowsher-prior was investigated, starting from the same fast Bowsher-reconstruction. The influence on the GM activity and the lesion signal of the starting image and of different values for the JE-weight, the number of bins for the PET- and for the MR-image (might be different) and the width of the Parzen window to be applied to the PET intensities (see [2]) was evaluated using equation (3).

The Bowsher-reconstruction was started from the same, fast

MLEM reconstruction image as A-MAP. Tuning of this algorithm was limited to varying the weight of the prior, the size of the neighborhood and the number of selected neighbors. Two approximately spherical neighborhoods were compared, namely the 18 closest neighbors in a 3×3×3 neighborhood, and the 80 closest neighbors in a 5×5×5 neighborhood. All neighbors were assigned the same weight.

As anatomical information, A-MAP needs an accurate segmentation of the GM, WM and CSF. Either the exact segmentation was used, or a more realistic segmentation was computed from the corresponding simulated MR-image from the BrainWeb database using SPM8 [31]. All remaining (partial and complete) voxels were classified as 'other'. The parameters were only tuned for the case where exact segmentation data was used.

The parameter tuning was also limited to reconstructions based on the noise-free MR-image. In the final evaluation, a more realistic, noisy MR-image (9% noise calculated relative to the brightest tissue) was used during the different MAP-reconstructions.

F. Comparison Quantitative Accuracy

To decide which algorithms perform best in terms of quantitative accuracy, the reconstructions of the SORTEO-simulated PET data were evaluated using the FOM from equation (3). To yield a single value for each algorithm, enabling ranking, the average was taken over the 60 FOMs. With Bonferroni-corrected sign-tests (using Statistica (StatSoft), $p < 0.05/45$), it was verified whether or not the differences in FOM obtained with different reconstruction algorithms are statistically significant. Furthermore, the effect of the smoothing parameter on the FOM was checked by calculating the FOM for two additional sets of reconstructions, i.e. for MLEM with 0.5 and 1.5 times the optimal FWHM for the postsmoothing filter, and for the MAP-algorithms with 0.1 and 10 times the optimal prior weight.

G. Bias-Noise Analysis

To further compare the various reconstruction algorithms, a bias-noise analysis was performed for the reconstructed gray matter activity and for the activity and difference in activity (which we call signal) in the lesion VOIs, based on the reconstruction images of the two sets of 60 noisy PET-SORTEO sinograms. Bias-noise curves were generated by varying the FWHM of the postsmooth filter of MLEM ($0.5\text{FWHM}_{optimized}$/$\text{FWHM}_{optimized}$/$1.5\text{FWHM}_{optimized}$) and by varying β for all MAP-reconstructions ($0.1\beta_{optimized}$/$\beta_{optimized}$/$10\beta_{optimized}$).

To study the image quality in the GM voxels, the percentage (%) mean absolute bias was calculated, by first averaging the bias in every GM80 voxel over all noisy reconstructions, and next taking the mean over the absolute values of all GM80 voxels:

$$\%\text{mean}|\text{bias}_{\text{GM,N}}| = \frac{\frac{100}{n_{\text{GM80}}} \sum_{j\in\text{GM80}} \left| \frac{\sum_{s=1}^{60}\left(\lambda_j^{\text{N},(s)} - \lambda_j^{\text{N,true}}\right)}{60} \right|}{\text{GMactivity}} \tag{4}$$

with $\lambda_j^{\text{N},(s)}$ the activity in voxel j in the reconstruction image of the s-th noisy sinogram of the normal brain. The % noise was calculated as:

$$\%\text{noise}_{\text{GM,N}} = \frac{100\sqrt{\frac{\sum_{j\in\text{GM80}}\sum_{s=1}^{60}\left(\lambda_j^{\text{N},(s)} - \frac{\sum_{s=1}^{60}\lambda_j^{\text{N},(s)}}{60}\right)^2}{59\cdot n_{\text{GM80}}}}}{\text{GMactivity}} \tag{5}$$

Similar expressions were used for the brain with lesions.

The % bias and % noise on the activity in each lesion VOI l are defined as:

$$\%\text{bias}_{l,\text{N}} = \frac{100(\frac{1}{60}\sum_{s=1}^{60}\bar{\lambda}_l^{\text{N},(s)} - \bar{\lambda}_l^{\text{N,true}})}{\bar{\lambda}_l^{\text{N,true}}} \tag{6}$$

$$\%\text{noise}_{l,\text{N}} = \frac{100\sqrt{\frac{1}{59}\sum_{s=1}^{60}\left(\bar{\lambda}_l^{\text{N},(s)} - \frac{1}{60}\sum_{s=1}^{60}\bar{\lambda}_l^{\text{N},(s)}\right)^2}}{\bar{\lambda}_l^{\text{N,true}}} \tag{7}$$

for the normal brain reconstructions, and in a similar way for reconstructions of the brain with lesions.

In addition, the signal-to-noise ratio was calculated in each lesion VOI as

$$\text{SNR}_l = \frac{\frac{1}{60}\sum_{s=1}^{60}(\bar{\lambda}_l^{\text{L},(s)} - \bar{\lambda}_l^{\text{N},(s)})}{\sqrt{\left(\text{stddev}([\bar{\lambda}_l^{\text{L},(s)}]_{s=1}^{60})\right)^2 + \left(\text{stddev}([\bar{\lambda}_l^{\text{N},(s)}]_{s=1}^{60})\right)^2}} \tag{8}$$

with $\text{stddev}([X]_{s=1}^{60})$ denoting the standard deviation of variable $X = \{X_1, \ldots, X_{60}\}$. To study whether the differences in SNR in the lesions were statistically significant (Bonferroni-corrected $p < 0.001111$), sign-tests were performed between the results of every two algorithms (with optimized reconstruction parameters) using Statistica (StatSoft). The low- and high-noise data were combined to increase the number of data points from 20 to 40.

III. Results

A. Reconstruction Images

Based on the FOM of equation (3), reconstruction parameters have been tuned to yield optimal quantification accuracy. Example reconstruction images of a PET-SORTEO simulated 5-min and 1-min scan of the brain with hypointense lesions are displayed in Fig. 4 and 5, respectively. These were obtained with the above described reconstruction algorithms using the optimal parameter settings. The locations of the lesions are indicated in the true activity distribution images Fig. 4(a) and 5(a) and are identical to those displayed in Fig. 2 (bottom), because the same slices were selected for visualization. All images were displayed using PET reconstruction voxels.

The optimal isotropic Gaussian postsmooth filter for MLEM was found to have a FWHM of 4 mm for the low-noise and 6 mm for the high-noise data. Fig. 4(b) and (c) show three orthogonal slices through the unprocessed and postsmoothed MLEM reconstruction image of a PET-SORTEO simulated 5-min FDG PET-scan of the brain phantom with 20 lesions. The former image is rather noisy, whereas the latter is fairly smooth. This difference can be appreciated much better in

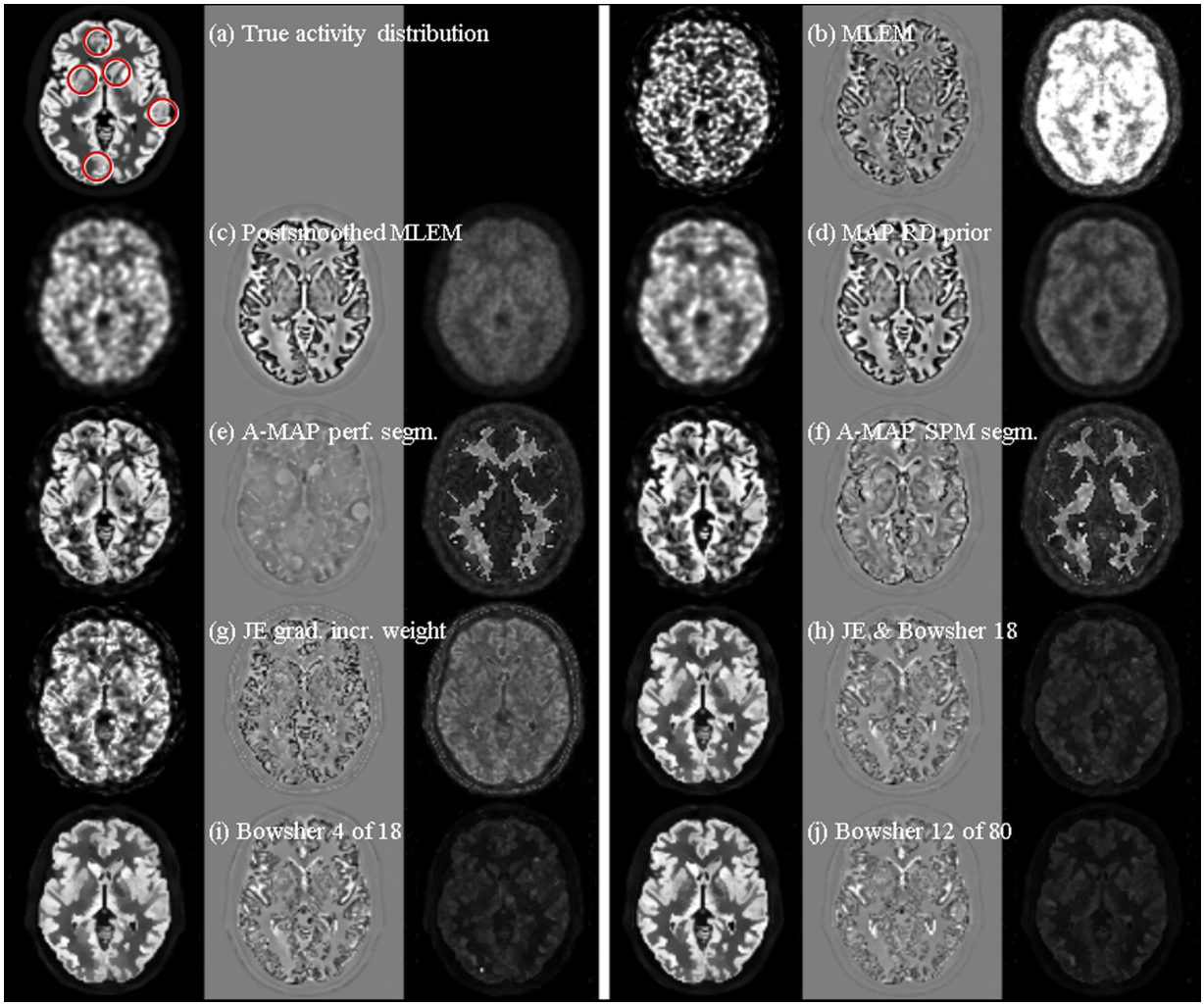

Fig. 5. True activity distribution (a) and example reconstruction images (first columns) and the corresponding bias (second columns) and standard deviation images (third columns) of a PET-SORTEO simulated 1-min FDG PET-scan of a brain with hypointense lesions using the simulated noisy MRI. The following reconstruction algorithms were used: (b) MLEM, (c) postsmoothed MLEM (6 mm FWHM), (d) MAP with a quadratic prior, (e) A-MAP using perfectly segmented MRI information, (f) A-MAP using SPM segmented MRI information, (g) JE with gradually increased weight, (h) JE + Bowsher using 4 out of 18 neighbors, (i) Bowsher using 4 out of 18 neighbors, and (j) Bowsher using 12 out of 80 neighbors. Both the bias images and the standard deviation images were multiplied by 3 for visualization reasons. Next, the bias images were clipped at -GM activity and GM activity and the standard deviation images were clipped at GM activity (as done for the reconstruction images).

Fig. 5(b) and (c), which represent the corresponding reconstruction images of a 1-min FDG PET-scan. The MAP-reconstructions with the RD-prior (Fig. 4(d) with smoothing parameter $\beta = 1$ and 5(d) with $\beta = 3$) look very similar to those obtained with postsmoothed MLEM.

As expected, visually appealing images were obtained with A-MAP-reconstruction using the perfect MRI segmentation, even in the presence of a high amount of noise (Fig. 4(e) and 5(e)). The different tissue classes can clearly be discriminated, because the activity was encouraged to be spread within the tissue classes only. To yield this image, the prior weights for GM, WM, CSF and other were gradually increased to 200/450 (for low/high noise level), 30/25, 125/200 and 30/30. A $3\times3\times3$ Markov neighborhood with 26 neighbors with equal weight was used. Increasing the weight for WM would make its activity distribution more smooth at the cost of the image quality of the GM. Using the SPM segmentation instead of the exact one obviously degrades the image quality, as the activity boundaries do no longer coincide exactly with the anatomical ones (Fig. 4(f) and 5(f)).

Using the JE-prior also enhances the visual contrast, detail and quantification of the GM activity compared to postsmoothed MLEM. When tuned for good performance in the GM, the JE-prior tends to produce a relatively noisy reconstruction of the WM. This is clearly visible in Fig. 4(g) and 5(g), where the image was started from a homogeneous activity distribution and the JE-weight was linearly increased to 1.0×10^8 and 6.2×10^8 for the 5- and 1-min data, respectively. The JE-histogram was computed using 6 bins for the PET image, 25 bins for the MRI image and a Gaussian Parzen window with a standard deviation of 0.9 pixels.

We also reconstructed the data with the JE-prior starting from a fast, optimized Bowsher reconstruction image (1 iteration of 36 subsets, using 4 out of 18 neighbors and a Markov weight of 225/275 for low/high noise level). Either only the JE-prior (weight $8.0 \times 10^7/6.4 \times 10^8$) was used

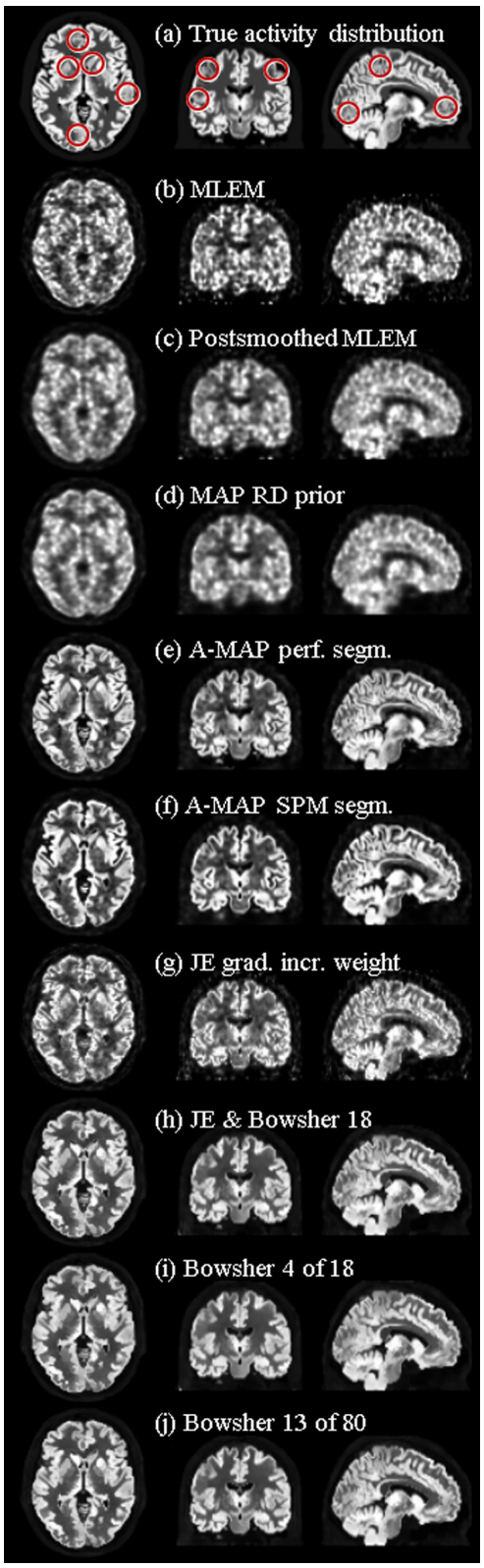

Fig. 4. True activity distribution (a) and example reconstruction images (b)-(j) of a PET-SORTEO simulated 5-min FDG PET-scan of a brain with hypointense lesions using the simulated noisy MRI. The following reconstruction algorithms were used: (b) MLEM, (c) postsmoothed MLEM (4 mm FWHM), (d) MAP with a quadratic prior, (e) A-MAP using perfectly segmented MRI information, (f) A-MAP using SPM segmented MRI information, (g) joint entropy (JE) with gradually increased weight, (h) JE + Bowsher using 4 out of 18 neighbors, (i) Bowsher using 4 out of 18 neighbors, and (j) Bowsher using 13 out of 80 neighbors.

from the initial image onwards, or both the JE- (weight 4.0 $\times$ 10^7/3.4 $\times$ 10^8) and the Bowsher-prior (weight 45/110) were used simultaneously to combine their global and local properties (Fig. 4(h) and 5(h)). The former look very similar to the original JE-reconstructions (not shown); the latter are less noisy and very alike the Bowsher-only reconstructions (Fig. 4(i) and 5(i)).

The simulated data were finally reconstructed with a small and large Bowsher neighborhood. For the former, the 4 most similar (in the MRI) of the 18 neighbors were selected and a Markov weight of 100/225 (for low/high noise) was assigned (Fig. 4(i) and 5(i)). For the latter, 13/12 of the 80 neighbors were chosen in combination with a Markov weight of 10/25 (Fig. 4(j) and 5(j)). In both cases, the images show a high level of detail in the GM and a fairly homogeneous activity distribution in WM and CSF. Using a larger neighborhood slightly improves the resolution in the GM. Most lesions in the GM can be discerned from both images.

In Fig. 5, the transaxial slices of the reconstructed 1-min scans (left columns) were displayed next to the corresponding bias (central columns) and standard deviation images (right columns). The latter two were multiplied by three for visualization purposes. The most obvious bias is visible at the edges of the GM in the postsmoothed MLEM and the MAP with RD-prior image. All anatomical priors successfully reduced this bias with the A-prior using perfect segmentation information as the clear winner and the Bowsher-prior using a large neighborhood as runner-up. The highest noise level was found in the unprocessed MLEM image. Postsmoothing as well as using the RD-prior resulted in a less intense, rather smooth standard deviation image. The noise in the A-MAP images was most concentrated in the WM voxels (recall the low WM weight). In the JE-only images the noise is again spread over the entire brain (somewhat lower in WM and CSF voxels, especially when started from a Bowsher reconstruction (not shown)) and of the same level as in the postsmoothed MLEM or MAP with RD-prior image. In the joint JE-Bowsher and in the Bowsher-images the noise is low in the entire image and most concentrated in the GM voxels.

B. Comparison Quantitative Accuracy

In Fig. 6 the FOM, averaged over the 60 noise realizations, is plotted against the normalized regularization parameter. The size of the symbols was increased with the magnitude of the FWHM (2/4/6 mm for low noise and 3/6/9 mm for high noise) or the β-value ($0.1\beta_{optimized}$/$\beta_{optimized}$/$10\beta_{optimized}$), hence the central points indicate the results obtained with the parameters yielded from the parameter tuning.

The curves in Fig. 6 confirm that the parameters obtained in the parameter tuning stage are still yielding minimal FOM when used to reconstruct more realistic data sets. The same ranking was found for the algorithms with optimized reconstruction parameters in case of low- and high-noise data. From best to worst (quantified by the FOM of eq. (3)): A-MAP with perfect segmentation, Bowsher with large neighborhood (nbh.), combined use of the JE- and Bowsher-prior (small nbh.), Bowsher with small nbh., A-MAP with SPM segmen-

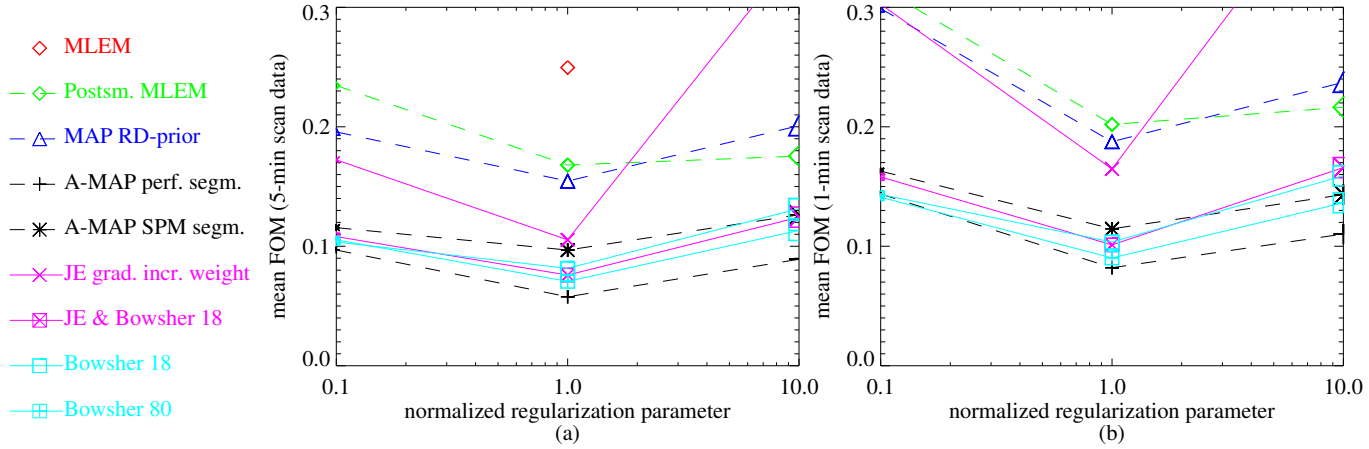

Fig. 6. Plot of the FOMs for (a) the 5-min scan data and (b) the 1-min scan data as a function of the normalized regularization parameter (1 = optimal value). For postsmoothed MLEM the optimized FWHM was multiplied with 0.5, 1 and 1.5 (x-axis used with linear scale in this case). For all MAP-algorithms the optimized β was multiplied by 0.1, 1 and 10 (x-axis with logarithmic scale).

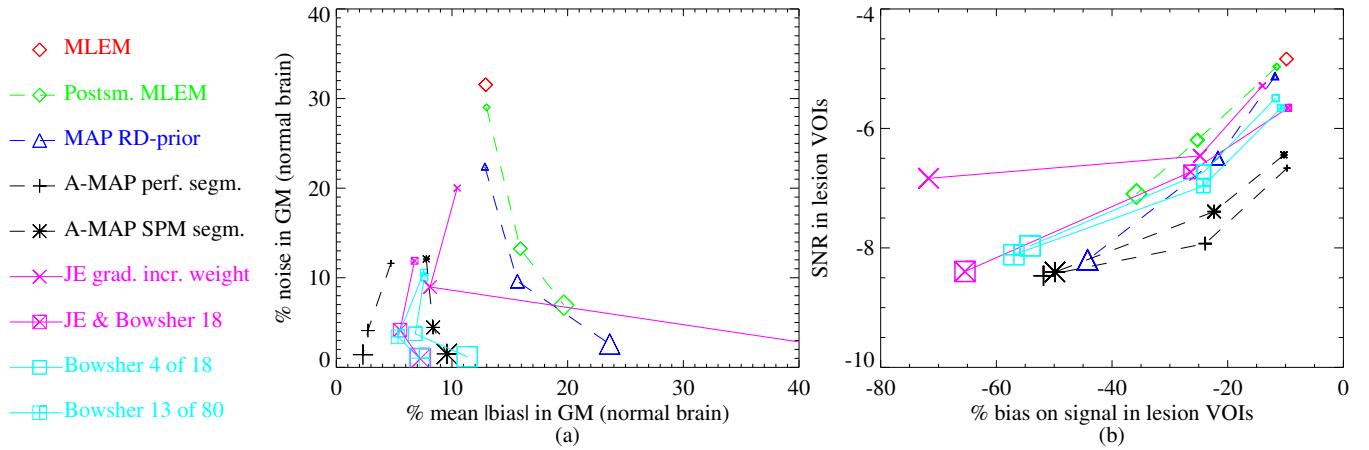

Fig. 7. Bias-noise analysis in the reconstruction images of the 5-min scan simulated data obtained with the different evaluated reconstruction algorithms. In (a) the percentage noise in the GM voxels is plotted with respect to the percentage mean |bias| in these voxels. In (b) the signal-to-noise ratio in the lesion VOIs is plotted with respect to the percentage bias on the signal in the lesion VOIs. The signal was defined as the numerator of equation (8).

tation, the two JE-only reconstruction sets, MAP with RD-prior, postsmoothed MLEM, and finally unprocessed MLEM. Only JE started from a fast Bowsher-reconstruction switched places with A-MAP with SPM segmentation when the low-noise data set was considered rather than the high-noise data set. From the statistical tests, we found that all differences in FOM were significant, except for the differences between the JE-reconstructions using a gradually increased weight and the JE-reconstructions starting from a fast Bowsher-reconstruction when the 1-min data were considered.

C. Bias-Noise Analysis GM

The results of the bias-noise analysis of the activity in the GM voxels are shown in Fig. 7(a) and 8(a) for the various reconstructions of the 5-min and 1-min measurements, respectively. For each algorithm a bias-noise curve was plotted. As in Fig. 6, the size of the symbols was related to the amount of regularization, hence the central points were obtained with the optimized parameters. The main conclusions from this bias-noise analysis are the same for low and high noise.

The unprocessed MLEM-image suffered from a relatively high mean absolute bias and high noise level in the GM voxels (◇). Postsmoothing the image slowly increased the bias while significantly reducing the noise (– ◇ –). The MAP-algorithm with the RD-prior (– △ –) slightly outperformed postsmoothed MLEM.

The optimized A-MAP-reconstruction based on perfectly segmented MRI data (– + –) has much lower mean |bias| and lower noise compared to the optimized postsmoothed MLEM. Using imperfect (SPM) segmentation data instead, greatly increased the mean |bias|, but had similar noise reduction properties (– ∗ –). The JE-prior with gradually increased weight (–×–) also yielded a reduction in mean |bias|, but was less capable of suppressing the noise than the A-priors. Nevertheless, in comparison with postsmoothed MLEM, it still reduced the mean |bias| by about 35-50%, while reaching about the same or a lower noise level. Starting from a fast Bowsher-reconstruction (using a small neighborhood) introduced a comparable image improvement (results not shown in order not to overload the graphs). Combining the JE- and

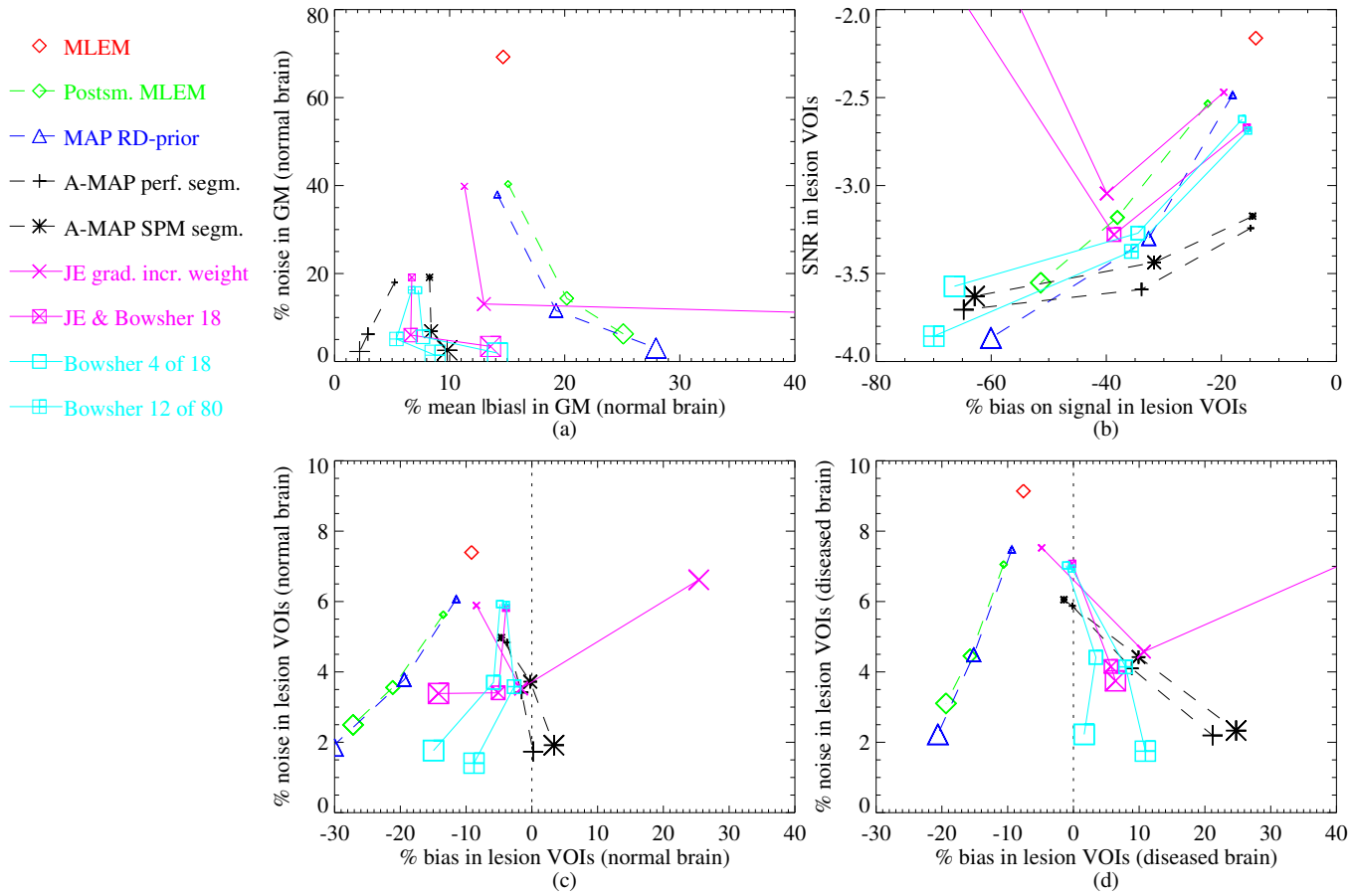

Fig. 8. Bias-noise analysis in the reconstruction images of the 1-min scan simulated data obtained with the different evaluated reconstruction algorithms. In (a) the percentage noise in the GM voxels is plotted with respect to the percentage mean |bias| in these voxels. In (b) the signal-to-noise ratio in the lesion VOIs is plotted with respect to the percentage bias on the signal in the lesion VOIs. In (c) and (d) the percentage noise in the lesion VOIs in the normal brain (c) and in the brain with lesions (d) is plotted with respect to the percentage bias in these VOIs.

the Bowsher-prior with small neighborhood (–⊠–) yielded somewhat lower bias than and the same noise as A-MAP with SPM segmentation information (– ∗ –), and hardly outperformed the Bowsher-only reconstruction (–□–). Both optimized Bowsher reconstructions (–□– & –⊞–) reached bias-noise trade-off points in between those of the two A-MAPs. Increasing the Bowsher-neighborhood only resulted in a minor bias reduction (–⊞–).

Fig. 7(a) and 8(a) show the results for the normal brain. Those for the brain with lesions are almost identical, and were therefore omitted.

D. Analysis Lesion SNR

Fig. 7(b) and 8(b) show the SNR in the lesion VOIs (see eq. (8)) versus the % bias on the signal in these VOIs for reconstructions of low- and high-noise data, respectively. The symbols depict the mean value averaged over the 20 lesions.

MLEM (◇) yields the lowest bias on the signal. The bias induced by each of the (optimized) anatomical priors is similar to that obtained with postsmoothed MLEM (– ◇ –) or MAP-RD (– △ –) with an optimized regularization parameter, and does not vary a lot between the algorithms. As expected, the smoothing effect of an increased FWHM of the postsmooth filter or of an increased β-value decreases the (absolute) signal, resulting in more negative bias, while decreasing the noise. As the noise tends to drop faster than the bias increases, this results in an improvement in SNR (at least initially).

In table I, the outcome of the statistical sign test is provided with in the lower triangle the p-values, and in the upper triangle the percentage lesions having a better SNR when reconstructed with algorithm B instead of A. Bold figures correspond to significant differences. All evaluated algorithms yielded a significantly better SNR in the lesions compared to unprocessed MLEM (◇). In addition, the SNRs in the A-MAP-images (– + – & – ∗ –), in the Bowsher-reconstructions using a large nbh. (–⊞–), and in the joint JE-Bowsher-reconstructions (–⊠–) were statistically better than those in the images reconstructed with postsmoothed MLEM (– ◇ –) and MAP with the JE-only priors (–×–). Furthermore, MAP with a regular RD-prior (– △ –) outperformed postsmoothed MLEM (– ◇ –) in terms of SNR, whereas no significant differences in SNR were found when comparing it to the Bowsher- and JE-priors. Finally, A-MAP-reconstruction using perfect segmentation information (– + –) resulted in significantly better SNRs than all other reconstructions.

TABLE I

RESULTS OF THE SIGN-TEST APPLIED TO VERIFY WHETHER THE SNR VALUES OBTAINED WITH DIFFERENT RECONSTRUCTION ALGORITHMS ARE SIGNIFICANTLY DIFFERENT OR NOT. THE LOWER TRIANGLE OF THE TABLE SHOWS THE P-VALUES. SIGNIFICANT DIFFERENCES ($p < 0.0011$) ARE IN BOLD. IN THE UPPER TRIANGLE, THE PERCENTAGE LESIONS WITH A BETTER SNR WHEN USING ALGORITHM B RATHER THAN A IS PROVIDED.

Algorithm B ⟶ / Algorithm A ↓	MLEM	Postsm. MLEM	MAP RD-prior	A-MAP perf. segm.	A-MAP SPM segm.	JE grad. incr. weight	JE start Bowsher 18	JE & Bowsher 18	Bowsher 18	Bowsher 80
MLEM		**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Postsm. MLEM	**0.0000**		**92.5**	**95**	**85**	50	55	**85**	75	**87.5**
MAP RD-prior	**0.0000**	**0.0000**		**92.5**	**82.5**	27.5	25	62.5	62.5	75
A-MAP perf. segm.	**0.0000**	**0.0000**	**0.0000**		**10**	**0**	**2.5**	**7.5**	**12.5**	**12.5**
A-MAP SPM segm.	**0.0000**	**0.0000**	**0.0001**	**0.0000**		**7.5**	**7.5**	**22.5**	27.5	32.5
JE grad. incr. weight	**0.0000**	0.8744	0.0072	**0.0000**	**0.0000**		72.5	**85**	**77.5**	**87.5**
JE start Bowsher 18	**0.0000**	0.6353	0.0027	**0.0000**	**0.0000**	0.0072		**85**	75	**90**
JE & Bowsher 18	**0.0000**	**0.0000**	0.1547	**0.0000**	**0.0009**	**0.0000**	**0.0000**		62.5	**77.5**
Bowsher 18	**0.0000**	0.0027	0.1547	**0.0000**	0.0072	**0.0009**	0.0027	0.1547		75
Bowsher 80	**0.0000**	**0.0000**	0.0027	**0.0000**	0.0398	**0.0000**	**0.0000**	**0.0009**	0.0027	

E. Bias-Noise Analysis Lesion VOIs

In Fig. 8(c) and (d), the % noise in the lesion VOIs is plotted with respect to the % bias for the normal brain and the brain with lesions measured during 1 min. Because the conclusions that can be drawn from the long and the short scan data are again very similar, results of the 5 min scans were omitted. As before, the symbols represent the mean value averaged over the 20 lesions. The dotted vertical lines denote zero bias.

The trends in the noise behavior of the various algorithms are similar to those observed for the signal in the lesion VOIs (not shown). Again, MLEM (◇) suffers from the highest noise level in the lesion VOIs. The optimized MAP-reconstructions all approach the lower noise level of the postsmoothed MLEM-images (– ◇ –) and of the MAP-RD-images (– △ –). The noise behavior is very similar in the lesion VOIs of the normal brain and those of the brain with lesions.

For MLEM, a similar negative bias is observed in the two images, and postsmoothing introduced extra bias, as did regularization with the RD-prior. However, the bias behaves somewhat differently depending of the presence of lesions when using an anatomical prior. They all produce very limited negative bias in the normal brain and small to moderate positive bias in the brain with lesions, indicating that some of the surrounding GM activity is smoothed into the lesion VOIs. This results in more (negative) bias on the signal of the lesions when compared to postsmoothed MLEM and MAP-RD (see Fig. 8(b)). Increasing β typically increases the negative/positive bias on the lesion activity while reducing the noise. The JE-reconstructions, especially the JE-only reconstructions (–×–), are more sensitive to an excessively high β than the A-MAP- and the Bowsher-reconstructions.

IV. DISCUSSION

In these simulations, we have assumed that the registration errors are negligible, which is a reasonable assumption for rigid registration of FDG PET and MRI in brain imaging. The reconstruction uses a shift-invariant resolution model, while the PET-SORTEO simulator produces a realistic shift-variant resolution as illustrated in Fig. 3. A realistic amount of scatter (simulated with PET-SORTEO for a 30-min sinogram) was added to the noisy sinograms, and an accurate scatter model was used during reconstruction. For A-MAP, realistic segmentation errors were obtained by segmenting the most noisy MR-image from the Brainweb database with the SPM software package. For these reasons, we believe that the results of this study are clinically relevant, even though the algorithms were not validated on real patient data. In addition, we studied a realistic, but challenging case of brain imaging where small to intermediate sized lesions with rather low reduction in FDG-uptake, and which were invisible on the MR-image, needed to be reconstructed with a high quantitative accuracy from noisy to very noisy data. Typically, a clinical acquisition lasts 20 minutes to minimize the noise effect. It is to be expected that if the lesions were visible on MRI, the advantage of the use of anatomical information would be even larger. For future work, it could be interesting to correlate the lesion contrast in MRI with the image quality of anatomy-based PET reconstruction, and to study the effect of misregistration.

Tuning of the reconstruction parameters was performed based on the quantitative accuracy of the reconstruction of a single, analytically simulated, noisy measurement. A different noise realization might have yielded slightly different optimal parameters. In addition, the FOM used for tuning, equation (3), balances between high GM quantification accuracy and accurate lesion signal estimation. Depending on what is considered more important, one could include a weighting factor. Alternatively, one could, e.g., impose a certain minimal accuracy for the lesion signal estimation, and determine an optimal parameter setting based on the GM FOM. While analyzing the results one should keep in mind that increasing the smoothing parameters of the reconstruction algorithms often yields an increase in GM accuracy while decreasing the signal accuracy and vice versa. An advantage of determining optimal reconstruction parameters beforehand using an objective FOM, is that no complex bias-noise analysis of a large variety of operating points is required. To show how the bias and noise characteristics of the different reconstruction algorithms vary with the smoothing parameter, two additional operating points were calculated in the final comparison studies.

Because the JE-prior often gets stuck in a local optimum, involves many parameters, and is very dependent on the starting image, parameter optimization is tedious, if not

impossible. Therefore, the number of bins and the width of the Parzen window were only tuned for the JE-reconstruction of the 1-min scan with the homogeneous starting image and gradually increased weight. For the other reconstructions, only an optimal prior weight was determined. Possibly, superior JE-reconstructions could be obtained with more tuning. However, an extensive investigation does not seem useful, since the optimal parameters are very case- and noise-dependent and such thorough parameter tuning is impractical for use in clinical routine.

As illustrated in Fig. 7 and 8, the anatomical priors yield an excellent reconstruction of the gray matter, with a bias-noise performance clearly superior to that of postsmoothed MLEM and regular MAP-reconstruction with the RD-prior. In the normal image, there is a perfect match between gray matter activity and anatomy. Consequently, by further increasing the weights of the prior, bias and noise can sometimes be further decreased. However, this would then result in a poorer reconstruction of the cold gray matter lesions in the diseased brain. These cold lesions are invisible in the anatomical image, and with increased weight of the priors, there will be increased smoothing over the lesion boundary within the gray matter. As an example, one can look at the results for the A-MAP-reconstructions using perfect segmentation information ($-+-$, Fig. 7 and 8), where 10β yields lower bias and noise in the GM, but higher bias on the lesion activity estimate and on the lesion signal.

Also note that a lower bias can be achieved when using an anatomical prior than when using MLEM without regularization, which is supposed to be an unbiased estimator (at least asymptotically). This is because we model the resolution. Even after many iterations, resolution recovery is imperfect, and the images suffer from Gibbs over- and undershoots, which cause bias. The priors suppress these over- and undershoots, yielding reduced bias.

The reconstruction image quality of the lesions can best be appreciated in Fig. 7(b) and 8(b), which show the SNR in the lesions, averaged over the 20 lesions. All evaluated algorithms yielded a better SNR than unprocessed MLEM. A-MAP with ideal segmentation information outperformed all other methods in terms of SNR. As the A-prior was specially designed for FDG-PET brain imaging of patients with hypointense lesions, this is in agreement with the expectations. Furthermore, the Bowsher-prior using a large neighborhood was also found to result in an improved SNR compared to postsmoothed MLEM and all JE-algorithms, and a similar SNR as A-MAP using an SPM segmentation, the Bowsher-algorithm using a small neighborhood and MAP-RD. At the same time, all anatomical priors yield clearly better bias-noise characteristics in the GM voxels than the conventional algorithms (Fig. 7(a) and 8(a)). Therefore, the modified Bowsher-prior can be considered a good alternative for A-MAP in applications where accurate segmentation is not possible. Both are superior to MAP-RD.

Based on Fig. 7(b) and 8(b) one could argue that β should be increased. Indeed, SNR improves for most algorithms when the prior information is given more weight. However, the focus of this work was on quantitative accuracy, which obviously decreases with increased regularization (i.e. more bias is observed in the GM, the lesion VOIs and on the signal). Comparing the algorithms for their lesion detectability performance would require a different FOM to be optimized and evaluation with an observer study, which was considered out of the scope of this work. Previous studies reported either improved lesion detection when using anatomical information or no significant difference, depending on the anatomical prior and the detection task (see discussion in [8] for an overview).

After tuning the JE-parameters carefully, the JE-prior clearly outperformed postsmoothed MLEM and regular MAP in terms of gray matter quantification (see Fig. 7(a) and Fig. 8)(a), but to a lesser extent than the Bowsher- and the A-prior. Similarly, Tang and Rahmim recently reported improved contrast versus noise tradeoff with the JE-prior compared to using a regular MAP reconstruction algorithm if a very hot lesion is present in the brain [34]. Not only is the difference in activity in our simulated lesions small in comparison with that in the healthy GM, the number of lesion voxels is also much smaller than the number of healthy GM voxels, which makes cluster separation very hard. We were unable to find combinations of parameters (number of bins, width of the Parzen window, weight of the prior) that produced a better signal than those shown in Fig. 7(b) and Fig. 8(b). If the number of histogram bins for the PET image was increased to obtain a better lesion signal, noise propagation deteriorated, because the prior then accepted the noise as small isolated clusters in the joint histogram. Similarly, the homogeneity of the GM activity decreased with increasing number of PET bins. Increasing the width of the Parzen window accordingly, cancels these effects. An additional problem of the JE-prior is the presence of local maxima. Hence, determining adequate parameter settings is very tedious. For this brain image, the best results were obtained by starting the iterations from a low-noise Bowsher-image, or from a uniform image and gradually increasing the weight of the prior, rather than by starting from an MLEM-reconstruction. Unfortunately, the quantitative accuracy in the JE reconstruction images was also found to be quite sensitive to the chosen parameters and to the noise in the MRI and in the emission data (see Fig. 6).

Note that the JE-prior operates on all image intensities simultaneously, disregarding their positional information, while the Bowsher-prior operates locally. In this work, we briefly examined the effect of combining the global properties of the JE-prior and the local ones of the Bowsher-prior. The global FOM was only marginally better than that obtained with the Bowsher-prior alone. As the combined use requires a very difficult parameter tuning due to the local maxima, this tract was abandoned. Other research groups are currently exploring the advantage of maximizing the joint entropy of other features, like boundaries, in addition to the JE of the intensities. Somayajula *et al.* [35] used a scale space approach, whereas Tang and Rahmim [36] resort to wavelets. Van de Sompel *et al.* recently proposed methods to increase robustness against local optima in [37].

In contrast to the JE-prior, the Bowsher-prior behaves more predictably, and its parameters are much more easily tuned. In addition, this prior always converged in practice. Furthermore, it does not require a segmentation, whereas A-MAP relies

heavily on the accuracy of the segmentation. Consequently, we will further study the (modified) Bowsher-prior for clinical applications. As no segmentation is needed, it might also be promising for PET/CT reconstruction in oncology in cases where a good registration is available.

V. Conclusion

Three anatomical priors were evaluated in a realistic simulation study of PET brain imaging with ^{18}F-FDG, co-registered to an MR-image. All three priors outperformed postsmoothed MLEM and MAP with the relative difference prior in terms of quantitative accuracy. The prior previously proposed by Bowsher *et al.* and recently modified by our group yielded comparable results for this realistic and, in our opinion clinically relevant simulation experiment as the dedicated A-MAP algorithm proposed by Baete *et al.*, and both obtained improved partial volume correction without degradation of the lesion signal-to-noise ratio. Not relying on an accurate segmentation, the modified Bowsher-algorithm is expected to be a good choice for PET-MRI brain imaging applications where the regions-of-interest are hard to segment.

Acknowledgment

The authors would like to thank Michel Koole for performing the point source measurements and Frédéric Noo and Patrick Dupont for the very interesting discussions.

Appendix

In this appendix, details are provided about the implementation of the Bowsher-prior. In fact, the Bowsher-prior is rather a neighbor selection mechanism than a type of prior. Any Markov prior could be used. The log-prior M, which is a function of the emission image $\boldsymbol{\lambda}$, is typically written as

$$M(\boldsymbol{\lambda}) = \sum_j \sum_k w_{jk} M_{jk} \qquad \text{(A-1)}$$

with M_{jk} the logarithm of the Markov prior to be applied to voxel j and its neighbor k, and w_{jk} the weighting factor (1 if $k \in N_j$, 0 otherwise). For MAP reconstruction, we iterate over the gradient ascent algorithm described in [27], which requires the calculation of the first derivative of M

$$\frac{\partial M}{\partial \lambda_j} = \sum_k w_{jk} \frac{\partial M_{jk}}{\partial \lambda_j} + \sum_k w_{kj} \frac{\partial M_{kj}}{\partial \lambda_j} \qquad \text{(A-2)}$$

In [3], we showed that the quantitative accuracy in FDG PET brain images improves when we change the definition of this gradient into

$$\frac{\partial M}{\partial \lambda_j} = \sum_k w_{jk} \frac{\partial M_{jk}}{\partial \lambda_j} \qquad \text{(A-3)}$$

Therefore, (A-3) was used in this work. The second derivative of M was changed accordingly. Although no function M exists that satisfies (A-3) and proving whether or not the algorithm converges is non-trivial, no convergence problems were experienced in practice.

References

[1] K Baete, J Nuyts, W Van Paesschen, P Suetens, P Dupont. "Anatomical-based FDG-PET reconstruction for the detection of hypo-metabolic regions in epilepsy", *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 510-519, 2004.

[2] J Nuyts. "The use of mutual information and joint entropy for anatomical priors in emission tomography", *IEEE Nucl. Sci. Symp. Conf. Record*, 2007, pp. 4149-4154.

[3] K Vunckx, J Nuyts. "Heuristic modification of an anatomical Markov prior improves its performance", *IEEE Nucl. Sci. Symp. Conf. Record*, 2010, pp.3262-3266.

[4] JE Bowsher, H Yuan, LW Hedlund, TG Turkington, G Akabani, A Badea, WC Kurylo, CT Wheeler, GP Cofer, MW Dewhirst, GA Johnson. "Using MRI information to estimate F18-FDG distributions in rat flank tumors", *IEEE Nucl. Sci. Symp. Conf. Record*, 2004, pp. 2488-2492.

[5] B Lipinski, H Herzog, E Rota Kops, W Oberschelp, HW Müller-Gärtner. "Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information", *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 129-136, 1997.

[6] A Rangarajan, IT Hsiao, G Gindi. "A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction", *J. Math. Imaging Vision*, vol. 12, no. 3, pp. 199-217, 2000.

[7] C Comtat, PE Kinahan, JA Fessler, T Beyer, DW Townsend, M Defrise, C Michel. "Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels", *Phys. Med. Biol.*, vol. 47, no. 1, pp. 1-20, 2002.

[8] S Kulkarni, P Khurd, I Hsiao, L Zhou, G Gindi. "A channelized Hotelling observer study of lesion detection in SPECT MAP reconstruction using anatomical priors", *Phys. Med. Biol.*, vol. 52, no. 12, pp. 3601-3617, 2007.

[9] X Ouyang, WH Wong, VE Johnson, X Hu, C-T Chen. "Incorporation of correlated structural images in PET image reconstruction", *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 627-640, 1994.

[10] K Baete, J Nuyts, K Van Laere, W Van Paesschen, S Ceyssens, L De Ceuninck, O Gheysens, A Kelles, J Van den Eynden, P Suetens, P Dupont. "Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET", *NeuroImage*, vol. 23, no. 1, pp. 305-317, 2004.

[11] S Somayajula, E Asma, RM Leahy. "PET image reconstruction using anatomical information through mutual information based priors", *IEEE Nucl. Sci. Symp. Conf. Record*, 2005, pp. 2722-2726.

[12] A Reilhac, C Lartizien, N Costes, S Sans, C Comtat, RN Gunn, AC Evans. "PET-SORTEO: A Monte Carlo-based simulator with high count rate capabilities", *IEEE Trans. Nucl. Sci.*, vol. 51, no. 1, pp. 46-52, 2004.

[13] A Reilhac. "Validation et exploitation d'un simulateur TEP de Monte Carlo", PhD-thesis, Institut National des Sciences Appliques de Lyon, 2007.

[14] website: "http://www.bic.mni.mcgill.ca/brainweb/".

[15] DL Collins, AP Zijdenbos, V Kollokian, JG Sled, NJ Kabani, CJ Holmes, AC Evans. "Design and construction of a realistic digital brain phantom", *IEEE Trans. Med. Imaging*, vol. 17, no. 3, pp. 463-468, 1998.

[16] J Cheng-Liao, J Qi. "PET image reconstruction with incomplete anatomical edge information using level set method", Presented at *IEEE NSS MIC*, Orlando, FL, 2009, M03-5.

[17] PJ Slomka, RP Baum. "Multimodality image registration with software: state-of-the-art", *Eur. J. Nucl. Med. Mol. Imaging*, vol. 36, suppl. 1, pp. 44-55, 2009.

[18] HPW Schlemmer, BJ Pichler, M Schmand, Z Burbar, C Michel, R Ladebeck, K Jattke, D Townsend, C Nahmias, PK Jacob, WD Heiss, CD Claussen. "Simultaneous MR/PET imaging of the human brain: feasibility study", *Radiology*, vol. 248, no. 3, pp. 1028-1035, 2008.

[19] K Mevel, B Desgranges, JC Baron, B Landeau, V de la Sayette, F Viader, F Eustache, G Ch'etelat. "Detecting hippocampal hypometabolism in Mild Cognitive Impairment using automatic voxel-based approaches", *Neuroimage*, vol. 37, no. 1, pp. 18-25, 2007.

[20] G Chételat, B Desgranges, B Landeau, F Mézenge, JB Poline, V de la Sayette, F Viader, F Eustache, JC Baron. "Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer's disease", *Brain*, vol. 131, part 1, pp. 60-71, 2008.

[21] BA Thomas, K Erlandsson, M Modat, L Thurfjell, R Vandenberghe, S Ourselin, BF Hutton. "The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease", *Eur. J. Nucl. Med. Mol. Imaging*, vol. 38, no. 6, pp. 1104-1119, 2011.

[22] K Goffin, W Van Paesschen, P Dupont, K Baete, A Palmini, J Nuyts, K Van Laere. “Anatomy-based reconstruction of FDG-PET images with implicit partial volume correction improves detection of hypometabolic regions in patients with epilepsy due to focal cortical dysplasia diagnosed on MRI”, *Eur. J. Nucl. Med. Mol. Imaging*, vol.37, no. 6, pp. 1148-1155, 2010.

[23] V Frouin, C Comtat, A Reilhac, MC Grégoire. “Correction of partial-volume effect for PET striatal imaging: Fast implementation and study of robustness”, *J. Nucl. Med.*, vol. 43, no. 12, pp. 1715-1726, 2002.

[24] M Soret, PM Koulibaly, J Darcourt, I Buvat. “Partial volume effect correction in SPECT for striatal uptake measurements in patients with neurodegenerative diseases: impact upon patient classification”, *Eur. J. Nucl. Med. Mol. Imaging*, vol. 33, no. 9, pp. 1062-1072, 2006.

[25] J De Reuck, D Decoo, MC Hasenbroekx, B Lamont, P Santens, P Goethals, K Strijckmans, I Lemahieu. “Acetazolamide vasoreactivity in vascular dementia: a positron emission tomographic study”, *Eur Neurol.*, vol. 41, no. 1, pp. 31-36, 1999.

[26] P Besson, F Andermann, F Dubeau, A Bernasconi. “Small focal cortical dysplasia lesions are located at the bottom of a deep sulcus”, *Brain*, vol. 131, no. 12, pp. 3246-3255, 2008.

[27] J Nuyts, D Bequé, P Dupont, L Mortelmans. “A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography”, *IEEE Trans. Nucl. Sci.*, vol. 49, no. 1, pp. 56-60, 2002.

[28] B Sadzot, R Debets, P Maquet, C Comar, G Franck. “PET studies of patients with partial epilepsy: visual interpretation vs. semi-quantification/quantification”, *Acta Neurologica Scandinavica*, vol. 89, no. S152, pp. 175178, 1994.

[29] J Qi. “Comparison of lesion detection and quantification in MAP reconstruction with Gaussian and non-Gaussian priors”, *Int. J. Biomed. Imaging*, vol. 2006, Article ID 87567, 10 pages.

[30] J Nuyts, C Michel, L Brepoels, L De Ceuninck, C Deroose, K Goffin, FM Mottaghy, S Stroobants, J Van Riet, R Verscuren. “Performance of MAP reconstruction for hot lesion detection in whole-body PET/CT: An evaluation with human and numerical observers”, *IEEE Trans. Med. Imaging*, vol. 28, no. 1, pp. 67-73, 2009.

[31] Software package, Wellcome department of cognitive neurology (London, UK), http://www.fil.ion.ucl.ac.uk/spm

[32] LS Shepp, Y Vardi. “Maximum likelihood reconstruction for emission tomography.”, *IEEE Trans. Med. Imaging*, vol. MI-1, pp. 113-122, 1982.

[33] HM Hudson, RS Larkin. “Accelerated image reconstruction using ordered subsets of projection data.”, *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 601-609, 1994.

[34] J Tang, A Rahmim. “Bayesian PET image reconstruction incorporating anato-functional joint entropy”, *Phys. Med. Biol.*, vol. 54, no. 23, pp. 7063-7075, 2009.

[35] S Somayajula, C Panagiotou, A Rangarajan, Q Li, SR Arridge, RM Leahy. “PET Image Reconstruction Using Information Theoretic Anatomical Priors.”, *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 537-549, 2011.

[36] J Tang, A Rahmim. “Anatomy assisted MAP-EM PET image reconstruction incorporating joint entropies of wavelet subband image pairs”, *IEEE Nucl. Sci. Symp. Conf. Record*, 2009, pp. 3741-3745.

[37] D Van de Sompel, M Brady. “Robust joint entropy regularization of limited view transmission tomography using Gaussian approximations to the joint histogram”, *Information Proc. in Med. Imaging*, 2009, vol. 5636, pp. 638-650.