

# A non-parametric bootstrap approach for analysing the statistical properties of SPECT and PET images

**Irène Buvat**

U494 INSERM, CHU Pitié-Salpêtrière, 91 Boulevard de l'Hôpital, 75 634 Paris, Cedex 13, France

E-mail: [buvat@imed.jussieu.fr](mailto:buvat@imed.jussieu.fr)

Received 5 December 2001

Published 2 May 2002

Online at [stacks.iop.org/PMB/47/1761](http://stacks.iop.org/PMB/47/1761)

## Abstract

Knowledge of the statistical properties of reconstructed single photon emission computed tomography (SPECT) and positron emission tomography (PET) images would be helpful for optimizing acquisition and image processing protocols. We describe a non-parametric bootstrap approach to accurately estimate the statistical properties of SPECT or PET images whatever the noise properties in the projections and the reconstruction algorithm. Using analytical simulations and real PET data, this method is shown to accurately predict the statistical properties, including the variance and covariance, of reconstructed pixel values for both linear (filtered backprojection) and non-linear (ordered subset expectation maximization) reconstruction algorithms.

## 1. Introduction

A general method for characterizing the noise properties (amplitude, spatial distribution and correlation) of single photon emission computed tomography (SPECT) and positron emission tomography (PET) images would be useful to optimize acquisition and processing protocols, by studying the effect of factors such as the injected dose, the reconstruction algorithm or various corrections (e.g., scatter, attenuation, random). Noise estimates in reconstructed images could also be introduced into quantitative data analysis to account for the uncertainty of region of interest (ROI) values (Huesman and Mazoyer 1987) or when using algorithmic observers for evaluation purposes (Barrett 1990, Myers *et al* 1985). For images reconstructed with filtered backprojection (FBP) in emission tomography, analytical derivations have been proposed to predict the variance of pixel or ROI values given the statistical properties of the projections (e.g., Alpert *et al* 1982, Budinger *et al* 1977, 1978, Carson *et al* 1993, Huesman 1984, Palmer *et al* 1986, Wilson *et al* 1994). For non-linear reconstruction algorithms such as maximum likelihood expectation maximization (MLEM) and block-iterative reconstruction algorithms, approximations are required to analytically

deduce the statistical properties of the reconstructed images from those of the projections (Barrett *et al* 1994, Fessler 1996, Kadmas *et al* 1999, Qi and Leahy 2000, Soares *et al* 2000, Wang and Gindhi 1997, Wilson *et al* 1994). For FBP as well as for iterative reconstruction, many approaches assume Poisson noise in the projections. As some processing steps (e.g., scatter or random correction, interpolation) make the noise non-Poisson, their statistical effects have sometimes been incorporated in the analysis of error propagation, especially in PET, using simplifying assumptions (e.g., Carson *et al* 1993, Qi and Leahy 2000). A bootstrap approach has been suggested to investigate the statistical properties of PET images but it also assumed that the recorded counts were Poisson distributed (Haynor and Woods 1989).

Numerical simulations or empirical measurements of several replicated sinograms have been used for tracking the propagation of statistical errors in specific configurations (e.g., Alpert *et al* 1991, Budinger *et al* 1978, Riddell *et al* 2001, Wilson and Tsui 1993). However, for realistic simulated data such as Monte Carlo simulations or for real data, the number of replicates has to be kept small from a practical point of view, while several hundred replicates would be needed for achieving a good statistical precision. Furthermore, for clinical data, usually only one realization is available.

In this study, we propose a non-parametric bootstrap method to characterize the statistical properties of SPECT or PET images from a small number of replicates, or even from a single acquisition, for any linear or non-linear reconstruction method and any type of noise in the projections. The non-parametric nature of the proposed bootstrap method makes it possible to study the statistical properties of SPECT and PET images when the sinograms used for the reconstruction are no longer Poisson distributed due to preprocessing steps.

## 2. Theory

### 2.1. The bootstrap approach

The bootstrap approach is a computer-based statistical method for determining the accuracy of a statistic  $\theta$  (e.g., median) estimated from experimental data (Efron and Tibshirani 1993). It requires an experimental sample  $\mathbf{x} = (x_1, \dots, x_N)$  whose empirical distribution estimates an unknown distribution  $F$ . In this sample, each measurement  $x_i$  is considered as an independent random realization of the variable that follows distribution  $F$ . Under its simplest form, the bootstrap uses what is called a plug-in principle:

- Given the empirical sample  $\mathbf{x} = (x_1, \dots, x_N)$ , draw  $B$  independent bootstrap samples  $\mathbf{x}^{b*} = (x_1^{b*}, \dots, x_N^{b*})$  of  $N$  elements  $x_i^{b*}$  each. Each element  $x_i^{b*}$  is obtained by randomly drawing with replacement one element  $x_i$  from the original empirical sample  $\mathbf{x}$ . Note that the number of elements in each bootstrap sample is identical to the number of elements in the original empirical sample.
- For each bootstrap sample  $\mathbf{x}^{b*}$ , calculate the statistic of interest  $\theta(\mathbf{x}^{b*})$ , which is called a bootstrap replication of  $\theta$ .
- The set of bootstrap replications  $\{\theta(\mathbf{x}^{b*})\}_{b=1, B}$  yields the bootstrap distribution of  $\theta$ , from which the statistical behaviour of  $\theta$  can be inferred. For instance, the bootstrap variance  $M_2$  (moment of order 2) of  $\theta$  is

$$M_2 = \left[ \sum_{b=1}^B (\theta(\mathbf{x}^{b*}) - M_1)^2 / (B - 1) \right] \quad (1)$$

where  $M_1 = \sum_{b=1}^B \theta(\mathbf{x}^{b*}) / B$  is the mean of  $\theta$  over the  $B$  bootstrap replications.

For more details about the bootstrap approach, we refer the reader to Efron and Tibshirani (1993).

## 2.2. Using the bootstrap concept to generate bootstrap sub-sinograms

When characterizing the statistical properties of a specific SPECT or PET image or of a parameter derived from it, the statistic of interest,  $\theta$ , is the reconstructed image itself. Using the bootstrap approach therefore requires obtaining  $B$  bootstrap reconstructed images,  $\theta(\mathbf{x}^{b*})$ , to derive the statistical distribution of the reconstructed image or of the parameter derived from it. These  $B$  bootstrap reconstructed images can be obtained from  $B$  bootstrap sinograms,  $\mathbf{x}^{b*}$ . The problem therefore lies in generating bootstrap sinograms representative of sinograms that would be obtained if one could repeat an acquisition a large number of times. In the conventional bootstrap approach, the trick is to consider an experimental sample,  $\mathbf{x} = (x_1, \dots, x_N)$ , whose empirical distribution estimates the unknown distribution  $F$  of  $x_i$ , to generate bootstrap samples (cf section 2.1). Similarly, a sample of empirical sinograms, whose empirical distribution will estimate the unknown distribution of the sinograms, is required to generate bootstrap sinograms. Such a sample can be obtained by collecting the data in a gated fashion, so as to distribute the  $C$  counts detected during the whole acquisition over  $N$  gates to obtain  $N$  statistically independent realizations of sinograms of about  $C/N$  counts (Riddell *et al* 2001). These realizations are called sub-sinograms in the following as they include  $N$  times fewer counts than the total number of acquired counts. Alternatively, the data can be acquired in list-mode and reformatted into  $N$  statistically independent sub-sinograms. This set of sub-sinograms gives a rough estimate of the statistical distribution of the sub-sinograms that will be taken advantage of in the non-parametric bootstrap approach.

Each empirical sub-sinogram of about  $C/N$  counts is an  $(A, K)$  matrix, where  $A$  and  $K$  are the numbers of projection angles and acquisition bins, respectively. A bootstrap sub-sinogram is generated by randomly drawing each row  $j$  (corresponding to a specific angle,  $j = 1, \dots, A$ ) among the  $N$  realizations of row  $j$  given by the  $N$  empirical sub-sinograms. Rows  $j$  and  $j'$  of a bootstrap sub-sinogram can thus come from different empirical sub-sinograms. Drawing rows as a whole instead of pixel values is a key step to properly account for the noise correlation potentially present within a row without having to make any assumption about the type of noise correlation that may exist. On the other hand, as noise is not correlated from one row to another, row  $j'$  can be drawn independently of row  $j$ . Using this method, any number,  $B$ , of bootstrap sub-sinograms can be randomly generated among the  $N^A$  possible combinations.

## 2.3. Deriving the statistical properties of reconstructed SPECT and PET images

Bootstrap sub-sinograms can be readily used to derive the statistical properties of the corresponding sub-images ('sub' means that they include only about  $C/N$  counts) using the plug-in principle (Efron and Tibshirani 1993). Each of the  $B$  bootstrap sub-sinograms is first reconstructed independently. The resulting  $B$  sub-images can then be used to derive any information regarding the statistical properties of the sub-image reconstructed from a sub-sinogram. For instance, the moment of order  $k$ ,  $M_k(i)$ , in pixel  $i$  is

$$M_k(i) = \left[ \sum_{b=1}^B (\theta(\mathbf{x}^{b*})_i - \underline{\theta}^*_i)^k / (B - 1) \right], \quad (2)$$

where  $\theta(\mathbf{x}^{b*})_i$  is the value of pixel  $i$  in the sub-image reconstructed from the bootstrap sub-sinogram  $b^*$  and  $\underline{\theta}^*_i$  is the mean value of pixel  $i$  over the  $B$  reconstructed sub-images.

To estimate the statistical properties of the image reconstructed from the total acquisition of  $C$  counts,  $B/N$  bootstrap sinograms of about  $C$  counts—as opposed to  $B$  bootstrap sub-sinograms of about  $C/N$  counts—are first calculated by summing the bootstrap sub-sinograms  $N$  by  $N$ .  $B/N$  bootstrap images of about  $C$  counts are then reconstructed from these  $B/N$  bootstrap sinograms. Using the resulting  $B/N$  reconstructed bootstrap images, the moment of

different orders can be estimated using equation (2), in which the average is calculated over the  $B/N$  reconstructed images only.  $M_2(i)$  for instance gives the pixel-by-pixel variance of the reconstructed pixel values.

When applied to the sub-images, the proposed bootstrap approach does not include any assumption about the statistical properties of the projections or about the propagation of errors during reconstruction. It is thus valid whatever the statistical properties of the projections and the linear or non-linear reconstruction algorithm. However, it only yields the statistical properties of the sub-image for which replicated sub-sinograms were acquired or simulated, and not of the image reconstructed from the sum of the replicated sub-sinograms.

When applying the bootstrap approach to the images (as opposed to the sub-images), the total acquisition of  $C$  counts does not have to be replicated. However, it is assumed that a sinogram of  $C$  counts is identical to a sum of  $N$  sinograms of  $C/N$  counts that would be acquired in the same conditions. This implies that the on-line corrections (such as randoms correction in PET) that might be applied to the detected events have to be additive (e.g., the number of randoms subtracted from a  $C$  count acquisition should be identical to the sum of the number of randoms subtracted from each of the  $N$  acquisitions of  $C/N$  counts). This assumption is needed to estimate sinograms of  $C$  counts from the measured sub-sinograms of  $C/N$  counts. Indeed, when using non-linear reconstruction algorithms, the statistical properties of reconstructed images including  $C$  counts cannot be easily deduced from the statistical properties of reconstructed images including  $C/N$  counts as is the case for linear reconstruction algorithms. They thus have to be derived from the reconstruction of sinograms including  $C$  counts, hence such sinograms have to be calculated from those of  $C/N$  counts. If some sinogram processing is non-additive, the bootstrap should be applied to the sub-sinograms before these undergo any non-additive correction, and non-additive corrections should be applied to the  $C$ -count bootstrap sinograms resulting from the summing of  $N$  uncorrected  $C/N$ -count bootstrap sub-sinograms.

### 3. Materials and methods

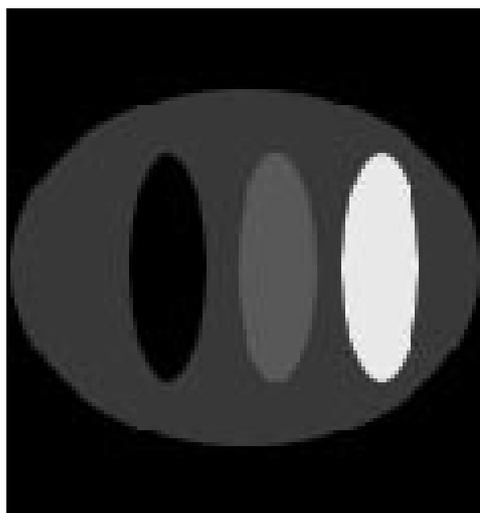
The proposed bootstrap approach was validated by studying whether it accurately predicted the statistical properties of both sub-images and images. Analytical simulations for which the statistical properties of the noise could be fully controlled and PET data including empirical noise were used and three different reconstruction schemes were considered.

#### 3.1. Analytical simulations

A SPECT acquisition (128 parallel projections over  $360^\circ$ , 128 bins per projection) of a 2D elliptical object with three elliptical regions of uniform activity (figure 1) was simulated using the RECLBL library (Huesman *et al* 1977). The activity ratios in the three inner ellipses with respect to the background corresponding to the largest ellipse were 0:1, 2:1 and 4:1, respectively. Attenuation, scatter and detector response function were not simulated. The noise-free  $128 \times 128$  sinogram included 77911 counts, with an average value ( $\pm 1$  standard deviation) in non-zero pixels of  $5.6 \pm 2.9$ . Thirty noisy realizations of the sinogram were simulated using Poisson noise. Thirty noisy realizations of the sinogram were also obtained using Gaussian noise with a constant variance over all pixels equal to 8.

#### 3.2. Real PET data

A cardiac phantom (Data Spectrum, Chapel Hill, NC) was considered, with FDG activity in the left ventricle (LV) wall ( $99.9 \text{ MBq ml}^{-1}$ ), no activity in the LV cavity and the lungs and



**Figure 1.** Simulated object. Activity ratios with respect to the background (large ellipse) were 0:1 (left ellipse), 2:1 (middle ellipse) and 4:1 (right ellipse).

background activity in the rest of the phantom representing the soft tissues ( $13.32 \text{ Mq ml}^{-1}$ ). A 15 min static PET acquisition of the phantom was performed on a GE-Advance machine operating in 2D mode. The data were acquired as a 40 image gated sequence, i.e. not gating on the cardiac cycle but using a pulse generator with a 1 s period. The resulting images thus represented 40 statistically independent replicates of a 22.5 s acquisition. The sinograms were corrected for dead time and random. A 32 min long transmission scan of the phantom was also acquired after the emission scan to get the attenuation coefficient factors that were used to correct the sinograms for attenuation before reconstruction. For this study, the sinogram (281 acquisition bins and 336 projection directions) corresponding to a single slice through the cardiac compartment of the phantom was considered, without and with attenuation correction, yielding two data sets presenting different noise properties, as attenuation correction alters the noise properties. Without attenuation correction, the total number of counts in the sinograms was 144 500 with an average pixel value in non-zero pixels of  $2.5 \pm 1.8$ . After attenuation correction, the average pixel value in non-zero pixels was  $25 \pm 21$ .

### 3.3. Statistical properties of the reconstructed sub-images

To check that the statistical properties of the sub-images were accurately predicted using the bootstrap approach, a gold standard was obtained by simulating 1000 noisy realizations of the analytical ellipsoidal phantom affected by Poisson noise and Gaussian noise. For each type of noise, the 1000 noisy sub-sinograms were reconstructed using three schemes: filtered backprojection with a Ramp filter (cut-off frequency of  $0.5 \text{ pixel}^{-1}$ ) (FBP-Ramp), FBP with a Hann filter (cut-off frequency of  $0.3 \text{ pixel}^{-1}$ ) (FBP-Hann) and ordered subset expectation maximization with eight subsets and three iterations (OSEM24). For each type of noise in the sinograms and each reconstruction scheme, the statistical properties of reconstructed pixel values were characterized from these 1000 reconstructed sub-images using: (1) the histograms of reconstructed pixel values in specific regions of interest; (2) the moment of order 2 (variance); (3) the moment of order 3 divided by  $M_2(i)^{3/2}$  which corresponds to the skewness coefficients characterizing the departure of the observed statistical distribution

from a symmetrical statistical distribution and (4) the one-dimensional (1D) local covariance at a specific image position  $i$  (Wilson *et al* 1994), obtained by plotting the elements of the covariance matrix

$$M_2(i, i + d) = \sum_{k=1}^K [\text{image}_k(i) - M_1(i)][\text{image}_k(i + d) - M_1(i + d)]/K, \quad (3)$$

where  $\text{image}_k(i)$  is the value of pixel  $i$  in the  $k$ th reconstructed sub-image,  $M_1(i)$  is the mean value of pixel  $i$  over all reconstructed sub-images,  $\text{image}_k(i + d)$  is the value of pixel  $i$  shifted by  $d$  pixels in the  $x$  direction and  $M_1(i + d)$  is the mean value over all reconstructed sub-images of pixel  $i$  shifted by  $d$  pixels in the  $x$  (horizontal) direction.  $K$  is the number of sub-images in the sample, i.e.  $K = 1000$  for the gold standard.

These figures of merit were compared to the corresponding figures of merit calculated from the images obtained using the bootstrap approach. For each type of noise, 30 noisy sub-sinograms were used to generate 1000 bootstrap sub-sinograms. These were reconstructed using the three schemes previously described and the statistical properties of the resulting bootstrap sub-images were characterized using the figures of merit.

For the PET cardiac data without and with attenuation correction, the 40 empirical replicates were reconstructed using the three reconstruction schemes. For each reconstruction scheme, the 40 resulting sub-images were used to get an estimate of the statistical properties of the reconstructed sub-images, providing an imperfect 'gold standard', due to the small number of available empirical replicates. These statistical properties were compared to those estimated using the bootstrap approach: for each of the two data sets (without and with attenuation correction), 1000 bootstrap sub-sinograms were calculated from the 40 sub-sinograms and reconstructed. The statistical properties of the resulting sub-images were then characterized.

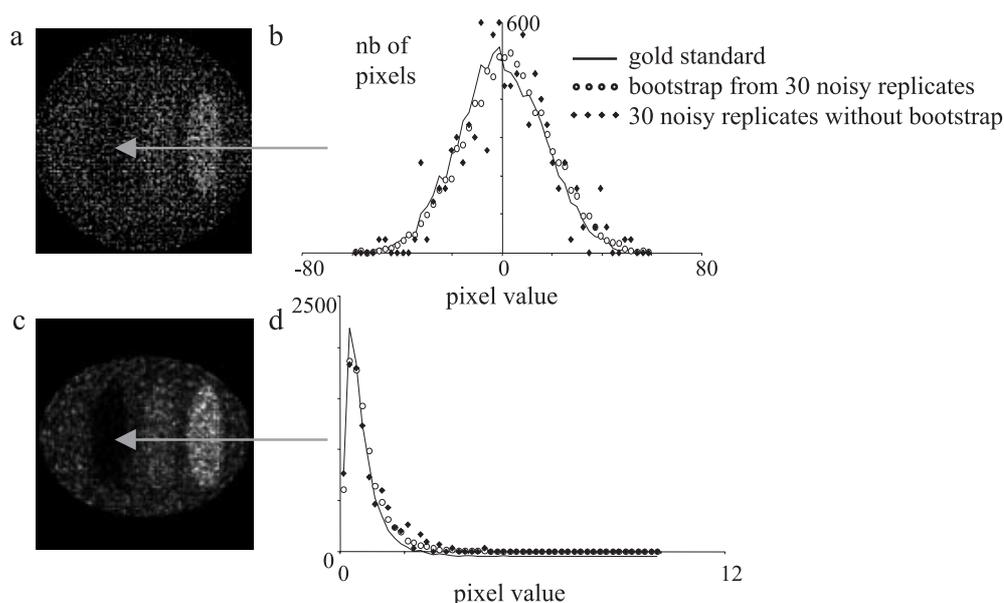
### 3.4. Statistical properties of the reconstructed images

Validation of the accurate prediction of the statistical properties of images (as opposed to sub-images) was performed using numerical simulations only, as only in that case could a gold standard be easily obtained. For Poisson or Gaussian noise, 1000 noisy realizations of a noise-free sinogram including 30 times more counts than the sinogram described in section 3.1 were generated. These noisy realizations were reconstructed using the three reconstruction schemes and the statistical properties of the resulting images were characterized using the moments corresponding to the variance and the skewness. On the other hand, 30 000 bootstrap sub-sinograms were calculated from the 30 noisy sub-sinograms and then grouped 30 by 30 to yield 1000 bootstrap sinograms. These 1000 bootstrap sinograms were reconstructed using the three reconstruction schemes and the statistical properties of the reconstructed images were derived.

## 4. Results and discussion

### 4.1. Statistical properties of reconstructed sub-images

**4.1.1. Simulations.** For the analytical simulation with Gaussian noise reconstructed with FBP-Ramp, figure 2(b) shows an example of the histograms of reconstructed pixel values obtained using the 1000 noisy sub-sinograms that were simulated (gold standard), using the 1000 bootstrap sub-sinograms calculated from 30 noisy sub-sinograms only and using 30 noisy sub-sinograms only without bootstrap. The first two histograms were calculated using 9000 pixel values corresponding to a 9-pixel region of interest drawn in the coldest ellipse and to the 1000 values obtained for each pixel from the 1000 reconstructed images. The last histogram



**Figure 2.** Examples of sub-images reconstructed using FBP-Ramp from the analytical simulation including Gaussian noise (a) and reconstructed using OSEM24 from the analytical simulation including Poisson noise (c). Parts (b) and (d) show histograms of pixel values in a 9-pixel region of interest located in the cold ellipse: the distribution of pixel values as estimated using the bootstrap from 30 noisy replicates (open circles) agreed well with the distribution of pixel values obtained using 1000 noisy replicates (plain line). The distributions of pixel values derived from 30 noisy sinograms only without using the bootstrap (diamonds) were much noisier than those obtained from the same 30 sinograms but using the bootstrap (open circles), demonstrating the enhanced statistical power resulting from the use of the bootstrap.

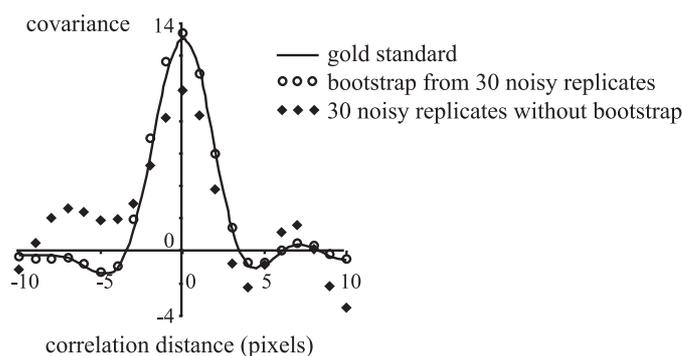
was calculated using 270 pixel values only (9-pixel region of interest times 30 reconstructed images) and was therefore normalized to be compared to the other two. There was an excellent agreement between the gold standard and the bootstrap histograms, demonstrating that the bootstrap approach accurately predicted the full statistical distribution of reconstructed pixel values using only 30 noisy sub-sinograms. Using 30 noisy sub-sinograms without the bootstrap yielded a much noisier estimate of the statistical distribution of reconstructed pixel values. Figure 2(d) shows the same results for the simulated sinograms including Poisson noise and reconstructed with OSEM24. Again, there was an excellent agreement between the gold standard and the bootstrap histograms, although these histograms were very different from those obtained for FBP-Ramp applied on the data affected by Gaussian noise. Similar to these examples, for all data sets (Poisson and Gaussian noise, all three reconstruction schemes), the bootstrap always accurately predicted the full statistical distribution of local pixel values.

Table 1 gives the average per cent differences in pixel standard deviation ( $\sigma$ ) estimated using the bootstrap approach and using the 1000 noisy simulations (gold standard):

$$100 \times (\text{estimated } \sigma - \text{gold standard } \sigma) / \text{gold standard } \sigma.$$

This average difference was calculated over all pixels inside the largest outer ellipse contour.

This table shows that the standard deviation value associated with each pixel value was accurately estimated using the bootstrap approach for the two types of noise and for the three reconstruction schemes. The higher variability of the per cent difference observed for



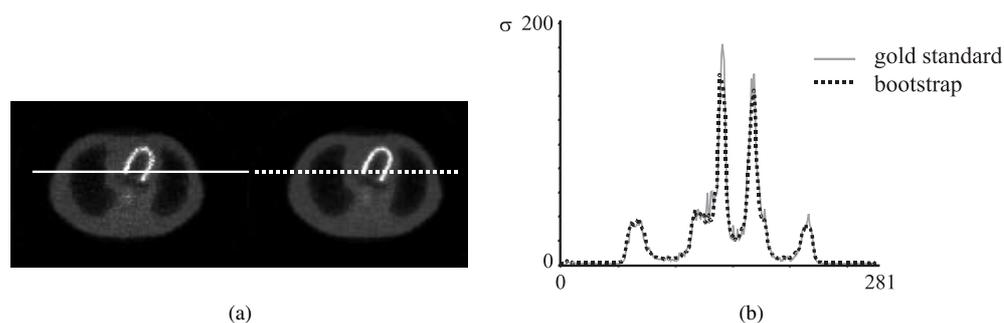
**Figure 3.** 1D local covariance for a pixel located in the background (largest ellipse) of the phantom, for the sinograms including Gaussian noise and reconstructed with FBP-Hann. The local covariance obtained using the 1000 images reconstructed from the 1000 noisy sub-sinograms that were simulated (gold standard—plain line) was accurately estimated by the local covariance measured using the 1000 images reconstructed from the 1000 bootstrap sub-sinograms (open circles) but poorly estimated when using only 30 images reconstructed from 30 sub-sinograms, without bootstrap (diamonds).

**Table 1.** Average per cent differences in standard deviation values associated with the reconstructed sub-image pixel values ( $\pm 1$  standard deviation) between the bootstrap estimates and the gold standard for the analytical simulations.

Reconstruction scheme	Poisson noise	Gaussian noise
FBP-Ramp	$-1.3 \pm 3.4$	$-1.4 \pm 3.4$
FBP-Hann	$-1.3 \pm 3.4$	$-1.9 \pm 3.3$
OSEM24	$-1.3 \pm 9.5$	$-1.7 \pm 11.4$

OSEM24 sub-images compared to FBP sub-images (standard deviation around 9 instead of 3 for Poisson noise) results from the fact that when sub-images are reconstructed with OSEM, the standard deviation in each pixel is strongly correlated with the reconstructed pixel value; thus low standard deviation values are observed in pixels with a theoretical zero value (Barrett *et al* 1994). For instance, for the Poisson simulated data reconstructed with FBP-Ramp, the standard deviation values were about 13 for pixels with reconstructed values around 1 and were about 18 (i.e., about 1.4 times higher) for pixels with reconstructed pixel values around 27, while for OSEM24, the standard deviations were about 1 and 8.5, respectively, for pixels with reconstructed values around 1 and 27, respectively. Therefore, with OSEM24, for those pixels with very low standard deviations in the ‘gold standard  $\sigma$  image’ (belonging to the ellipse with a theoretical zero value), a small error in the standard deviation estimate caused a high per cent difference in standard deviation values, hence the higher variability of the per cent difference between estimated and gold standard  $\sigma$  observed for OSEM24.

Figure 3 shows the 1D local covariance for a pixel located in the background (largest ellipse) of the phantom, for the sinograms including Gaussian noise and reconstructed with FBP-Hann. The three curves correspond to the local covariance obtained using the 1000 images reconstructed from the 1000 noisy simulated sub-sinograms (gold standard), using the 1000 images reconstructed from the 1000 bootstrap sub-sinograms and using only 30 images reconstructed from 30 sub-sinograms (no bootstrap). The local covariance function includes a central peak at  $d = 0$ , representing the variance, while the values corresponding to  $d$  different from zero represent the noise correlation at distance  $d$  from the considered pixel in the  $x$  (horizontal) direction. This figure demonstrates that for this particular pixel, the



**Figure 4.** (a) Standard deviation images associated with the PET phantom sub-images reconstructed using OSEM24 as estimated using the 40 empirical replicates (left) and the bootstrap approach (right), and (b) horizontal profiles through the standard deviation images.

**Table 2.** Average per cent differences in standard deviation values associated with the reconstructed sub-image pixel values ( $\pm 1$  standard deviation) between the bootstrap estimates and the empirical estimates for the PET phantom.

Reconstruction scheme	No attenuation correction	With attenuation correction
FBP-Ramp	$1.9 \pm 12.1$	$1.9 \pm 12.0$
FBP-Hann	$1.8 \pm 12.1$	$1.9 \pm 12.1$
OSEM24	$1.9 \pm 18.0$	$5.1 \pm 19.4$

bootstrap approach accurately estimated the gold standard local covariance, while the local covariance obtained using 30 sub-sinograms only without bootstrap was a poor estimate of the gold standard local covariance. The agreement between the 1D local covariance given by the gold standard and the 1D local covariance given by the bootstrap approach was expected, because as the bootstrap applies on entire rows of the sinograms (see section 2.2), nothing was supposed to make the spatial correlation present in the bootstrap sinograms different from that present in the original sinograms. Hence the correlation introduced by the reconstruction algorithm should yield identical correlation patterns in the images reconstructed from the gold standard and from the bootstrap sinograms. Similar agreements between gold standard and bootstrap local covariance were observed for other pixels randomly chosen in the image, for the two types of noise and for the three reconstruction algorithms that were considered (results not shown).

**4.1.2. Real PET data.** For the real PET data with sub-sinograms corrected for attenuation and reconstructed using OSEM24, figure 4(a) shows the standard deviation image associated with the reconstructed sub-images as estimated using the 40 empirical replicates and using the bootstrap approach. Profiles across these images (figure 4(b)) suggest that the bootstrap estimate is a low noise estimate of that obtained when using only the 40 empirical replicates.

The average differences in pixel standard deviation estimated from the 40 empirical replicates and from the bootstrap approach are summarized in table 2 for the three reconstruction schemes and for the sub-sinograms corrected for attenuation or not.

Table 2 shows that for real noise as observed in experimental PET data, which is neither Poisson nor Gaussian but most probably spatially correlated due to the corrections applied to the data (dead time, random and optionally attenuation), the bootstrap approach yielded an accurate estimate of the standard deviation value associated with each reconstructed pixel value. The variability of the per cent differences between estimated and 'gold standard' standard deviation

**Table 3.** Average per cent differences in standard deviation values associated with the reconstructed sub-image pixel values ( $\pm 1$  standard deviation) depending on the number of sub-sinograms available and on whether the bootstrap was used.

Number of samples	Bootstrap	FBP-Ramp	FBP-Hann	OSEM-24
5	No	$-15.4 \pm 30.8$	$-15.7 \pm 31.4$	$-20.2 \pm 39.0$
5	Yes	$-10.3 \pm 5.3$	$-10.0 \pm 4.7$	$-11.7 \pm 19.1$
10	No	$-7.9 \pm 22.0$	$-6.9 \pm 22.4$	$-10.8 \pm 31.0$
10	Yes	$-5.2 \pm 4.5$	$-4.6 \pm 4.3$	$-6.1 \pm 15.1$
30	No	$-2.0 \pm 12.8$	$-2.2 \pm 13.0$	$-3.3 \pm 19.7$
30	Yes	$-1.3 \pm 3.4$	$-1.3 \pm 3.4$	$-1.3 \pm 9.5$
100	No	$-0.5 \pm 6.7$	$-0.5 \pm 6.7$	$-0.8 \pm 11.1$

values was higher for the PET data (table 2) than for the analytical simulations (table 1). This is probably because for the PET data, the ‘gold standard’ was imperfect since it was derived from 40 experimental measurements only. It was thus noisier than the gold standard calculated for the analytical simulations for which 1000 noisy realizations of the sinograms were available.

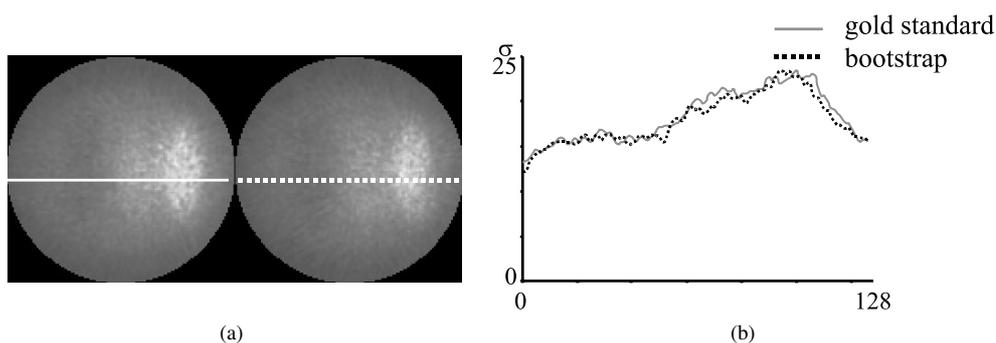
*4.1.3. Increased accuracy in standard deviation estimates resulting from the bootstrap approach.* The experiments performed using the simulations and the PET data demonstrate that when a small number of realizations of a specific configuration are available (30 for the simulations and 40 for the PET data in our examples), using the bootstrap makes it possible to predict the statistical properties of reconstructed sub-images with a high accuracy. Table 3 illustrates the gain in accuracy resulting from the use of the bootstrap ( $B = 500$  bootstrap realizations were used) as a function of the number of available empirical samples for the Poisson simulation reconstructed using FBP-Ramp, FBP-Hann and OSEM-24. For a given number of empirical samples, using the bootstrap always reduces the bias with which the standard deviations associated with reconstructed pixel values are estimated and also reduces the variability of this bias by a factor between 2 and 5. Using 30 empirical replicates and the bootstrap yields an estimate of the pixel-by-pixel standard deviations as good as and even much better than (in terms of variability of the bias) that obtained using 100 empirical replicates without the bootstrap.

For simulations and phantom experiments, the bootstrap is thus a useful adjunct to increase the power of the statistical tests that might be performed using the reconstructed image pixel values, since it yields accurate estimates of the standard deviations associated with reconstructed pixel values.

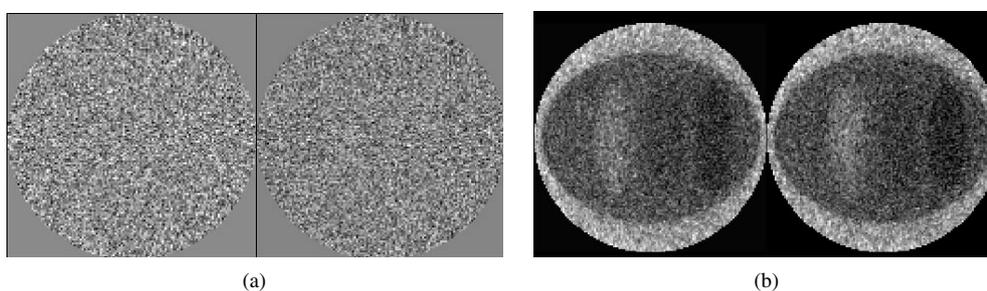
#### *4.2. Statistical properties of reconstructed images*

For the analytical Poisson simulation reconstructed using FBP-Hann, figure 5(a) shows the images of pixel standard deviations estimated using the 1000 noisy realizations of the sinograms and using the bootstrap performed from the 30 noisy realizations of sub-sinograms. The profiles drawn across the images confirm that the bootstrap accurately estimated the standard deviation associated with reconstructed pixel values (figure 5(b)).

Table 4 lists the average differences in pixel standard deviation  $\sigma$  estimated using the bootstrap approach and using the 1000 noisy replicates of the sinogram, for the three reconstruction schemes and for the two types of noise. This table demonstrates that the bootstrap approach using 30 noisy sub-sinograms obtained by splitting the total acquisition into



**Figure 5.** (a) Standard deviation images associated with the image reconstructed using FBP-Hann from the Poisson analytical simulation as estimated using 1000 noisy replicates (gold standard, left) and using the bootstrap approach from 30 noisy replicates of sub-sinograms (right), and (b) horizontal profiles through the standard deviation images.



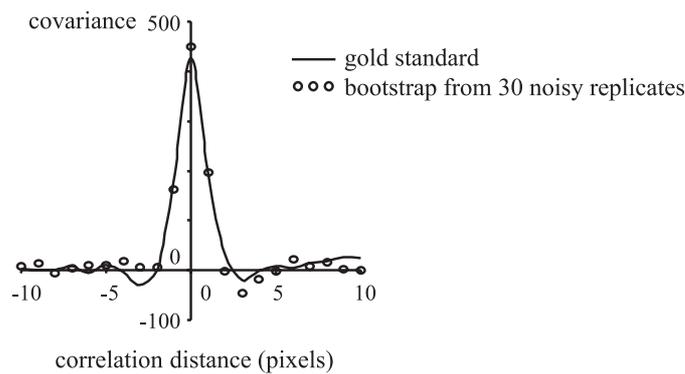
**Figure 6.** Images of the skewness coefficients as estimated from the noisy replicates (gold standard, left) and from the bootstrap (right) for the Poisson simulation reconstructed using FBP-Ramp (a), and OSEM24 (b). The two images in each pair are represented on a grey scale with the same minimum and maximum values.

**Table 4.** Average per cent differences in standard deviation values associated with the reconstructed image pixel values ( $\pm 1$  standard deviation) between the bootstrap estimates and the gold standard for the analytical simulations.

Reconstruction scheme	Poisson noise	Gaussian noise
FBP-Ramp	$-1.2 \pm 3.4$	$-1.5 \pm 3.5$
FBP-Hann	$-1.3 \pm 3.3$	$-2.0 \pm 3.4$
OSEM24	$-1.2 \pm 9.4$	$-2.6 \pm 12.5$

30 accurately predicted the standard deviations associated with the pixel values reconstructed from the total acquisition.

Figure 6 shows the pixel-by-pixel skewness coefficients obtained using the noisy replicates and using the bootstrap approach for the Poisson simulation reconstructed using FBP-Ramp and OSEM24. The bootstrap approach properly showed that for FBP-Ramp, the statistical distribution of the reconstructed pixel values was symmetrical in all pixels, while it was skewed in low count pixels with OSEM24. These results therefore demonstrate that the bootstrap also allows for an accurate prediction of the third order moment of the reconstructed image. This was true for the three reconstruction schemes and for both types of noise (Poisson and Gaussian). The asymmetrical statistical distributions of reconstructed pixel values in the



**Figure 7.** 1D local covariance for a pixel located in the background (largest ellipse) of the phantom, for the sinograms including Poisson noise and reconstructed with OSEM24. The local covariance obtained using the 1000 images reconstructed from the 1000 noisy sinograms that were simulated (gold standard—plain line) was well estimated by the local covariance measured using the 1000 images reconstructed from the 1000 bootstrap sinograms (open circles).

low count regions observed with OSEM24 are a consequence of the non-negativity constraint included in the OSEM algorithm.

Figure 7 displays the 1D local covariance for a pixel located in the background (largest ellipse) of the phantom, for the sinograms including Poisson noise and reconstructed with OSEM24. The local covariance in the reconstructed image as estimated using the bootstrap approach properly estimated the gold standard local covariance, confirming the potential of the bootstrap approach to provide not only accurate estimates of the noise variance in the reconstructed images, but also good estimates of the noise correlation. Similar agreements between gold standard and bootstrap covariance estimates were observed for all types of noise and reconstruction algorithms, and for various pixels randomly chosen within the reconstructed images (results not shown).

These results show that the bootstrap approach can be used to determine the statistical properties of the image reconstructed from a single acquisition, provided this acquisition can be split into  $N$  sub-acquisitions or acquired in list-mode. It is then appropriate for determining the variance image associated with images reconstructed from clinical studies for which gating using an electronic pulse generator or list-mode acquisition is feasible. It could also be used for analysing noise correlation in such clinical images, which might be useful for image quality analysis based on observer studies, in which noise correlation has been shown to play an important role (e.g., Myers *et al* 1985).

#### 4.3. Impact of the number of empirical sub-sinograms for estimating the statistical properties of a reconstructed image

To estimate the statistical properties of a reconstructed image, the bootstrap approach requires the acquired data to be split into  $N$  sub-sinograms (section 2.2). For a fixed total number of counts, the number of sub-sinograms to be considered has to be studied, as the larger the number of sub-sinograms, the smaller the number of counts in each sub-sinogram. Using the Poisson simulations and 500 bootstrap realizations, we found that the average differences in pixel standard deviation ( $\pm 1$  standard deviation) estimates using the bootstrap approach and the 1000 noisy replicates of the sinogram (gold standard) were  $-10.0 \pm 5.1$ ,  $-4.8 \pm 4.5$ ,  $-2.4 \pm 4.2$ ,  $-1.2 \pm 3.4$  and  $-1.1 \pm 4.1$  for  $N = 5, 10, 20, 30$  and  $40$ , respectively, for FBP-Ramp. The

corresponding values were  $-9.6 \pm 9.2$ ,  $-4.9 \pm 9.5$ ,  $-3.1 \pm 9.4$ ,  $-1.2 \pm 9.4$  and  $-1.3 \pm 9.5$  for OSEM24. This suggests that the larger the number of sub-sinograms that can be acquired or simulated, the smaller the bias affecting the standard deviation estimates. This is consistent with the fact that low count sinograms including more noise best describe the noise properties of the acquired data. In our examples, using 20 or best 30 sub-sinograms yielded quite accurate estimates of the standard deviation of the reconstructed pixel values.

#### 4.4. Impact of the number of bootstrap realizations

When studying the statistical properties of reconstructed images, the number,  $B$ , of bootstrap realizations to be generated from the  $N$  sub-sinograms available must be appropriately chosen. Using the Poisson simulations and  $N = 30$  sub-sinograms, we found that the averaged differences in pixel standard deviation estimates derived from the bootstrap approach and from the 1000 noisy replicates of the sinogram (gold standard) were  $-3.4 \pm 12.0$ ,  $-1.9 \pm 7.4$ ,  $-1.5 \pm 5.6$ ,  $-1.4 \pm 4.0$  and  $-1.3 \pm 3.4$  for  $B = 30, 100, 200, 500$  and  $1000$ , respectively, for FBP-Ramp. The corresponding values were  $-4.5 \pm 21.6$ ,  $-2.1 \pm 14.5$ ,  $-1.7 \pm 12.0$ ,  $-1.4 \pm 10.1$  and  $-1.3 \pm 9.5$  for OSEM24. Using  $B = 200$  bootstrap realizations thus already yields an accurate estimate of the standard deviation distributions associated with pixel values and obviously, the larger the  $B$ , the better the accuracy of the standard deviation estimates. The precise number to be considered should then be chosen as a function of the computational burden associated with the reconstruction of  $B$  bootstrap sinograms.

#### 4.5. Parametric versus non-parametric bootstrap

The bootstrap approach we propose can be qualified as non-parametric as it does not include any assumption regarding the statistical properties of the sinograms. An alternative bootstrap approach proposed in PET (Haynor and Woods 1989) consists in considering the list-mode of the  $C$  acquired events before any correction so that they are still Poisson distributed.  $B$  bootstrap list-modes are then created from this list-mode, each bootstrap list-mode being obtained by randomly drawing with replacement  $C$  events from the original list-mode. The  $B$  bootstrap list-modes then undergo the correction and reconstruction processing to yield  $B$  reconstructed images. Although this bootstrap approach is accurate, its use is both less practical (need to access the list-mode before any correction, large storage space required to store the  $B$  bootstrap list-modes, need to apply corrections to the  $B$  created bootstrap list-modes, large computational burden for processing the  $B$  bootstrap list-modes) and more restricted (need for the list-mode data to be Poisson distributed) than that of the non-parametric bootstrap approach we propose, which processes directly the sinograms whatever their statistical distributions.

## 5. Conclusion

We introduced and validated a non-parametric bootstrap approach to estimate the statistical properties of PET and SPECT reconstructed images whatever the noise properties in the projections and the reconstruction algorithm. This approach can be used when several empirical replicates of the same configurations are available, for instance, for simulated data or phantom experiments: it then greatly increases the accuracy and reduces the variability with which the statistical properties of the reconstructed images (especially the variance and covariance of pixel values) can be estimated. The proposed bootstrap approach can also be used when a single acquisition is available, for instance in clinical studies, provided this can be stored in list-mode format or split into a number of sub-acquisitions. It then makes it possible

to obtain the variance associated with the reconstructed image (or any parameter related to the statistical properties of the image) or with any parameter derived from the reconstructed image. The noise correlation is also accurately predicted.

Use of this bootstrap approach to characterize the statistical properties of SPECT and PET images as a function of the processing scheme (correction and reconstruction) is in progress (Buvat *et al* 2001).

## Acknowledgments

The author thanks Steve Bacharach for kindly providing the PET data and Habib Benali, Robert Di Paola and Cyril Riddell for many helpful discussions.

## References

- Alpert N M, Barker W C, Gelman A, Weise S, Senda M and Correia J A 1991 The precision of positron emission tomography: theory and measurement *J. Cereb. Blood Flow Metab.* **11** A26–A30
- Alpert N M, Chesler D A, Correia J A, Ackerman R H, Chang J Y, Finklestein S, Davis S M, Brownell G L and Taveras J 1982 Estimation of the local statistical noise in emission computed tomography *IEEE Trans. Med. Imaging* **1** 142–6
- Barrett H H 1990 Objective assessment of image quality: effects of quantum noise and object variability *J. Opt. Soc. Am.* **7** 1266–78
- Barrett H H, Wilson D W and Tsui B M W 1994 Noise properties of the EM algorithm: I. Theory *Phys. Med. Biol.* **39** 833–46
- Budinger T F, Derenzo S E, Greenberg W L, Gullberg G T and Huesman R H 1978 Quantitative potentials of dynamic emission computed tomography *J. Nucl. Med.* **19** 309–15
- Budinger T F, Derenzo S E, Gullberg G T, Greenberg W L and Huesman R H 1977 Emission computer tomography with single-photon and positron annihilation emitters *J. Comput. Assist. Tomogr.* **1** 131–45
- Buvat I, Hapdey S, Riddell C and Soret M 2001 Noise properties of reconstructed SPECT images as a function of the reconstruction scheme *J. Nucl. Med.* **42** 138P
- Carson R E, Yan Y, Daube-Witherspoon M E, Freedman N, Bacharach S L and Herscovitch P 1993 An approximation formula for the variance of PET region-of-interest values *IEEE Trans. Med. Imaging* **12** 240–50
- Efron B and Tibshirani R J 1993 *An Introduction to the Bootstrap* (New York: Chapman and Hall)
- Fessler J A 1996 Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography *IEEE Trans. Image Process.* **5** 493–506
- Haynor D R and Woods S D 1989 Resampling estimates of precision in emission tomography *IEEE Trans. Med. Imaging* **8** 337–43
- Huesman R H 1984 A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography *Phys. Med. Biol.* **29** 543–52
- Huesman R H, Gullberg G T, Greenberg W L and Budinger T F 1977 RECLBL library users manual: Donner algorithms for reconstruction tomography *Lawrence Berkeley Laboratory, University of California Publication* 214
- Huesman R H and Mazoyer B M 1987 Kinetic data analysis with a noisy input function *Phys. Med. Biol.* **32** 1569–79
- Kadrmas D J, DiBella E V R, Huesman R H and Gullberg G 1999 Analytical propagation of errors in dynamic SPECT: estimators, degrading factors, bias and noise *Phys. Med. Biol.* **44** 1997–2014
- Myers K J, Barrett H H, Borgstrom M C, Patton D D and Seeley G W 1985 Effect of noise correlation on detectability of disk signals in medical imaging *J. Opt. Soc. Am. A.* **2** 1752–9
- Palmer M R, Bergstrom M, Pate B D and Beddoes M P 1986 Noise distribution due to emission and transmission statistics in positron emission tomography *IEEE Trans. Nucl. Sci.* **33** 439–42
- Qi J and Leahy R M 2000 Resolution and noise properties of MAP reconstruction for fully 3D PET *IEEE Trans. Med. Imaging* **19** 493–506
- Riddell C, Carson R E, Carrasquillo J A, Libutti S K, Danforth D N, Whatley M and Bacharach S L 2001 Noise reduction in oncology FDG PET images by iterative reconstruction: a quantitative assessment *J. Nucl. Med.* **42** 1316–23
- Soares E J, Byrne C L and Glick S J 2000 Noise characterization of block-iterative reconstruction algorithms: I. Theory *IEEE Trans. Med. Imaging* **19** 261–70

- 
- Wang W and Gindhi G 1997 Noise analysis of MAP-EM algorithms for emission tomography *Phys. Med. Biol.* **42** 2215–32
- Wilson D W and Tsui B M W 1993 Noise properties of filtered-backprojection and ML-EM reconstructed emission tomographic images *IEEE Trans. Nucl. Sci.* **40** 1198–203
- Wilson D W, Tsui B M W and Barrett H H 1994 Noise properties of the EM algorithm: II. Monte Carlo simulations *Phys. Med. Biol.* **39** 847–71