The Validation Problem of Joint Emission/Transmission Reconstruction From TOF-PET Projections

I. Introduction

About six years ago, it was shown with experiments and with mathematical analysis that with the introduction of time-of-flight (TOF), the data acquired with positron emission tomography (PET) were significantly richer than before [1]–[3]. In particular, it was found that if the tracer distribution is wider than the TOF uncertainty, the TOF-PET data provide enough information to jointly estimate the activity image and the attenuation sinogram up to a global scale factor. Since then, many algorithms have been proposed to exploit this feature to perform attenuation correction without additional transmission measurements or to improve available attenuation images or sinograms. An excellent recent review of these methods has been published by Berker and Li [4]. Some of these methods rely on the emission data only, other methods use additional information to determine the scale factor and/or to suppress noise propagation. This additional information can be obtained from an anatomical image acquired with the MR-component in a PET/MR system or with the CT-component in a PET/CT system.

These contributions confirm that joint estimation of the activity and attenuation indeed works on realistic simulations and on real data [5], [6]. However, it is also often observed that it does not work as well on real data as predicted by theoretical analysis and as observed in simulations, and additional tricks or careful tuning of dedicated penalties are often used to obtain clinically acceptable results. As an example, Mehranian and Zaidi [7] observed cross-talk between the activity and attenuation reconstructions obtained with the MLAA algorithm [8], while according to the theory, the reconstructed activity image should be accurate up to a multiplicative constant. These findings hinder validation of joint estimation against the current reference method, which is maximum likelihood (ML) reconstruction using a CT-based estimate of the attenuation.

In this paper, we argue that these problems are at least partly due to the sensitivity of the reconstructed images to inaccuracies in the system matrix and in the estimates of scatter or randoms contributions. Such errors make the data inconsistent, and in that case, conventional ML reconstruction and a joint estimation algorithm will converge to different solutions, which both differ from the true solution. Thus, it seems that for good performance of joint estimation algorithms, the calibration of PET systems must be improved. Our preliminary results indicate that also the conventional ML reconstruction will benefit from such an improved calibration.

II. PET System Calibration Problems

In simulation experiments, the data are often (almost) consistent for the system matrix and the scatter and/or randoms estimates which are used during reconstruction. In contrast, real scans are typically far less consistent, because there are always inaccuracies in the system matrix and the scatter estimate. This is an important complication for the validation of reconstruction algorithms that jointly estimate the activity and attenuation from TOF-PET data. Conventional ML-reconstruction maximizes the likelihood by estimating the activity image only. In contrast, joint estimation methods not only estimate the activity but also the attenuation, which makes their set of possible solutions larger than that of ML-reconstruction. As a result, for inconsistent data the joint estimation algorithm will converge to a different solution than the ML-reconstruction, and that solution typically has a higher likelihood. The estimate of both algorithms will have bias, and it is not always obvious which of the two is the better one. In the following, this problem is illustrated with a toy example where the reconstructions are done either with a correct or with an inaccurate estimate of the scatter contribution.

Fig. 1 shows the true activity and attenuation images for the 2-D toy experiment, which simulates a cardiac TOF-PET study. The diameter of the PET field of view was 60 cm. The simulated TOF sinogram has a TOF resolution of 420 ps FWHM, 150 radial bins of 4 mm, 100 angular bins covering 180°, and 54 TOF-bins of 105 ps. The simulation included the effect of attenuation and of a scatter distribution with a...
scatter to primary ratio of 0.9. Fig. 1 shows images of the TOF-integrated sinograms of the prompts (i.e., primary + scatter) and the scatter only. Other effects such as Poisson noise, randoms, singles, finite spatial resolution, etc., were ignored.

Two ML expectation maximization (MLEM) [9] and two MLAA [8] reconstructions were computed from these TOF-sinograms. For all four reconstructions, 30 iterations with 20 subsets were applied. MLEM uses the known attenuation image for the attenuation correction, whereas MLAA computes ML estimates of both the attenuation and the activity image. The first reconstruction was based on an exact model of the acquisition. In the second reconstruction, the scatter contribution was overestimated by 20%. To determine the unknown scale factor, the MLAA activity image was scaled after each iteration such that its total activity was equal to that of the corresponding MLEM reconstruction, facilitating the comparison of the images. The first reconstruction was based on an exact model of the acquisition. In the second reconstruction, the scatter contribution was overestimated by 20%. To determine the unknown scale factor, the MLAA activity image was scaled after each iteration such that its total activity was equal to that of the corresponding MLEM reconstruction, facilitating the comparison of the images. The figures show that when an exact model is used, the difference between the reconstructed and true images is very small for both algorithms, as is the difference between the MLEM and MLAA activity images. In contrast, when the scatter was overestimated during reconstruction, MLEM and MLAA produced different images: the MLEM image has negative bias everywhere and shows some cross-talk with the attenuation (the bone shows up although it has the same activity as the surrounding tissue), whereas the bias of the MLAA image is more concentrated near the heart and does not show that cross-talk.

Of course, the bias of MLAA will depend on how the scale factor is determined. Since here the MLAA image was scaled to have the same total activity as the MLEM image, it has the same mean negative bias. If MLAA would be used in a clinical application, a different method to determine the scale factor would be required. A popular approach is to choose the scale factor such that the correct attenuation in soft tissue is obtained. This was done to produce Fig. 4, revealing a very different bias of the MLAA images and illustrating that the problem of the scale factor further complicates the validation of MLAA against MLEM.

For the first simulation (the one of Figs. 2 and 3) we computed the difference between the achieved log-likelihood and the highest possible value of the log-likelihood (i.e., when the forward projection is identical to the sinogram) for each reconstruction, which produced for the ideal model the values $-0.23$ (MLEM) and $-1.75$ (MLAA), and for the invalid model $-321.8$ (MLEM) and $-294.1$ (MLAA). In the ideal case, both algorithms achieved a very high likelihood, but MLEM did slightly better because it converged faster than MLAA (its activity updates were more effective because it used the exact attenuation from the start, whereas MLAA had to estimate it). In the invalid model case, both algorithms reached a significantly lower likelihood because both failed to fit the data well, but MLAA approached the data better than MLEM because it could also manipulate the attenuation image. The two algorithms produced different reconstructions, which were both clearly biased.

### III. Refining the Calibration

In working with real scanner data, we have struggled to validate MLAA against the reference method MLEM with CT-based attenuation. For that reason, we have attempted to refine the PET system calibration, and our current results...
indicate that this indeed results in better agreement between MLAA and MLEM. Possible sources of data inconsistencies include inaccuracies in the scatter estimate used for Compton scatter correction, in the time calibration of the detectors, in the estimate of the TOF resolution, and in the estimated detector efficiencies [11], [12].

As illustrated in the example above, ML-reconstruction reaches a lower likelihood when the data are inconsistent. Therefore, it seems feasible to estimate the value of some additional calibration parameters by further increasing the data consistency, or equivalently, the likelihood. This is expected to be more effective when the number of parameters is small.

In PET, the scatter contribution is typically estimated with accelerated Monte Carlo simulation based on the attenuation image and a preliminary reconstruction of the activity image, as originally proposed by Ollinger [13]. This method has been extended to TOF-PET [14]. The method is approximate and to correct for modeling errors, the scatter estimate is improved by scaling it such that it better agrees with the available TOF-PET sinograms. The scaling is usually done by fitting the estimated scatter profile to the (randomly corrected) measured data for lines-of-response (LORs) not intersecting the patient body, because these counts are almost entirely due to scattered photon pairs. To further improve on that, several authors have proposed to not only use these “scatter tails” but instead the photon pairs. To further improve on that, several authors have proposed to not only use these “scatter tails” but instead the entire sinogram [15]–[18]. We have used the scatter scaling method of [17] to scale the scatter estimate produced with the vendor software. This algorithm estimates the scatter scale factor during ML reconstruction using the CT-based attenuation correction.

We have also assumed that there might be small inaccuracies in the timing synchronization, the estimates of the TOF-resolution, and the detector efficiencies. These inaccuracies affect all the LORs of the sinogram, and there is a huge number of them. However, by parameterizing these errors as a function of the detectors rather than of the detector pairs, the number of parameters becomes relatively small.

In TOF-sinograms, the difference in detected photon arrival time is used to determine the coordinate of the annihilation point along the LOR. Werner and Karp [19] assumed that for the two detectors \( j \) and \( k \), the arrival times for event \( i \) are determined as \( t_{i,j} − t_{0,j} \) and \( t_{i,k} − t_{0,k} \), respectively, where \( t_{0,j} \) is a reference time associated with detector \( j \). Subtraction of these time measurements produces the desired arrival time difference if both detectors are perfectly synchronized, i.e., if \( t_{0,j} = t_{0,k} \). Therefore, synchronization errors will cause (signed) errors in the positioning of TOF-events along the LOR, equal to \( c(t_{0,k} − t_{0,j})/2 \), where \( c \) is the speed of light (and assuming that the TOF-axis points from detector \( j \) to detector \( k \)). In [20], we followed an approach inspired by that of [19] to estimate the TOF-reference time offsets \( t_{0,j} \) for every detector \( j \).

A very similar approach was followed to estimate the TOF-resolution along every LOR and the detector sensitivity for each LOR. It was assumed that both can be modeled as a combination of features of the two detectors contributing to each LOR. Therefore, we estimate the TOF-resolution of each detector, and compute the TOF-resolution along the LOR as

\[
\text{FWHM}_{j,k} = \sqrt{\text{FWHM}_{j}^2 + \text{FWHM}_{k}^2}
\]

where \( \text{FWHM}_{j,k} \) is the TOF-resolution along the LOR connecting detectors \( j \) and \( k \), expressed as the full width at half maximum, and \( \text{FWHM}_{j} \) is the TOF-resolution of detector \( j \). The vendor system software provides a normalization sinogram which contains the sensitivity correction for each LOR. We have estimated corrections \( \epsilon_j \) to the efficiencies by assuming

\[
s_{j,k} = \epsilon_j \epsilon_k/n_{j,k}
\]

where \( s_{j,k} \) is the final sensitivity, \( \epsilon_j \) is the efficiency correction for detector \( j \), and \( n_{j,k} \) is the vendor normalization for the LOR connecting detectors \( j \) and \( k \). This way of parameterizing detector pair sensitivities has often been used before [21]–[24].

We first estimate the scatter scale during an ML-reconstruction as described in [17]. Using this improved scatter scale, a new ML-reconstruction is started with joint estimation of all the other corrections. This approach was followed because in our experience, the scatter scale estimate is hardly affected by the other corrections, whereas there is a considerable inter-dependence between the estimates of detector efficiencies, TOF-resolutions, and TOF-offsets.

IV. PHANTOM EXPERIMENT

A uniform cylindrical phantom (the so-called WCC phantom with diameter of 19 cm) was filled with 21 MBq \(^{18}\text{F}-\text{FDG}\) in water and scanned for 10 min on the GE Signa PET/MR system. This system has a TOF resolution of approximately 390 ps (excluding the effect of binning). Standard MLEM reconstructions were made using the vendor calibrations. For attenuation correction, this reconstruction relies on a template attenuation map stored in the system, which can be applied here because the position of the phantom is known, thanks to the use of a dedicated phantom holder. Additional MLEM corrections were made by including one or several of the corrections described above. Similarly, MLAA reconstructions were computed without and with additional corrections. During the MLAA reconstructions, the phantom attenuation template was incorporated in the system matrix. Since MLAA estimates a non-negative attenuation image, this means that MLAA could increase the attenuation everywhere, but it could not make it lower than in the template. The reconstructed images were postsmoothed with a 3-D Gaussian filter of \( \sim 8 \text{ mm FWHM} \). Fig. 5 shows the MLEM and MLAA reconstructions. The MLAA reconstructions are clearly affected by the corrections to the system calibration. The effects on the MLEM reconstructions seem less obvious, but they are not negligible. As more corrections are included, the MLAA and MLEM reconstructions become more similar.

Fig. 6 shows images of the estimated detector TOF-offsets, the TOF resolutions, and the efficiency corrections. Most of these corrections are fairly small. There are five detector blocks which have a significantly higher FWHM for the TOF-kernel. The PET system has two ways of determining the scintillation time of an event. If for some reason the first way
Fig. 5. MLEM and MLAA reconstructions using standard vendor corrections (left), and with additional corrections for scatter, detector efficiencies, TOF-offsets, and TOF-resolutions as described in the text. From left to right more corrections are applied, so the images at the far right are produced using all proposed corrections.

Fig. 6. Images of the estimated crystal offsets (intensity range $[-50 \text{ ps}, 50 \text{ ps}]$), TOF resolutions (range $[176 \text{ ps}, 389 \text{ ps}]$), and efficiency corrections (range $[0.75, 1.25]$). The images show the values for all crystals, which are organized as 45 rings (column index), with 448 crystals (row index) per ring.

Fig. 7. Plots of the total squared difference between MLAA and MLEM for reconstructions with standard corrections (black), and when improving the calibration by adding scatter scaling (red), efficiency correction (green), TOF-offset correction (blue), and TOF-resolution (purple). Note that in this list the corrections are accumulated, so the purple curve is for all corrections combined.

Table I

<table>
<thead>
<tr>
<th>S</th>
<th>E</th>
<th>O</th>
<th>R</th>
<th>MLEM Mean Squared Difference</th>
<th>MLAA Mean Squared Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>E</td>
<td>O</td>
<td>R</td>
<td>MLAA - MLEM</td>
<td>MLAA - MLEM</td>
</tr>
<tr>
<td>scatter</td>
<td>efficiency</td>
<td>TOF-offset</td>
<td>resolution</td>
<td>3645</td>
<td>3043</td>
</tr>
<tr>
<td>414</td>
<td>431</td>
<td>33</td>
<td>0</td>
<td>MLAA - MLEM</td>
<td>MLAA - MLEM</td>
</tr>
<tr>
<td>2122</td>
<td>1718</td>
<td>54</td>
<td>0</td>
<td>MLAA - MLEM</td>
<td>MLAA - MLEM</td>
</tr>
<tr>
<td>2312</td>
<td>1362</td>
<td>178</td>
<td>0</td>
<td>MLAA - MLEM</td>
<td>MLAA - MLEM</td>
</tr>
<tr>
<td>219</td>
<td>3798</td>
<td>219</td>
<td>0</td>
<td>MLAA - MLEM</td>
<td>MLAA - MLEM</td>
</tr>
</tbody>
</table>

Table I lists the mean sum of squared differences between various reconstructions to illustrate the effect of the different corrections. Only the first 70 planes (of 89 total planes) were used in the calculation, to eliminate the influence of the mismatch between the template and the true object near the phantom holder. The log-likelihood values for each of the MLEM and MLAA reconstructions with different calibration corrections are shown in Fig. 8. If no corrections are applied, MLAA obtains a better likelihood despite its slower convergence, and the likelihood increases if more corrections are applied.

V. DISCUSSION

Many authors have reported on the evaluation of algorithms that jointly estimate the activity and attenuation from TOF-PET data, comparing the reconstructed images to those of the reference method, which is MLEM or OSEM using the CT-based attenuation map for attenuation correction. Although promising results have been obtained, some authors report that strong regularization is required to suppress cross talk artifacts and obtain correct attenuation values. In theory, there should be no cross talk artifacts, and since the activity image and attenuation sinogram are determined up to a constant factor, a gentle constraining should suffice to drive the algorithm to the correct solution.

The simulation and phantom data above indicate that inconsistencies in the data drive conventional MLEM and MLAA to different solutions, which are both wrong. This is in agreement with a few recent publications on the sensitivity of MLAA to
system modeling errors. Zhu et al. [12] investigated the sensitivity of MLAA to various calibration factors and reported a high sensitivity to timing offset errors. Cheng et al. [26] found that the use of a matched TOF kernel caused cross-talk artifacts in MLAA reconstructions.

Table I shows that even on a state-of-the-art TOF-PET scanner, this effect is not negligible. By applying a data-driven calibration approach, the acquisition model was modified. We believe that these modifications are improvements and therefore lead to a reduction of the data inconsistencies. This in turn makes the MLEM and MLAA activity images more similar. The table shows that the impact of the scatter scaling was larger for MLEM than for MLAA, whereas the offset correction affected MLAA more than MLEM. However, a more systematic investigation of the effect of each correction must be done, because the corrections interfere: if a particular correction is not applied, the estimates of the other corrections will be affected by that. Therefore, Table I does not provide enough information to accurately assess the relative importance of each individual correction, but it clearly shows that MLEM and MLAA are both affected significantly by the full refined calibration.

The need for improved modeling of the acquisition physics is illustrated here with a phantom measurement on the GE Signa PET/MR system. It seems likely that a similar need exists for other TOF-PET scanners too, as illustrated by the improved similarity between MLEM and MLAA we observed for Siemens mCT TOF-PET data after estimating the scatter scale factor using all the LORs [18].

In our experiment with real data, MLAA used a system matrix that already accounted for the phantom attenuation. Consequently, MLAA could only reconstruct additional attenuation, which should have been zero if the data were perfect. Although this is a significant constraint, it gave MLAA enough freedom to produce reconstructions that were different from those of MLEM. In experiments where MLAA was allowed to estimate the total attenuation (not shown here), we obtained again improved similarity between MLEM and MLAA when more corrections were applied. However, the final similarity was not as good as the one reported here, indicating that there are still other sources of inconsistencies waiting to be identified and corrected.

VI. Conclusion

Introduction of MLAA in clinical practice requires a careful validation by comparing the MLAA-images to those produced by the current reference method, MLEM with CT-based attenuation. We find that this requires a refinement of the TOF-PET system calibration. With the current calibration accuracy, the algorithms converge to different solutions, which are both different from the ground truth. Therefore, it is necessary to refine the calibration, not only to improve the similarity between the MLAA and MLEM images (as predicted by theory) but also to reduce the errors in the current clinical images, reconstructed by the current method of choice, i.e., MLEM with CT-based attenuation correction.

ACKNOWLEDGMENT

The authors would like to thank T. Deller, S. Ahn, and F. Jansen from General Electric and V. Panin from Siemens for helpful discussions. This work was supported by the Research Foundation Flanders (FWO) under Project G.0275.14N and Project 12T7118N.

JOHAN NYUITS
Department of Imaging and Pathology
Nuclear Medicine & Molecular Imaging, Medical Imaging Research Center
KU Leuven – University of Leuven
3000 Leuven, Belgium

AHMADREZA REZAEI
Department of Imaging and Pathology
Nuclear Medicine & Molecular Imaging, Medical Imaging Research Center
KU Leuven – University of Leuven
3000 Leuven, Belgium

MICHEL DEFRISSE
Department of Nuclear Medicine
Vrije Universiteit Brussel
1090 Brussels, Belgium

REFERENCES


