Evaluation of lesion detectability in positron emission tomography when using a convergent penalized likelihood image reconstruction method

Kristen A. Wangerin
Sangtae Ahn
Scott Wollenweber
Steven G. Ross
Paul E. Kinahan
Ravindra M. Manjeshwar

Evaluation of lesion detectability in positron emission tomography when using a convergent penalized likelihood image reconstruction method

Kristen A. Wangerin,^{a,b,*} Sangtae Ahn,^a Scott Wollenweber,^c Steven G. Ross,^c Paul E. Kinahan,^{b,d} and Ravindra M. Manjeshwar^a

^aGeneral Electric Global Research Center, 1 Research Circle, Niskayuna, New York 12309, United States

^bUniversity of Washington, Department of Bioengineering, 3720 15th Avenue NE, Seattle, Washington 98195, United States

^cGeneral Electric Healthcare, 3000 North Grandview Boulevard, Waukesha, Wisconsin 53188, United States

^dUniversity of Washington, Department of Radiology, 1959 NE Pacific Street, Seattle, Washington 98195, United States

Abstract. We have previously developed a convergent penalized likelihood (PL) image reconstruction algorithm using the relative difference prior (RDP) and showed that it achieves more accurate lesion quantitation compared to ordered subsets expectation maximization (OSEM). We evaluated the detectability of low-contrast liver and lung lesions using the PL-RDP algorithm compared to OSEM. We performed a two-alternative forced choice study using a channelized Hotelling observer model that was previously validated against human observers. Lesion detectability showed a stronger dependence on lesion size for PL-RDP than OSEM. Lesion detectability was improved using time-of-flight (TOF) reconstruction, with greater benefit for the liver compared to the lung and with increasing benefit for decreasing lesion size and contrast. PL detectability was statistically significantly higher than OSEM for 20 mm liver lesions when contrast was ≥ 0.5 ($p < 0.05$), and TOF PL detectability was statistically significantly higher than TOF OSEM for 15 and 20 mm liver lesions with contrast ≥ 0.5 and ≥ 0.25 , respectively. For all other cases, there was no statistically significant difference between PL and OSEM ($p > 0.05$). For the range of studied lesion properties, lesion detectability using PL-RDP was equivalent or improved compared to using OSEM. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.4.1.011002]

Keywords: positron emission tomography imaging; lesion detection; penalized likelihood; maximum likelihood; model observers.

Paper 16124SSR received Jun. 24, 2016; accepted for publication Oct. 18, 2016; published online Nov. 22, 2016.

1 Introduction

Positron emission tomography (PET) imaging is clinically used in the detection and quantitation of tracer-avid lesions or tumors in cancer. Ordered subsets expectation maximization (OSEM)¹ is currently a standard image reconstruction algorithm for clinical PET imaging. The OSEM algorithm is not convergent, and noise increases rapidly with iteration. Noise is controlled in OSEM by stopping after a few iterations and postfiltering the image. Despite its popularity, the behavior of the algorithm is unpredictable and the joint performance of detection and quantitation using OSEM is not well understood. In contrast, convergent penalized likelihood (PL) image reconstruction algorithms² show predictable behaviors, and the parameters can be tuned to control the trade-off between image noise and resolution.^{3–7} These algorithms, therefore, have the potential to provide well-defined image quality for tasks such as lesion detectability^{8,9} and region of interest (ROI) quantitation accuracy¹⁰ when the penalty functions and parameters have been optimized. Although PL algorithms have been in the research community for over three decades,¹¹ they have not been adopted clinically due to the different and unfamiliar image appearance compared to OSEM, as well as due to increased computation time. For example, nonquadratic edge-preserving penalties often produce undesirable patchy or “mosaic”-like images.^{12,13}

In previous work, we developed a PL algorithm¹⁴ based on the relative difference prior (RDP),¹⁵ which used a block sequential regularized expectation maximization (BSREM) optimizer.^{16,17} The RDP applies activity-dependent smoothing and suppresses image noise in low-activity background regions. The degree of edge preservation is also controlled. Studies comparing PL-RDP and OSEM have demonstrated that PL-RDP achieves better lesion quantitation accuracy with higher signal-to-noise ratio (SNR)^{14,18,19} while maintaining subjective image quality^{19–23} compared to OSEM. Studies have also shown that PL-RDP produces higher and more accurate lesion standardized uptake values (SUVs)^{14,18–25} than OSEM, suggesting there could be a potential improvement in lesion detectability; indeed, recent clinical studies have reported that the PL-RDP produced higher lesion detectability scores than OSEM in visual image quality evaluation studies.^{19,21}

These existing studies used relatively high-contrast lesions, and we were concerned about whether the edge-preserving RDP could smooth out low-contrast lesions, degrading lesion detectability. A study has not yet been performed to evaluate lesion detectability of low-contrast lesions with varying sizes and contrasts for the PL-RDP algorithm¹⁴ in comparison to OSEM. In a preliminary study,²⁶ we performed a two-alternative forced choice detectability study^{27,28} using PL-RDP and OSEM and calibrated a channelized Hotelling observer (CHO) model against human observers.²⁹ In this work, we expanded the

*Address all correspondence to: Kristen A. Wangerin, E-mail: wangerin@uw.edu

Fig. 1 Synthetic lesions are inserted into the raw patient data after accounting for all PET system physics effects. Examples of a lesion-absent (top right) and lesion-present (bottom right) image reconstruction are shown for a lesion contrast of 4.0.

previous study by using the CHO model to compare detectability as a function of lesion location, size, and contrast, as well as for algorithm parameters and time-of-flight (TOF) and non-TOF data. For the remainder of this paper, the PL-RDP algorithm is referred to as simply PL.

2 Methods

2.1 Image Generation

We obtained 81 TOF list-mode patient datasets using¹⁸ F-FDG acquired on a Discovery PET/CT 690 scanner (General Electric, Waukesha, Wisconsin) at the University Hospital of Zürich, Switzerland. The TOF timing resolution of the system was 544 ps.³⁰ From each list-mode dataset, two projection datasets were generated for both TOF and non-TOF imaging scenarios by binning alternating events into the two sinograms and histogramming with nonoverlapping time windows. Dividing the data provided independent and identically distributed (i.i.d.) noise realizations for the lesion-absent and lesion-present images.

Synthetic spherical lesions with a diameter of 10, 15, and 20 mm were generated in both the liver and lung regions using the procedure shown in Fig. 1. The synthetic lesions were then forward projected through the scanner model, where patient effects (attenuation) and scanner effects (resolution, geometric efficiency, and detector efficiency variations) were considered. The synthetic lesion emission projection data were weighted based on the desired contrast, Poisson noise was added, and the noise-added lesion projection data were added to patient emission projection data to form the hybrid emission projection data. The lesion contrasts for the liver and lung lesions studied are shown in Table 1. The lesion contrast was

defined as $[(\text{lesion activity}) - (\text{local background activity})] / (\text{local background activity})$. The contrasts were chosen so that the lesions were near the limit of detectability.

The two different sinogram noise realizations were used for the lesion-absent and lesion-present cases. Both sinograms were reconstructed using both OSEM and PL algorithms for TOF and non-TOF cases including point spread function (PSF) modeling.³¹ The reconstructed voxel size was $3.65 \times 3.65 \times 3.27$ mm.

We evaluated and compared lesion detectability using clinically representative parameters for each algorithm. The OSEM algorithm was run with 24 subsets and three iterations and postfiltered with a $[1/6, 4/6, 1/6]$ three-point axial filter and 6 mm transaxial filter. PL was implemented using the BSREM algorithm^{16,17} as in Ahn et al.¹⁴ The PL algorithm was run with a penalty strength parameter β of 250, 350, and 450 and an RDP edge-preserving parameter $\gamma = 2$. The RDP is defined as

$$\beta R(x) = \beta \sum_{j,k:\text{neighbors}} \alpha_{jk} \frac{(x_j - x_k)^2}{(x_j + x_k) + \gamma |x_j - x_k|} \quad (1)$$

and is calculated between each pixel j and its k neighbor pixels. The weights α_{jk} were used as in Ahn et al.¹⁴ The range of the β parameters was found to be clinically relevant in Ahn et al.,¹⁴ and the edge-preservation parameter $\gamma = 2$ was previously chosen as a reasonable trade-off between visual image quality and lesion quantitation accuracy.^{14,32} No postfiltering was applied to PL images. We found the image roughnesses from the first three OSEM iterations with postfiltering to most closely match the image roughnesses from the three PL β s, which allowed us to compare the lesion detectability between algorithms based on matched image noise.

Considering all of the parameters (81 patients, TOF/non-TOF, liver/lung, lesion absent/present, OSEM/PL, three lesion sizes, three reconstruction parameters with different resolution and noise trade-offs, and three lesion contrasts), a total of 46,656 images were generated and analyzed. Examples of lesion-absent and lesion-present image patches are shown in Fig. 2. Detailed results are displayed for two iterations of OSEM and a β value of 350 for PL, as these are the most commonly applied clinical parameters.

Table 1 Lesion contrasts for the liver and lung lesion locations and three spherical lesion diameters.

	10 mm			15 mm			20 mm		
Liver	0.75	1.0	1.25	0.25	0.5	0.75	0.25	0.5	0.75
Lung	1.25	1.75	2.25	0.75	1.0	1.25	0.5	0.75	1.0

Fig. 2 Lesion-absent and lesion-present image patches reconstructed using OSEM and PL algorithms. The liver and lung lesion examples are 20 mm in diameter with a contrast of 1 and 2.25, respectively. OSEM images are shown for two iterations with postfiltering, and PL images are shown for $\beta = 350$.

2.2 Model Observer Study

The detectability was evaluated in a signal known statistically and background known statistically task. For each lesion size and contrast, although the true lesion signal to local background was kept constant among images, the reconstructed lesion signal magnitude as well as the true lesion signal magnitude varied across patients. Therefore, the signal magnitude was only statistically known. Similarly, the background of every patient was different and only statistically known.

We applied a two-dimensional CHO model using two types of radially symmetric channel models to the reconstructed images using the center image slice through each lesion. The channel models were four constant- Q channels³³ and three sparse difference-of-Gaussian (DOG) channels.³⁴ The channels, U , were represented in a n by (30×30) matrix, where n is the number of channels, and the columns described each channel's impulse response centered on the tumor location. The constant- Q channels were defined by the frequency intervals $[\rho_c Q^{n-1}, \rho_c Q^n]$, where $\rho_c = 1/32$, $Q = 2$, and $n = 1:4$. The DOG channels were defined by

$$u_n(\rho) = \exp\left[-\frac{1}{2}\left(\frac{\|\rho\|}{Q\sigma_n}\right)^2\right] - \exp\left[-\frac{1}{2}\left(\frac{\|\rho\|}{\sigma_n}\right)^2\right], \quad (2)$$

where ρ is the frequency vector and $Q = 2$. The channel standard deviation, σ_c , was $\alpha^c \sigma_0$ with $\sigma_0 = 0.015$ and $\alpha = 2$. The channels, as shown in Fig. 3, were applied individually to the 81 lesion-absent and lesion-present images as Uf to extract the $n \times 81$ feature vectors, v_0 and v_1 .

The lesion template was derived from an ROI using all 81 lesion-absent and 81 lesion-present images for a given parameter set (lesion size, location, contrast, and reconstruction parameters), described by

$$w = K^{-1}U(\bar{f}_1 - \bar{f}_0), \quad (3)$$

where K is the $n \times n$ unconditional covariance matrix of the channel output and $\bar{f}_0$ and $\bar{f}_1$ are the mean of the lesion-absent images and lesion-present images, respectively.³³ The covariance matrix, K , was calculated using

$$K = \frac{1}{2}(K_0 + K_1), \quad (4)$$

where K_0 and K_1 are the conditional covariance matrices of the channel output for the lesion-absent and lesion-present cases, respectively. The lesion was centered in this 30×30 pixel ROI.

The test statistics, λ_0 and λ_1 for the lesion-absent and lesion-present images, were calculated for each image using

$$\lambda_0 = w^T v_0 \quad \text{and} \quad \lambda_1 = w^T v_1. \quad (5)$$

The more similar the image was to the template, the higher the output test statistic was. The SNR, often called the detectability index, d' , was calculated using

$$d' = \text{SNR} = \frac{\bar{\lambda}_1 - \bar{\lambda}_0}{\sqrt{\frac{1}{2}(\sigma_1^2 + \sigma_0^2)}}, \quad (6)$$

where σ_0^2 and σ_1^2 are the variances of the test statistics for lesion-absent and lesion-present images, respectively. The standard error of d' was calculated using propagation of error of the SNR expression.³⁵ Detectability was also compared using d_A , which is obtained from the area under the receiving operator characteristic curve (AUC) using the transformation

$$d_A = 2 \operatorname{erf}^{-1}(2 \text{AUC} - 1). \quad (7)$$

A more detailed explanation of Eqs. (5)–(7) can be found in Abbey et al.³⁵

The detectabilities were then compared as a function of contrast as well as mean image roughness (IR) in the liver,³⁶ defined as the coefficient of variation of the pixel values as

$$\text{IR} = \frac{\sqrt{\frac{1}{I-1} \sum_{i \in \text{ROI}} (f_i - m)^2}}{m}, \quad (8)$$

where I is the number of pixels in the ROI, f_i is the image value of pixel i , and m is the mean pixel value in the ROI.

We applied our CHO model to OSEM and PL reconstructed images with varying simulation parameters, and studied the

Fig. 3 The (a) four constant- Q and (b) three DOG channel models applied in the CHO model.

Fig. 4 CHO SNR as a function of contrast for liver and lung lesion locations for the three lesion sizes and non-TOF and TOF reconstructions. The p -values indicate the statistical significance of the difference in CHO SNR between non-TOF OSEM and PL and between TOF OSEM and PL. Statistically significant p -values (<0.05) are highlighted in orange.

impact of lesion location, lesion size, lesion contrast, TOF, PSF, transaxial and axial postfiltering, and OSEM iteration number or PL β value on detectability.

2.3 Statistical Analysis

The significance of the differences in lesion detectability between images reconstructed using the OSEM and PL algorithms were evaluated using a resubstitution bootstrapping method.³⁷ To generate 1000 samples, each of which consisted of 81 patient image pairs, for each combination of parameters (TOF/non-TOF, liver/lung, OSEM/PL, lesion size, lesion contrast, and reconstruction parameter such as iteration number for OSEM and β parameter for PL), images were randomly sampled from the 81 patient images; for each sampling, some patients were selected one or more times and some patients were not selected at all. Because images were reconstructed from the same sinogram noise realization for both algorithms, the same random sampling was used for both algorithms to maintain that correlation. Each random sample of lesion detectability for OSEM was subtracted from that for PL, and the results were histogrammed to form a difference distribution. If such a difference distribution centered on zero, e.g., it would indicate no difference in lesion detectabilities between the two algorithms. Confidence intervals were calculated using the standard deviation of the difference bootstrap distribution. The p -value for each case was determined by calculating the fraction of the difference distribution that crossed zero, and multiplying that fraction by two to represent a two-tailed test.

3 Results

We only present results for the constant- Q channel model, as the results using the DOG channel model showed no substantial

differences to those using the constant- Q channel model (data not shown). Results are also presented only for detectabilities calculated using the SNR, and the detectabilities calculated using the AUC also showed no substantial difference compared to detectabilities calculated using the SNR (data not shown).

OSEM and PL SNR detectability results are compared as a function of contrast in Fig. 4 for lesion location, lesion size, and TOF and non-TOF image reconstruction. Results are shown for clinically relevant reconstruction parameters of two iterations of OSEM with a $[1/6,4/6,1/6]$ three-point axial postfilter and 6 mm transaxial postfilter and for PL with a β value of 350.¹⁴ Detectabilities were higher in the liver than the lung for matched contrast. Detectability also increased with increasing contrast and lesion size, as well as for TOF compared to non-TOF, as expected. Detectability using PL showed more of a dependence on lesion size than did detectability using OSEM; e.g., for a contrast of 0.75 in the liver, the detectability was 186% higher for the 20 mm lesion compared to the 10 mm for PL and 164% higher for OSEM for the TOF case. The percent improvement in TOF versus non-TOF SNR increased with decreasing lesion size and contrast, as shown in Fig. 5. TOF also provided more benefit in the liver than the lung, increasing detectability by an average of 22.0% compared to 11.8%, respectively, when averaging over algorithm, lesion size, and contrast. Modeling PSF also improved lesion detectability (data not shown), as other studies have found.^{38,39}

The 95% confidence intervals and p -values in Fig. 4 were determined from the bootstrapping statistical analysis. Example bootstrap distributions using 1000 samples are shown in Fig. 6(a). The same patient sinograms were used for both algorithms, resulting in a significant correlation between the detectabilities, as seen in Fig. 6(b). We were interested in the performance difference beyond these correlations, so we

Fig. 5 Percent improvement of TOF over non-TOF image reconstruction as a function of contrast for both reconstruction algorithms and lesion locations.

Fig. 6 (a) Bootstrap distributions of CHO SNR for OSEM (o's) and PL (x's) for liver and lung lesion locations. A Gaussian model is fit to each distribution. (b) CHO SNR for PL versus OSEM for each of the 1000 bootstrap samples, showing the detectability correlation between algorithms. Data points above or below the line of identity indicate improved performance of PL or OSEM, respectively. (c) PL and OSEM difference distribution, where statistical significance increases the further away the distribution is from zero.

evaluated the difference in the bootstrap distributions. The distribution of the performance difference between algorithms is shown in Fig. 6(c). PL detectability performance improved for increasing lesion size and contrast. PL SNR was statistically significantly higher than OSEM for 20 mm lesions in the liver when contrast was ≥ 0.5 ($p < 0.05$ two-sided test), and TOF PL SNR was statistically significantly higher than TOF OSEM for 15 mm lesions in the liver with contrast ≥ 0.5 and 20 mm lesions with contrast ≥ 0.25 . For all other cases, there was no statistically significant difference between PL and OSEM ($p > 0.05$).

Detectability as a function of image noise as measured by IR is shown in Fig. 7, where IR is impacted by PL β value or OSEM iteration number and postfiltering. The impact of the OSEM axial and transaxial filters is shown separately. Smoother images were seen to be favorable for detection in the liver for both PL and OSEM; detectability increased with decreasing image noise. Lung detectabilities showed less dependence on IR. Transaxial postfiltering with a 6 mm full-width at half-maximum Gaussian produced smoother images with generally

slightly higher detectability compared to no postfiltering. Axial postfiltering also produced smoother OSEM images while improving detectability.

4 Discussion

We demonstrated in previous studies^{14,32} that our convergent PL algorithm improved quantitation accuracy of lesion SUV compared to OSEM at matched IR, particularly in cold background regions such as the lung, due to the edge-preserving property of RDP and the slow convergence rates of OSEM in cold regions. While the properties of the RDP were favorable for quantitation and higher contrast lesions, we questioned how properties of the RDP would impact lesion detectability for low contrasts. In particular, we were concerned that the RDP might smooth out low-contrast lesions, degrading detectability. Therefore, we performed this study to evaluate lesion detectability over a range of lesion and reconstruction algorithm parameters, including lesion location, size, and contrast, TOF/non-TOF OSEM and PL image reconstruction, and image postfiltering.

Fig. 7 CHO SNR as a function of IR in the liver, and impact of transaxial and axial filtering applied to OSEM images on detectability. Transaxial and axial filtering results are shown independently. Representative results are shown for TOF reconstruction and one contrast. Each OSEM curve has three markers representing three, two, and one iterations from left to right, and each PL curve has three markers representing $\beta = 450, 350$, and 250 from left to right.

Our goal was not to develop a model observer but, rather, to model the human observer results. It is well known in the community that CHO more consistently predicts human performance for a variety of correlations.^{40,41} The CHO observer has become a popular model observer within the nuclear medicine community due to its flexible and predictive characteristics.⁴² Further, others have used the CHO model for both TOF/non-TOF and without and with PSF modeling.^{38,39} The choice of model observer parameters, such as the channel model, is important, and there are many channel models that have been extensively evaluated in other works.⁴³ In our previous study,²⁹ we evaluated this model observer using constant- Q channels and found good agreement with humans. In addition to the constant- Q model, we conducted this study using a DOG channel model, and we found the same ranking between PL and OSEM algorithms. Although it is not uncommon for model observer results to outperform human results,⁴⁴ this difference was small in our previous study, potentially due to the template estimation (observer training),³⁷ ROI, or image noise properties. While internal noise can be added to reduce the model observer performance, we did not find this step necessary in our previous study.

We chose lesion contrasts so that the lesions were at the limit of detectability. Lesion recovery is lower for lesions in low-activity background regions, such as the lungs, than for those in high-activity background regions such as livers. This is due to slow convergence rate of OSEM in the lungs and more smoothing in the lungs applied by the penalty function, RDP, used in PL. As a result, the lesion detectability is also lower in the lungs, and we used higher contrasts for lung lesions than for liver lesions.

We chose the SNR as our detectability metric so that we could directly evaluate the impact of signal means and variances. We evaluated the model observer test statistics and found that they are generally Gaussian, indicating that the performance using the SNR and AUC should be similar. To further confirm this, we calculated the detectability performance of PL and OSEM using the AOC, and we found similar performances between the algorithms using the two metrics.

First considering lesion location, lesion detectability was higher in the liver than the lung at equal contrast (Fig. 4), due to higher contrast recovery in the liver compared to the lung for an equivalent number of iterative updates in OSEM and for an equivalent β parameter in PL. It is known that the lung region converges more slowly,¹⁴ which was likely the reason for the lower detectability in the lung for OSEM; to control image noise, we ran the algorithm for only three iterations with 24 subsets and postfiltered the image. For PL, lower detectability in the lung was likely due to the RDP applying more smoothing in this low-activity region.

Considering TOF versus non-TOF image reconstruction (Fig. 5), detectability was improved by TOF for all cases, with more improvement for smaller lesions and lesion location in the liver compared to the lung. The trend of detectability improvement as a function of lesion size and location suggests varying impact of TOF reconstruction on the underlying signal and noise properties. When considering the impact of postfiltering in OSEM on detectability (Fig. 7), the detectabilities at matched IR improved with axial filtering but decreased with transaxial filtering for the filters used in the comparison. In the axial case, the benefit of reduced noise likely drove the improvement in CHO SNR, whereas for the transaxial case,

Fig. 8 CRC for a wide range of inserted lesion contrasts in the liver and lung for the three lesion sizes, showing the improved performance of the RDP with increasing contrast, particularly for the 10-mm sized lesion.

the decrease in signal from the smoothing outweighed the improvement in noise.

In this lesion detection study, we found PL performed better in the liver than in the lung in comparison to OSEM. In contrast, our previous quantitation study¹⁴ showed that PL performed better in the lung than the liver in comparison to OSEM. To investigate this difference, we calculated the contrast recovery coefficient (CRC) over a wide range of lesion contrasts as shown in Fig. 8. The CRC showed greater improvement for PL with increasing contrast, particularly for the smallest lesion size. The difference in algorithm performance for low- versus high-contrast lesion properties illustrates the contrast dependence of the RDP and explains the different results seen in the detection versus quantitation studies. These results highlight the important difference between detection and quantitation tasks, which should be well defined and considered in study design.

Considering lesion size, PL showed greater relative performance improvement with increasing lesion size, a trend also seen in Qi.⁴⁵ PL performance compared to OSEM was lowest for the 10 mm lesion, though the difference between algorithms was not statistically significant. For the 15 and 20 mm lesion sizes and certain lesion contrasts, PL outperformed OSEM. For the small lesions, it is likely that the edge-preserving nature of the prior is less sensitive. The detectability of lesions smaller than 10 mm should be investigated in a future study, especially for higher resolution scanner that will improve contrast recovery.

Finally, we considered detectability as a function IR (Fig. 7), which is a function of OSEM iteration and postfiltering or PL β value. The detectability in the liver increased with decreasing IR (lower OSEM iteration or higher PL β), a trend observed by many previous studies.^{46–49} Detectability in the lung was less dependent on IR, and for the 10 mm lesion, PL detectability decreased with decreasing IR, a trend also seen in previous studies.^{50,38} These opposite trends illustrate the dependence of

lesion detectability on different noise correlations that arise with different reconstruction algorithms and parameters. Further, these trends indicate that OSEM postfiltering or PL β should be applied cautiously for smaller lesion sizes and lower contrasts, due to the potential to blur the lesion into the background.

In a previous relevant study, Qi⁵¹ compared lesion detectability for PL with quadratic penalties and edge-preserving non-quadratic penalties, such as Huber and Geman-McClure, using a CHO, and showed the edge-preserving property of the nonquadratic penalties did not improve lesion detectability. In another similar study, Nuyts et al.⁴⁹ found that detectability for PL with RDP using a small or large edge-preserving parameter ($\gamma = 0$ or $\gamma = 200$) was superior to postsmoothed maximum likelihood expectation maximization, and when $\gamma = 10$ or $\gamma = 20$, RDP-based PL yielded reduced detectability.

Prior to our previous preliminary study and this current work, there had not yet been a comparison of the detectability of the RDP-based PL to the standard clinical algorithm, OSEM. Additionally, our choice of the RDP parameter, $\gamma = 2$, which resulted in superior lesion quantitation accuracy compared to OSEM,^{14,18–23} was different from $\gamma = 0$ and $\gamma = 10$, which showed very different performances from each other in the study of Nuyts et al.⁴⁹ Due to these differences, it was not straightforward to extrapolate previous study results to the performance of our PL algorithm. The results in this study indicate that, for clinically relevant count density and range of reconstruction parameters and lesion properties considered, PL does not decrease lesion detectability compared to OSEM. PL maintained equivalent detectability in the lung and improved detectability in the liver for some cases.

There were some limitations in this study. The model observers were applied to a single axial slice of cropped PET images (30×30 pixel ROIs). We did not explore larger volumes or location uncertainty. We averaged the results over all patients and did not perform analyses or separate the results based on

patient characteristics. Finally, we emphasize that we performed this study using clinically relevant parameters for each algorithm. We acknowledge that the algorithm parameters, such as OSEM iteration number and transaxial and axial filtering, could be separately optimized in the task of detection, and that the results could be different from those found in this study. However, it is not clear if it makes sense to optimize the CHO SNR solely; Fig. 7 shows one iteration of OSEM results in better SNR than two and three iterations, which contradicts the general belief that one iteration yields overly smooth images. This is due, at least in part, to the multiple tasks images are used for (e.g., detection, localization, characterization, and estimation), and may also be due to the limitations with the model observer we used and perhaps the over-simplified detection task (e.g., known lesion location) we considered. Future model observer studies could be conducted to evaluate a broader range of both OSEM and PL algorithm parameters for optimization of lesion detectability.

5 Conclusions

We evaluated lesion detectability in images reconstructed using a convergent PL algorithm with RDP and compared the results to detectability in images reconstructed using the clinical standard OSEM algorithm. A CHO model analysis showed that at matched IR, PL-RDP achieved matched, or improved lesion detectability compared to OSEM for the clinically relevant parameter space investigated.

Acknowledgments

We thank the University Hospital of Zürich for the patient data, and Dan Hippe for his insightful guidance and discussion on the statistical analyses. This work was partially supported by NIH Grants Nos. R01CA160253 and U01CA148131 and a research grant from GE Healthcare.

References

1. H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imaging* **13**(4), 601–609 (1994).
2. J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.* **51**, R541–R578 (2006).
3. J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography," *IEEE Trans. Image Proc.* **5**(3), 493–506 (1996).
4. J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *IEEE Trans. Image Proc.* **5**(9), 1346–1358 (1996).
5. J. Qi and R. M. Leahy, "Resolution and noise properties of map reconstruction for fully 3-D PET," *IEEE Trans. Med. Imaging* **19**(5), 493–506 (2000).
6. S. Ahn and R. M. Leahy, "Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET," *IEEE Trans. Med. Imaging* **27**(3), 413–424 (2008).
7. J. Dutta, S. Ahn, and Q. Li, "Quantitative statistical methods for image quality assessment," *Theranostics* **3**(10), 741–756 (2013).
8. J. Qi, "Analysis of lesion detectability in Bayesian emission reconstruction with nonstationary object variability," *IEEE Trans. Med. Imaging* **23**(3), 321–329 (2004).
9. J. Qi and R. H. Huesman, "Penalized maximum-likelihood image reconstruction for lesion detection," *Phys. Med. Biol.* **51**(16), 4017–4029 (2006).
10. J. Qi and R. H. Huesman, "Theoretical study of penalized-likelihood image reconstruction for region of interest quantification," *IEEE Trans. Med. Imaging* **25**(5), 640–648 (2006).
11. S. Geman and D. E. McClure, "Bayesian image analysis: an application to single photon emission tomography," *Proc. Am. Stat. Assoc.* 12–18 (1985).
12. W. Cholewicki, F. Hermansen, and S. B. Hansen, "Noise reduction and convergence of Bayesian algorithms with blobs based on the Huber function and median root prior," *Phys. Med. Biol.* **49**, 4717–4730 (2004).
13. E. Asma and R. M. Manjeshwar, "Analysis of organ uniformity in low count density penalized likelihood PET images," in *IEEE Nuclear Science Symp. Conf. Record (NSS 2007)*, pp. 4426–4432 (2007).
14. S. Ahn et al., "Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET," *Phys. Med. Biol.* **60**, 5733–5751 (2015).
15. J. Nuyts et al., "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," *IEEE Trans. Nucl. Sci.* **49**(1), 56–60 (2002).
16. A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," *IEEE Trans. Med. Imaging* **20**, 280–288 (2001).
17. S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Trans. Med. Imaging* **22**, 613–626 (2003).
18. E. J. Teoh et al., "Novel penalised likelihood reconstruction of PET in the assessment of histologically verified small pulmonary nodules," *Eur. Radiol.* **26**(2), 576–584 (2015).
19. E. J. Teoh et al., "Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q. Clear on an LYSO PET/CT system," *J. Nucl. Med.* **56**(9), 1447–1452 (2015).
20. H. Ma et al., "Clinical evaluation of penalized likelihood reconstruction in whole-body PET studies," in *Annual Congress of the European Association of Nuclear Medicine* (2013).
21. B.-R. Sah et al., "Clinical evaluation of a new block sequential regularized expectation maximization (BSREM) reconstruction algorithm in PET/CT studies," *J. Nucl. Med.* **55**(S1), 2097 (2014).
22. S. Passalacqua et al., "Qualitative and quantitative evaluation of regularized PET image reconstruction," *J. Nucl. Med.* **55**(S1), 579 (2014).
23. J. Lantos et al., "Standard OSEM vs. regularized PET image reconstruction: qualitative and semi-quantitative comparison," *J. Nucl. Med.* **56**(S3), 1805 (2015).
24. N. Parvizia et al., "Does a novel penalized likelihood reconstruction of 18F-FDG PET-CT improve signal-to-background in colorectal liver metastases?," *Eur. J. Radiol.* **84**(10), 1873–1878 (2015).
25. S. Y. Kang et al., "Increased detectability of small lesions by BSREM reconstruction of PET in breast cancer," *J. Nucl. Med.* **56**(S3), 1832 (2015).
26. K. A. Wangerin et al., "Improving lesion detectability in PET imaging with a penalized likelihood reconstruction algorithm," *Proc. SPIE* **9416**, 94160W (2015).
27. C. K. Abbey, M. P. Eckstein, and F. O. Bochud, "Estimation of human-observer templates in two-alternative forced-choice experiments," *Proc. SPIE* **3663**, 284–295 (1999).
28. A. E. Burgess, "Comparison of receiver operating characteristic and forced choice observer performance measurement methods," *Med. Phys.* **22**(5), 643–655 (1995).
29. J. Yao and H. H. Barrett, "Predicting human performance by a channelized Hotelling observer model," *Proc. SPIE* **1768**, 161–168 (1992).
30. V. Bettinardi et al., "Physical performance of the new hybrid PET/CT Discovery-690," *Med. Phys.* **38**(10), 5394–5411 (2011).
31. A. M. Alessio et al., "Application and evaluation of a measured spatially variant system model for pet image reconstruction," *IEEE Trans. Med. Imaging* **29**, 938–949 (2010).
32. E. Asma et al., "Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images," in *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 4062–4066 (2012).
33. H. C. Gifford et al., "Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J. Nucl. Med.* **41**(3), 514–521 (2000).
34. H. C. Gifford et al., "Evaluation of multiclass model observers in PET LROC studies," *IEEE Trans. Nucl. Sci.* **54**, 116–123 (2007).
35. C. K. Abbey, H. H. Barrett, and M. P. Eckstein, "Practical issues and methodology in assessment of image quality using model observers," *Proc. SPIE* **3032**, 182–194 (1997).

36. S. Tong, A. M. Alessio, and P. E. Kinahan, "Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation," *Phys. Med. Biol.* **55**(5), 1453–1473 (2010).
37. A. Wunderlich et al., "Exact confidence intervals for channelized Hotelling observer performance in image quality studies," *IEEE Trans. Med. Imaging* **34**(2), 453–464 (2015).
38. D. J. Kadrmas et al., "Impact of time-of-flight on PET tumor detection," *J. Nucl. Med.* **50**(8), 1315–1323 (2009).
39. J. Schaefferkoetter et al., "Clinical impact of time-of-flight and point response modeling in pet reconstructions: a lesion detection study," *Phys. Med. Biol.* **58**(5), 1465–1478 (2013).
40. H. H. Barrett et al., "Model observers for assessment of image quality," *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9758–9765 (1993).
41. K. Li, J. Garrett, and H. G. Chen, "Correlation between human observer performance and model observer performance in differential phase contrast CT," *Med. Phys.* **40**, 111905 (2013).
42. J. G. Brankov, "Evaluation of the channelized Hotelling observer with an internal-noise model in a train-test paradigm for cardiac SPECT defect detection," *Phys. Med. Biol.* **58**, 7159–7182 (2013).
43. X. He and S. Park, "Model observers in medical imaging research," *Theranostics* **3**(10), 774–786 (2013).
44. Y. Zhang, B. T. Pham, and M. P. Eckstein, "The effect of nonlinear human visual system components on performance of a channelized Hotelling observer in structured backgrounds," *IEEE Trans. Med. Imaging* **25**(10), 1348–1362 (2006).
45. J. Qi, "Investigation of lesion detection in map reconstruction with non-Gaussian priors," in *Nuclear Science Symp. Conf. Record 2005 IEEE*, pp. 1704–1708 (2005).
46. S. Surti and J. S. Karp, "Experimental evaluation of a simple lesion detection task with time-of-flight PET," *Phys. Med. Biol.* **54**(2), 373–384 (2009).
47. V. Y. Panin, M. Chen, and M. E. Casey, "Numerical observer evaluation of 3D statistical PET reconstruction method with system response modeling," in *2005 IEEE Nuclear Science Symp. Conf. Record*, pp. 2143–2147 (2005).
48. A. Petschke and P. J. La Rivière, "Comparison of photoacoustic image reconstruction algorithms using the channelized Hotelling observer," *J. Biomed. Opt.* **18**(2), 026009 (2013).
49. J. Nuyts et al., "Performance of map reconstruction for hot lesion detection in whole-body PET/CT: an evaluation with human and numerical observers," *IEEE Trans. Med. Imaging* **28**(1), 67–73 (2009).
50. G. El Fahkri et al., "Improvement in lesion detection with whole-body oncologic time-of-flight PET," *J. Nucl. Med.* **52**(3), 347–353 (2011).
51. J. Qi, "Comparison of lesion detection and quantification in MAP reconstruction with Gaussian and non-Gaussian priors," *Int. J. Biomed. Imaging* **2006**, 1–10 (2006).

Kristen A. Wangerin received her bachelor's and master's degrees in nuclear engineering from Purdue University and Rensselaer Polytechnic Institute, respectively. She spent seven years at General Electric's Global Research Center in upstate NY contributing to a wide variety of medical imaging technologies. She recently received her PhD in bioengineering as a member of the Imaging Research Laboratory at the University of Washington.

Sangtae Ahn is a research scientist at General Electric's Global Research, Niskayuna, New York, USA. He received his BS and MS degrees in mechanical engineering at Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea, and his MS and PhD degrees in electrical engineering systems from the University of Michigan, Ann Arbor, Michigan, USA. He currently works on PET image reconstruction algorithms and data correction for PET/CT and PET/MR scanners.

Biographies for the other authors are not available.