Statistical Analysis of Maximum Likelihood Estimator Images of Human Brain FDG PET Studies

Jorge Llacer, *Senior Member, IEEE,* Eugene Veklerov, *Senior Member, IEEE,*
Kevin J. Coakley, Edward J. Hoffman, *Senior Member, IEEE,* and Jorge Nunez

Abstract— The work presented in this paper evaluates the statistical characteristics of regional bias and expected error in reconstructions of real PET data of human brain fluoro-deoxiglucose (FDG) studies carried out by the maximum likelihood estimator (MLE) method with a robust stopping rule, and compares them with the results of filtered backprojection (FBP) reconstructions and with the method of sieves.

The task that we have investigated is that of quantifying radioisotope uptake in regions-of-interest (ROI's). We first describe a robust methodology for the use of the MLE method with clinical data which contains only one adjustable parameter: the kernel size for a Gaussian filtering operation that determines final resolution and expected regional error. Simulation results are used to establish the fundamental characteristics of the reconstructions obtained by our methodology, corresponding to the case in which the transition matrix is perfectly known. Then, data from 72 independent human brain FDG scans from four patients are used to show that the results obtained from real data are consistent with the simulation, although the quality of the data and of the transition matrix have an effect on the final outcome. The most important results are that, for equal resolution, expected pixel-by-pixel error in the MLE and sieves reconstructions are lower in the regions of low counts than in the regions of high counts, the lowest being for the MLE. In contrast, FBP reconstructions show an expected error that is high and nearly independent of the number of counts in a region. As a consequence, the determination of radioisotope uptake in ROI's of high activity has approximately the same standard deviation in MLE, sieves, and FBP reconstructions, while the standard deviation in ROI's of low uptake is substantially lower for MLE, while sieves take an intermediate value. The use of a well-constructed Monte Carlo transition matrix improves all the results with real data in a measurable way. We conclude that our proposed MLE methodology and the method of sieves have a definite advantage over FBP. There is a tradeoff between shorter computation time, a slight bias but lower standard deviation for MLE and longer computation time, a basically unbiased estimation but higher standard deviation for sieves.

I. INTRODUCTION

SINCE the introduction by Shepp and Vardi [1] of the maximum likelihood estimator (MLE) method of image reconstruction for positron emission tomography (PET) based on the expectation–maximization (EM) algorithm of Dempster *et al.* [2], considerable work has been done by a number of research groups with the aim of understanding its characteristics and controlling its shortcomings. After a few years of work in our group, we have developed a body of knowledge on that subject that we feel is now ready for presentation as a practical methodology for the utilization of the MLE in the reconstruction of clinical PET data. The measurement of the basic statistical characteristics of a reconstruction method can best be made by simulation since all the parameters are under the control of the experimenter. The transition matrix of the simulated tomograph, for example, is known exactly. In the present work, however, we have wanted to carry out the assessment of the reconstructions one step further, but using data that have originated in real clinical measurements. In that case, the transition matrix is not known exactly, there may be variations of radioisotope uptake as a function of time in scans obtained consecutively from the same patient, and patients are anatomically different from each other. We have endeavored to ascertain whether, under those conditions of variability, statistical characteristics of reconstructed images from real data sets are consistent with the results of analyzing simulated images.

We have concentrated our effort on the task of evaluating radioisotope uptake in regions-of-interest (ROI's), in a comparison between our MLE methodology, filtered backprojection (FBP) methods currently used in clinical practice, and with the method of sieves [3], [4]. For the reasons indicated above, assessment of bias and variance in uptake measurements has been carried out first with simulated data. Then, by using three different transition matrices with different degrees of accuracy and a components of variance model for statistical analysis, we have been able to show that the characteristics obtained from real human FDG brain data are consistent with the results of the simulation studies.

This paper will be divided into a number of sections, with references to our or other groups' work as appropriate. Section II defines the problem and the notation to be used in this paper. Section III follows the development of the EM form of the algorithm for the practical case in which corrections for absorption, detector gain variations, and random coincidences are included in the formulation, along with considerations

regarding the conservation of counts. A variation of the algorithm based on the successive substitution (SS) method [5], which allows substantial speeding up of the MLE calculations, will also be described in that section. Section IV briefly reviews the concept of feasibility and justifies the use of a stopping rule in the iterative process. It then introduces a robust stopping rule for real PET data based on likelihood cross validation, leading to our version of the MLE for PET reconstruction. Section V presents the methods, data, and results of a comparative bias and variance study in which the characteristics of the MLE and sieve images are shown to be superior to FBP images in a well-defined manner. In the Conclusion, we summarize the results obtained, discuss the relationship between our work and that of other groups, and what has been accomplished.

II. DEFINITION OF THE PROBLEM AND NOTATION

A. The Tomographic Instrument

We consider the case of a single-plane PET instrument with n_d detectors in a ring configuration. The number of projection angles possible with that system is also n_d, and we consider that each projection is divided into n_b bins or tubes. The total number of tubes is then $n_d \times n_b = J$. We will consider the results of one measurement to consist of one vector p with J elements, containing the number of counts detected in each detector pair from true γ-ray pair coincidences, contaminated by a certain fraction of accidental coincidences, which we will call "random counts" or simply "randoms." We will consider the case in which the PET electronics can obtain a spatial estimate of the distribution of randoms in vector p by the method of a delayed time window and recording a separate randoms vector r for the delayed coincidences [6]. In many installations, each count recorded in r would be subtracted automatically from p during the data acquisition process, but we will consider obtaining both vectors independently, as can be done easily in some instruments. A vector c will contain multiplicative correction factors incorporating the effects of gamma ray absorption in the patient and detector gain nonuniformity. Vectors p, r, and c are all of dimension J. Without loss of generality, we assume that the PET detector system is fixed, i.e., no wobbling is considered. If there is detector motion, the dimension J has to be modified accordingly to reflect the total number of tubes of the system. The image plane will be described by a centered, square array of $n_p \times n_p$ pixels, represented as a vector of activities a with $n_p \times n_p = I$ elements.

Finally, we define a transition matrix f, whose elements f_{ji} correspond to the expected number of counts detected per unit time in tube j per unit radioisotope activity in pixel i, in the absence of absorber, with all detectors having uniform gain. This matrix will have I columns, one per pixel, and J rows, one per tube, and will be sparse. The sum of all the elements in a column will give the total number of counts that the system will detect per unit time for a unit activity in the corresponding pixel, i.e., its sensitivity. The matrix values that we have used for the reconstructions with simulated data

have been obtained by the simple procedure of finding the area of the intersection of a circle of area equal to a pixel area, centered on that pixel, and the parallel-line tube that joins the faces of two detectors. The geometry of the CTI-831 Neuro-PET system has been used for the calculations. For the reconstruction of real PET data, we have used two different matrices: 1) the simple matrix indicated above, and 2) a preliminary matrix calculated by Monte Carlo (MC) methods [7] taking into consideration actual detector system geometry, detector efficiency, crystal penetration, and energy resolution for the detector system (CTI-831). The generation of the MC matrix will be discussed in Appendix B, and the relationship between matrix characteristics and the application of measured gain corrections [8] will be described in Appendix A.

B. Handling Random Coincidences

If a tomograph can provide an estimate of the randoms counted during a measurement in the form of vector r described above, we can postulate the existence of a fictitious pixel somewhere in space that emits coincidence gamma rays into the detector pairs of the tomograph with a spatial probability proportional to the elements of vector r. Then, it is an additional task of the MLE to estimate the activity a_b in that background pixel, along with the rest of the pixels in the image plane. The vector r, or a smoothed version of it, becomes one more column of the matrix f since it can be considered to contain elements that correspond to the number of background counts detected during the measurement time by each detector pair for an arbitrary unit of activity in the background pixel. This column will not be sparse, and it is fully corrected for absorptions, gain differences, etc. since it is the result of a measurement. The MLE estimation of a_b effectively subtracts the background from the reconstructed image in a manner statistically consistent with the rest of the estimation. Politte and Snyder in [9] treated the problem of randoms as a correction, assuming our "background pixel" to be a known constant, rather than a parameter to be estimated.

C. Summary of Notation

The notation that will be used in this paper can then be summarized as follows:

$p_j, j = 1, ..., J$ projection data, counts detected in tube j

$r_j, j = 1, ..., J$ randoms vector

$c_j, j = 1, ..., J$ absorption and gain corrections vector

$a_i, i = 1, ..., I$ activity in a pixel, the image being estimated

$a_i^{(k)}, i = 1, ..., I$ the activity estimated at the kth iteration

f_{ji} transition matrix elements, uncorrected

$f_{ji}' = f_{ji}c_j$ transition matrix elements, corrected

a_b activity in the fictitious background pixel

$h_j = \sum_{i=1}^{I} \frac{f_{ji}a_i}{c_j} + r_j a_b$ the projection of an image estimate.

In addition, a more concise representation of the forward-projection of an estimate a, fully corrected for absorption, gain variations, and random background, will be defined:

$$h_j = \sum_{i=1}^{I} \frac{f_{ji}a_i}{c_j} + r_j a_b = \sum_i f_{ji}' a_i. \tag{1}$$

We establish the convention that $\sum_i f'_{ji} a_i$ includes the fictitious pixel term $r_j a_b$ in all cases without explicitly mentioning it.

Two sensitivity expressions will be useful in the analysis: 1) the uncorrected sensitivity of the detector system for a pixel

$$q_i \sum_{j=1}^{J} f_{ji}, \qquad (2)$$

and 2) the sensitivity corrected for absorption and gain variations

$$q'_i = \sum_{j=1}^{J} \left(\frac{f_{ji}}{c_j} \right). \qquad (3)$$

III. ITERATIVE ALGORITHMS

The estimation of pixel values for the case in which the elements of the corrections vector c are all unity (i.e., no absorption or grain corrections) and the sensitivities q_i are all unity was treated by Shepp and Vardi [1]. In that case, the matrix elements f_{ji} are true probabilities, and those authors were able to develop an iterative formula based on the EM algorithm [2] that is known to converge towards a possibly unique maximum [10] and to preserve the number of counts. The case with corrections and constant randoms has been treated more recently by Politte and Snyder [9]. The introduction of attenuation corrections into the EM algorithm has also been treated by Hebert and Leahy [11]. In this paper, we present the derivation of the EM method with corrections, with additional consideration to the problem of random coincidences. A different formulation of the MLE based on the SS algorithm (4), which allows speeding up convergence, will also be presented.

A. EM Algorithm

For the development of the EM algorithm, we will make use of a vector x, the "complete" data set whose elements x_{ji} are the number of disintegrations from pixel i that are detected in tube j. The relationship between the complete data set and the "incomplete" set p is given by

$$p_j = \sum_i x_{ji}. \qquad (4)$$

We will use the EM algorithm [2] to estimate a by maximizing the log likelihood of x given an activity distribution a. The complete data set x has elements which are Poisson distributed with means $f'_{ji} a_i$, while the detected counts or incomplete data set vector p has elements which are Poisson variables with mean $\sum_i f'_{ji} a_i$. We consider only the case in which the data p are uncorrected for attenuation, etc., i.e., they are purely Poisson in nature.

The target function to be maximized is, from [1] and introducing the requirement of conservation of counts,

$$
\begin{aligned}
L(a) &= \log P(x|a) \\
&= \sum_j \sum_i \left[- f'_{ji} a_i + x_{ji} \log(f'_{ji} a_i) - \log(x_{ji}!) \right] \\
&\quad - \mu \left(\sum_i q'_i a_i - \sum_j p_j \right)
\end{aligned}
\qquad (5)
$$

where P is the conditional probability of a complete data set x having resulted from an activity distribution a. The Lagrange multiplier μ is one more parameter to be estimated by the algorithm, along with the distribution a. *E-step:* Shepp and Vardi show in [1] that, for Poisson random variables x_{ji}, the conditional expected value of x_{ji} given the data elements p_j for a current estimate of the parameter $a_i^{(k)}$ is

$$x_{ji}^{(k)} = E\{x_{ji}|p, a^{(k)}\} = \frac{a_i^{(k)} f'_{ji} p_j}{\sum_l f'_{jl} a_l^{(k)}}. \qquad (6)$$

A physical interpretation of this estimate can be seen by considering that if a tube has p_j counts and the pixel values are $a_i^{(k)}$, the expected fraction of p_j that was emitted by the ith pixel is $f'_{ji} a_i^{(k)} / \sum_l f'_{jl} a_l^{(k)}$.

We can then compute the expected value of the target function (5) given the incomplete data p and the current estimate $a^{(k)}$:

$$
\begin{aligned}
E\{L(a|p, a^{(k)})\} = \sum_j \sum_i \Big[&- f'_{ji} a_i + E\{x_{ji}|p, a^{(k)}\} \\
&\cdot \log(f'_{ji} a_i) - E\{\log(x_{ji}!)\} \Big] \\
&- \mu \left(\sum_i q'_i a_i - \sum_j p_j \right).
\end{aligned}
\qquad (7)
$$

Replacing from (6), we have

$$
\begin{aligned}
E\{L(a|p, a^{(k)})\} = \sum_j \sum_i \Bigg[&- f'_{ji} a_i + \frac{p_j f'_{ji} a_i^{(k)}}{\sum_l f'_{jl} a_l^{(k)}} \\
&\cdot \log(f'_{ji} a_i) + \text{terms independent of} a_i \Bigg] \\
&- \mu \left(\sum_i q'_i a_i - \sum_j p_j \right).
\end{aligned}
\qquad (8)
$$

M-step: We now maximize the expectation equation (8) with respect to the parameters a, which will lead to the new values $a^{(k+1)}$. Setting the partial derivatives with respect to a_i and μ equal to zero,

$$\frac{\partial E\{\}}{\partial a_i} = \sum_j \left[-f'_{ji} + \frac{p_j f'_{ji} a_i^{(k)}}{\sum_l f'_{jl} a_l^{(k)}} \frac{1}{a_i} \right] - \mu q'_i = 0 \quad (9a)$$

$$\frac{\partial E\{\}}{\partial \mu} = \sum_i q'_i a_i - \sum_j p_j = 0 \qquad (9b)$$

and solving for a_i in terms of μ, we get the next estimate of $a_i^{(k+1)}$:

$$a_i^{(k+1)} = \frac{1}{\mu + 1} \frac{a_i^{(k)}}{q'_i} \sum_j \left\{ \frac{p_j f'_{ji}}{\sum_{i'} f'_{ji'} a_{i'}^{(k)}} \right\}. \qquad (10)$$

To compute μ from (9b), we set the condition

$$\sum_i a_i^{(k+1)} q'_i = \sum_j p_j. \qquad (11)$$

Placing $a_i^{(k+1)}$ from (10) into (11) and solving for μ, we obtain the $(k+1)$ estimate of that parameter in terms of the image estimates at iteration (k) as

$$\mu = \frac{\sum_i a_i^{(k)} \sum_j \left\{ \frac{p_j f_{ji}'}{\sum_l f_{jl}' a_l^{(k)}} \right\}}{\sum_j p_j} - 1. \tag{12}$$

Interchanging the order of summation in the numerator of (12), it becomes evident that $\mu = 0$ for any set of values of $\boldsymbol{a}$ and any representation of the matrix $\boldsymbol{f}'$. Thus, conservation of counts is assured for the MLE problem of PET tomography solved iteratively by the EM algorithm.

The quotient in brackets in (10) can be simplified by multiplying the numerator and denominator by c_j so that the iterative step can be applied with the uncorrected matrix elements f_{ji}. Our MLE–EM procedure, explicitly including the fictitious background pixel estimation, is given by the following iterative formulas:

$$a_i^{(k+1)} = a_i^{(k)} \frac{1}{q_i'} \sum_{j=1}^{J} \left\{ \frac{p_j f_{ji}}{\displaystyle\sum_{l=1}^{I} f_{jl} a_l^{(k)} + r_j c_j a_b} \right\} \tag{13}$$

for the image pixels and

$$a_b^{(k+1)} = a_b^{(k)} \frac{1}{\displaystyle\sum_{j=1}^{J} r_j} \sum_{j=1}^{J} \left\{ \frac{p_j r_j}{\displaystyle\sum_{l=1}^{I} f_{jl} a_l^{(k)} + r_j c_j a_b} \right\} \tag{14}$$

for the background pixel. The final results are in activity concentration in the pixels. It may be desired to convert that value to counts by multiplying the resulting values a_i by the corresponding uncorrected sensitivity values q_i of (2). If the measured and smoothed sample of randoms $\boldsymbol{r}$ is a correct representation of the actual random coincidences that are included in the data $\boldsymbol{p}$, the value of a_b goes to unity as the iterative procedure advances. This is rarely found in practice, however.

B. Successive Substitution Method

The idea of modifying the EM solution for accelerated convergence has been treated extensively in the statistics literature, starting from Dempster et al. [2] in 1977, to Silverman et al. [12] in 1990. In the medical imaging literature, Lewitt and Muehllehner [13] and Kaufman [14] initiated practical modifications of the algorithm, and Tanaka [15] proposed the exponentiation of the EM correction term with renormalization after every iteration. More recently, Rajeevan et al. [16] have presented an elegant method of acceleration by vector extrapolation of a set of preliminary corrections. With the development to be described below, we justify the exponentiation form of acceleration, and place it in a form that can be used with practical data sets, although, unfortunately, a proof of convergence is not available.

The SS method can be described as follows: given a series of equations in the form

$$a_i = KF(\{a_m\}), \qquad i = 1, \cdots, I \tag{15}$$

where F is some function, $\{a_m\}$ is the complete set of variables $a_i, \cdots, a_I$, and K is a normalization constant, then (15) can be transformed into a recursive relation by

$$a_i^{(k+1)} = KF(\{a_m^{(k)}\}), \qquad i = 1, \cdots, I. \tag{16}$$

Each of the new values of $a_i^{(k+1)}$ is calculated from all the known I values of $a_m^{(k)}$ and the complete set $\{a_m\}$ is updated at once. The constant K may be obtained by invoking some conservation law, for example.

For the MLE case, we start with the log likelihood of obtaining a set of measurements $\boldsymbol{p}$ conditional on having a set of pixel activities $\boldsymbol{a}$:

$$L(\boldsymbol{a}) = \sum_j \left[-\sum_i f_{ji}' a_i + p_j \log \left(\sum_i f_{ji}' a_i \right) - \log(p_j!) \right]$$
$$- \mu \left(\sum_i q_i' a_i - N \right) \tag{17}$$

where μ is a Lagrange multiplier to be evaluated for the conservation of counts.

We then maximize (17) with respect to a_i by setting the partial derivatives equal to zero:

$$\frac{\partial L(\boldsymbol{a})}{\partial a_i} = \sum_j -f_{ji}' + \sum_j \frac{p_j f_{ji}'}{\displaystyle\sum_l f_{jl}' a_l} - \mu q_i' = 0. \tag{18}$$

Noticing that the first summation is equal to q_i', we divide by that quantity, move $\mu + 1$ to the right-hand side, raise the expression to the power n, multiply both sides by a_i, and obtain

$$a_i(\mu + 1)^n = a_i \left[\frac{1}{q_i'} \sum_j \frac{p_j f_{ji}'}{\displaystyle\sum_l f_{jl}' a_l} \right]^n. \tag{19}$$

Equation (19) has the form of (15), and we can then write a recursive formula by the method of (16). Multiplying the numerator and denominator of the summation over j by c_j, we obtain our MLE–SS procedure:

$$a_i^{(k+1)} = \frac{1}{(\mu + 1)^n} a_i^{(k)} \left[\frac{1}{q_i'} \sum_{j=1}^{J} \frac{p_j f_{ji}}{\displaystyle\sum_{l=1}^{I} f_{jl} a_l^{(k)} + r_j c_j a_b} \right]^n \tag{20}$$

for the image pixels and

$$a_b^{(k+1)} = \frac{1}{(\mu + 1)^n} a_b^{(k)} \left[\frac{1}{\sum_j r_j} \sum_{j=1}^{J} \frac{p_j r_j}{\sum_{l=1}^{I} f_{jl} a_l^{(k)} + r_j c_j a_b} \right]^n \tag{21}$$

for the background pixel. At the end of the iterative procedures, the estimated pixel activities can be converted to equivalent counts by multiplication by the corresponding q_i of (2). Formulas (20) and (21) are similar to the EM formulas (13) and (14), except for the exponent n and the presence of the $1/\mu + 1$ multiplier. Since both sets of iterative equations solve the same problem and the SS formulas hold for the case of $n = 1$, we conclude that μ has to be zero, as it was for the EM case.

In practice, we find that, in order to prevent instabilities of theSS solution, we must estimate the parameter μ during the iterative procedure of (20) and (21), rather than set it equal to zero from the very beginning. For that purpose, renormalizing the estimates a_i multiplied by the corresponding correction factors q_i' at the end of each iteration to the known total number of detected counts is equivalent to calculating μ. After three or four iterations, the value of μ has practically gone to zero. The exponent n in (21) controls the speed of convergence. Values between $1 \leq n \leq 3$ have been found to result in stable solutions for simulated data, and for $1 \leq n \leq 2$ for real PET data, with a speed-up factor over the EM method approximately proportional to n. Some damped oscillatory behavior is sometimes observed in the first iterations with the larger values of n. The difference between images resulting from the SS and EM algorithms stopped according to the same rule is found to be insignificant.

C. Starting Point for the Image Estimates

As Shepp and Vanderbei discuss in [10], the EM-based MLE algorithm converges towards a possibly unique maximum, and the image obtained at convergence should be independent of the initial point. However, if the procedure is stopped before maximum likelihood, the initial estimate can strongly influence the result. We have shown in [17] that the MLE-EM iterative procedure uses the initial estimate as prior information (in a Bayesian sense) for the calculation of the results of the first iteration, those results as prior for the second iteration, and so on. In that situation, in the absence of any prior information on the image that is to be reconstructed, the only reasonable initial estimate is a uniform, featureless field. All the reconstructions to be shown in this paper were started with uniform pixel values, equal to the average activity in the image plane, as determined from the number of counts in the data and the corrected sensitivities q_i'. The initial estimate for the fictitious background pixel that we have used is a small number, like 0.01.

D. Calculation of Change in Likelihood

The calculation of change in likelihood between iterations is considerably easier than the calculation of the actual likelihood, and for many purposes it is sufficient. For the case of a reconstruction with a corrections vector c, it is easy to derive that

$$\Delta L(a) = \sum_{j=1}^{J} \left\{ h_j^{(k)} - h_j^{(k+1)} + p_j \log \left(\frac{h_j^{(k+1)}}{h_j^{(k)}} \right) \right\} \tag{22}$$

with the definition of h_j given in Section II-C, which explicitly is a function of the corrections c_j.

IV. FEASIBILITY AND STOPPING RULES

We have defined feasible images as those images that, if they were true radioisotope distributions in a patient, could have given the measured data by the Poisson process that governs the emission tomography process [18]. We have postulated that feasible images are important as possible representations of the unknown activity distribution because they are consistent with the data. Fundamentally, a feasible image is one whose *residuals* $p_j - h_j$ have the correct average magnitude, so that each h_j could be the mean of a Poisson distribution of which the corresponding p_j is one realization. Not all the residuals should be too small or too large. The concept and the tests for feasibility have been discussed in depth in [18]–[21]. We only describe here the main results of that work.

Statistical hypothesis tests indicate that the MLE can be stopped in a rather wide range of iteration numbers within which the resulting images are feasible. Below and at the lower edge of feasibility, the images have not achieved full contrast. At the upper edge and above feasibility, they have achieved full contrast, but noise has increased sufficiently in the regions of high counts that the images appear not to be useful representations of the actual source. Iterating past feasibility and filtering with a Gaussian kernel brings the images back into feasibility, although iterating to near convergence of the MLE requires filtering with a kernel which is sufficiently large to result in loss of resolution.

Stopping the maximization of a functional before reaching its maximum has been found equivalent to obtaining a maximum of a regularized version of the original functional for a large family of iterative reconstruction algorithms [22], [23]. Images obtained in that manner can be considered similar to reconstructions with a smoothness constraint, whose strength depends on the particular iteration taken [24].

We have obtained feasible images by a number of methods [25], [26],but we will discuss here only one method that is simple to use and provides a point of departure for the work to be presented below. We call this method the MLE–PF (for postfiltering). It consists of iterating the MLE procedure past the point where the test first indicates feasibility by an undetermined number of iterations, to a point where the image is usually still feasible, but it begins to exhibit excess noise. Slight filtering with a two-dimensional Gaussian kernel is then used to reduce that noise, without departing from the region of feasibility. The choice of a Gaussian kernel is not arbitrary:

it is the only linear kernel that will not create image detail that did not exist before filtering [27].

There are two problems with the above procedure: 1) the region of feasibility ranges over a large number of iterations and there is no clear criterion for choosing a specific image as being the most desirable one, and 2) the hypothesis test of [19] fails to indicate feasibility with real data from a PET scanner. We have ascertained that the effect is due to the error with which the transition matrix f can be known for a real tomograph [18], and we have proposed a modification of the hypothesis test that takes into account a certain amount of error in the matrix f by the inclusion of one adjustable parameter. The modified test has turned out to be not robust enough for practical use, as that parameter has been found to depend to some extent on the image being reconstructed. We have turned, instead, to a likelihood cross-validation method of stopping the MLE procedure, which we have found to be adequately robust with real data. The concept and tests for feasibility in simulated images have allowed us to accept the need for stopping the MLE process and to know some of the characteristics that images will have at different points of the feasibility region. The cross-validation stopping rule complements the above concept and provides an answer to the question of which feasible images should be taken.

Coakley [28] has recently published a procedure for stopping the MLE that is based on the statistical concept of cross validation. In the case of tomographic image reconstruction, the procedure can be described as follows. Consider that the projection data have been split into two halves, A and B. With the activity in the patient tissues in a relative stable condition, this can be done, for example, by taking two consecutive data sets, each with half the desired time interval. The MLE is used to reconstruct data set A and, at the end of each iteration, the likelihood of data set B is computed with respect to the A image estimate. Initially, this cross likelihood will increase, but at some point in the iterative process, it will stop increasing and actually decrease. The algorithm is stopped at the maximum of the cross likelihood. Then the roles of the two data sets are reversed and the new reconstruction process is stopped again at the maximum. The two resulting images are then added. At the maxima, the estimated images are most consistent with the other halves of the data in a probabilistic sense.

From the point of view of feasibility, we can understand the function of the cross-validation technique by considering that, as the cross likelihood increases, the reconstructed image is acquiring features that are common to the two data sets, and hence to features that are more likely to be true. When it starts to decrease, the reconstructed image is acquiring features that are specific to the statistical fluctuations of one data set, and the procedure should be stopped. This is qualitatively similar to the feasibility test that requires that the residuals do not become too small, i.e., that the reconstructed image does not follow the statistical noise in the data too closely. We have found the cross-validation technique to work reliably with simulated images and with a wide range of real data sets [29]. Invariably, the EM algorithm stops at a point in which noise in the areas of high activity has begun to grow visibly, corresponding to images at the upper end of the feasibility region.

From a practical point of view, the PET data that are available to us are not usually split in two independent halves. However, the split can be made by the process of "thinning." The process consists of taking each count in each projection bin and assigning it to one or other of the new data halves, according to the outcome of a random process of equal probability for each of the two outcomes. The thinning process results in two new Poisson distributed data sets, with means that are one half of the means of the original data set.

The above is a standard result in statistics, and it can be shown by considering the thinning of a discrete random variable, that is, a process whereby each unit of its value is retained with a probability p and eliminated with the probability $1 - p$, and each unit is retained or eliminated independently of all other units. If the random variable x is obtained by thinning the Poisson distributed variable y with the parameter λ, x is also a Poisson distributed variable with the parameter $p\lambda$. By the definition of thinning.

$$\text{Prob}(x = k) = \sum_{i=k}^{\infty} \text{Prob}(y = i) \cdot \frac{i!}{k!(i-k)!} p^k (1-p)^{i-k}.$$
$$(23)$$

Indeed, in order for x to take the value k, y should take any value i greater than or equal to k, and the thinning should retain k out of i units, which is given by the binomial expressions above. After substituting

$$\text{Prob}(y = i) = e^{-\lambda} \lambda^i / i! \qquad (24)$$

and simple algebraic manipulations, one can show that

$$\text{Prob}(x = k) = e^{-p\lambda} (p\lambda)^k / k! \qquad (25)$$

which was to be proved.

The interpretation that at the point of maximum cross likelihood each half image is most consistent with all the data is strictly true if the matrix f is known exactly, which is not the case with real data. With real data sets with a very large number of counts (50M, for example), we have found that a maximum may not be reached, although it is always reached in simulations. It appears that the errors in the matrix overshadow the differences between the two half images. For the more real common data sets (100K–10M counts), we have found that the maximum is always attained, with a more accurate matrix resulting in somewhat earlier stopping. We have also found that for data sets with presubtracted background, i.e., in which the Poisson characteristics of the data have been disturbed, the stopping point occurs at later iterations. These effects will be discussed in Section V.

In practice, we reconstruct real PET data by the "data splitting" method described above, carrying out the two partial reconstructions in parallel in the same computer program and stopping when one of the cross likelihood reaches a maximum. We have called that the MLE–CV (for cross validation) method of reconstruction, and it can be based on the EM or on the accelerated SS iterative formulas described above.

V. BIAS AND VARIANCE ANALYSIS

In the last few years, a number of groups have reported their findings on the characteristics of MLE, sieves, and modified MLE for PET, compared to FBP reconstructions, using a variety of criteria for stopping the iterations. Rosenqvist et al. [30] used a substantially modified MLE algorithm proposed by Tanaka [31] to equalize the frequency response and increase the convergence speed. They find that the modified MLE offers no particular advantage in quantitation over FBP, and conclude that possible advantages of statistically based algorithms may come from the use of prior information. For the highly modified MLE algorithm, they decided to stop at a fixed number of iterations. Holte et al. [32] investigated the behavior of the MLE algorithm principally in cold spots. They found that noise in regions of low counts was lower than in regions of high counts, and that by iterating to a point where the high-count regions exhibited substantial noise, quantitation of cold spots was excellent. Llacer and Bajamonde [33] found similar results to the above two papers by analyzing real data from the Hoffman brain phantom. Those authors found that, in order to obtain unbiased ROI estimates (in regions of high and low counts), it was necessary to continue some arbitrary number of iterations past the onset of feasibility. They also reported a preliminary form of the main results of the current paper. Liow and Strother [34] carried out a very careful analysis of the relationship among noise, quantitation, and the number of iterations. For the purposes of the work reported here, their most important finding was that images stopped at the onset of feasibility "terminate the iteration prematurely, leading to significant errors." This finding is in agreement with [33]. Herman and Odhner [24] have studied the comparative behavior of the MLE algorithm for specific tasks. Particularly interesting from our point of view is their finding that at a large number of iterations, the "noisy" MLE images give very accurate estimates of ROI uptake, although point estimates are not good. We shall comment on their finding further below. Finally, Politte and Snyder showed in [9] that, for simulated data in a time-of-flight PET environment, the introduction of a statistically correct procedure to account for absorption and randoms yields MLE reconstructions with sieve constraints that are essentially unbiased. The standard deviations (expected errors averaged over pixels) are from two to three times lower than in reconstructions made by the confidence-weighted algorithm (linear method) for both white and gray matter areas of the phantom for equal spatial resolution. Those findings are in apparent disagreement with the work reported here.

In this section, we will report on a detailed comparative statistical study of MLE, sieves, and FBP reconstructions of data from simulations and from fluoro-deoxiglucose (FDG) human brain PET studies of four normal subjects (nontime-of-flight PET). Section V-A describes the data, reconstruction, and analysis methods, Section V-B describes the results of the statistical analysis for simulations, and Section V-C describes the results obtained with real data.
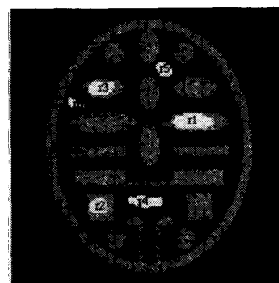

Fig. 1. Phantom used for the simulations. The high-intensity areas correspond to 100% activity; the interior background was set to 25%, with no activity in the outside background. The labeled white patches are the six ROI's chosen for the statistical analysis of reconstruction characteristics.

A. Data, Reconstruction, and Analysis Methods

1) Data: The data for the simulations were obtained by generating 24 independent sets of scan data based on the phantom of Fig. 1. The areas of high intensity correspond to 100% activity, the background in the interior of the phantom corresponds to 25%, with the exterior having no counts. For each of the data sets, 1.3 million (1.3M) counts were distributed by a random process into the different structures of the phantom, with means proportional to the designated activity. A second step distributed each of the counts by a multinomial probability to one of the 20 480 bins (320 angles, 64 bins per angle) of a single image plane of the CTI-831 tomograph. The multinomial probability was obtained from the "simple" transition matrix described in Section II-A. This two-step simulation allows us to measure the statistical characteristics of the source distribution. No attenuation, scattering, or background were included in the simulation.

For the tests with real data, scans from four subjects, $S1$–$S4$, scanned with a CTI-831 Neuro-PET system, have been used in this study. FDG data were taken from each subject in six sequential time intervals of 5 min, when the radioisotpe concentration in the brain was relatively stable, approximately 45 min after injection. The number of counts per data set was between 1.1 and 1.4 million per image plane, and the differences between the total number of counts in the six consecutive time frames of each subject were within 2–3%. From the 15-plane data sets for each subject, three were selected, one in the upper, one in the middle, and one in the lower brain. This corresponds to 18 data sets per subject. Data for subjects $S1$ and $S2$ were obtained with separate random coincidence files, while the randoms for $S3$ and $S4$ were presubtracted by the hardware. The customary absorption and gain corrections files were part of the data sets.

2) Reconstruction Methods: Both the simulation and the real data sets have been reconstructed by the FBP method with two different filters whose frequency pass characteristics are shown in Fig. 2. The Shepp–Logan filter has a cutoff frequency that has been selected through the experience of several years of clinical practice with FDG brain images, and the Butterworth filter exhibits improved response in frequencies in the region between 0.15 and 0.3 of the sampling frequency (3.18 cycles/cm), while reducing higher frequency noise. The
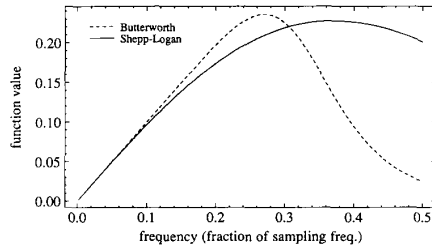
Fig. 2.   Shepp–Logan and Butterworth filter functions used in the FBP
reconstructions reported in this paper.

appropriateness of the parameters chosen for the Butterworth
filter was confirmed by analyzing the power spectral density
of a group of independent images of the same source distribu-
tion and determining the approximate cutoff frequency above
which there was mostly noise. Independent visual tests carried
out by physicians at the Department of Nuclear Medicine,
UCLA, have confirmed the appropriateness of the filter for
the brain data used. The filtering and the backprojection (BIN)
algorithms for the reconstructions were obtained from the
Donner Algorithms Package [35]. It has not been the intention
of the present work to compare the MLE–CV with the best
that can be obtained in linear methods (like a Wiener filtering
that has to be tailored to a specific class of images), but we
intend it to be a comparison with the methodology currently
used in clinical practice. Shepp–Logan filters are standard in
the software packages that are shipped with the scanners. The
Butterworth filter represents an easily implementable possible
improvement over that standard.

All the data sets were also reconstructed by the MLE–CV
method.The simulated data were reconstructed with our simple
matrix, which is exact for the reconstruction step. The patient
data sets were reconstructed with two different matrices, as
described in Section II-A. Details of the matrix generation by
MC methods are discussed in Appendix B. We consider the
work with the MC matrix still preliminary, but the results
obtained are already useful to illustrate the points that we
intend to make in this paper. For all the data sets, the MLE–CV
iterative process was stopped at the point of maximum cross
likelihood (iterations 35–45, except as noted below). The
resulting images were then filtered with a Gaussian kernel
to decrease the noise in the reconstructions. The size of the
kernels was obtained by requiring that the resolution of the
MLE-CV reconstructions be as close to that of the FBP
results as can be measured. The measurement of resolution
is discussed below. The resulting kernel sizes were $\sigma = 0.75$
pixels (0.1848 cm/pixel) for the simulations and for the data
of subjects $S1$ and $S2$, and $\sigma = 1.0$ for subjects $S3$ and
$S4$. The reconstructions for $S3$ and $S4$ always stopped at a
substantially larger number of iterations than in the case of
$S1$ and $S2$. Evidently, the late stopping is caused by the non-
Poisson nature of the data with presubtracted background. The
reconstructions for $S3$ and $S4$ were noisier and sharper than
those of $S1$ and $S2$, and could, therefore, be filtered with a
larger kernel. Real data reconstructions with the MC matrix
usually stopped a few iterations earlier than those with the
simple matrix. It appears, then, that the quality of a matrix will

also affect the stopping point of the cross-validation procedure.

It has been interesting to carry out the iterative process
to a large number of iterations and postfilter with a kernel
of sufficient size to bring the resolution to that of the FBP
results. This is equivalent to a method of sieves in which
the sieve and resolution kernels have the same size [3]. Note
that reconstructions at convergence of the MLE algorithm
are extremely sharp because there has been an attempt at
deconvolving the point response function of the tomograph.
The larger kernel filter brings the resolution to the FBP levels
and controls the high noise effectively. For both the simulated
and real data sets, we have obtained images at iteration 300
and have filtered them with kernels of $\sigma = 1.25$ pixels for the
simulations and real data of $S1$ and $S2$, and with $\sigma = 1.50$ for
$S3$ and $S4$. Real data sets were reconstructed in this manner
with the MC matrix only.

3) Measurement of Resolution: The definition of resolution
when discussing linear methods of reconstruction can be given
in a precise manner in terms of a modulation transfer function.
The same is not true when we are dealing with a nonlinear
method which may reconstruct different features in a different
way, depending on this intensity, location, shape, etc. In
addition, the images that we wish to compare will not register
exactly (different assumptions about point response function
variation over the image plane made for different methods of
reconstruction). Edge sharpness, however, if it is consistent
over a whole reconstruction, can be used to define resolution
when comparing images obtained by different methods. The
comparison will be meaningful if there is no artifactual ringing,
which would imply undesirable sharpness. In order to measure
edge sharpness under the less than optimal conditions that
we are facing, we have turned to an objective measurement
that comes from the field of vision and image analysis. We
can convolve the image of interest with a direction- and
position-invariant kernel that yields "edgeness" [36]–[38] or
edge strength. The kernel used for the convolution is given by

$$L_w = \left\{ \left(\frac{\partial G}{\partial x} \right)^2 + \left(\frac{\partial G}{\partial y} \right)^2 \right\}^{1/2} \qquad (26)$$

where G is a two-dimensional Gaussian with $\sigma_x = \sigma_y$ defining
the "scale" at which the operator acts. What the operation does
is measure the magnitude of the image gradient at a particular
point $[x, y]$ in the direction of maximum gradient. The mea-
surement is averaged with a Gaussian weight over a region of
dimensions given by the scale parameter. The operation is very
robust against noise in spite of taking derivatives because of
that averaging effect. After some preliminary experimentation,
a scale of 2.7 pixels was chosen as yielding clean edge strength
images by being sufficiently insensitive to image noise. An
edge strength image corresponding to the phantom of Fig. 1
is shown in Fig. 3. The intensity displayed is proportional to
edge strength. The postfiltering step after our reconstructions
effectively removes ringing, so that the results of edge strength
are meaningful. This will be demonstrated below.

4) Bias Analysis Methods: Bias has awell-defined meaning
in statistics and, in the imaging context, could be described as
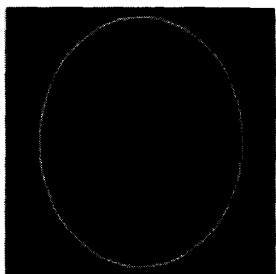differences between the expected value of reconstructed pixel

Fig. 3. Edge strength image corresponding to the phantom of Fig. 1. Brightness at a point is proportional to the magnitude of the gradient at that point in the direction of maximum gradient.

Fig. 4. Samples of ROI's selected for the analysis of statistical characteristics of real data. One gray matter and one white matter region have been selected from each of the three planes of each of the four subjects used in the study. The images shown are the mean images from reconstructing six independent data sets by the FBP–Butterworth method.

intensities and the correct values. In the case of tomographic reconstructions, in which filtering of high frequencies is a necessity, the above definition of bias is not practical since all edges in the image would show bias. A more practical definition that we devised in [33] is that of "regional bias" as the difference between the expected average value in extended uniform regions, away from edges, and the correct average values for those regions. In the case of simulated reconstructions, we have selected six ROI's ($r1$–$r6$) for the regional bias measurements. They are shown in Fig. 1 as labeled areas, both in the hot and in the cold regions of the image. The average of the 24 reconstructions for each of the six ROI's can then be compared to the known phantom values for each method of reconstruction as a measurement of bias.

For real data, we do not know what the correct region values are. As discussed in Appendix A, even the FBP reconstruction, in which all reasonably possible corrections have been incorporated by the standard procedures used in the clinic, a certain amount of bias can be expected due to the way the correction factors are obtained. However, the FBP reconstructions can be used as a reference for the MLE and sieve reconstructions for the purposes of discussing their results. For each of the groups of six data sets obtained from a given brain plane of each patient, two ROI's have been chosen, one in the gray matter and one in the white matter. Examples of the selections are given in Fig. 4. The mean values of each ROI in each reconstruction and the standard deviation from that mean have been measured, yielding 24 sets of values for each method of reconstruction. In order to simplify the presentation, the mean values have been processed so as to yield a set of global numbers giving one average ROI value for all white matter measurements and one ROI average for all gray matter, for each of the methods of reconstruction. There is a lack of registration between the reconstructions by different methods at the points where some of the ROI's are defined. By moving the ROI's by one pixel in different directions, it has been ascertained that the regional bias measured is not substantially affected by that lack of registration.

5) Estimation of Standard Deviation in ROI Intensity Measurements: For the case of simulated data, the procedure is straightforward, and the results can be plotted along with the estimated mean ROI intensities by means of error bars. For real data, however, there can be variability of radioisotope intensity in different structures of a human brain at different
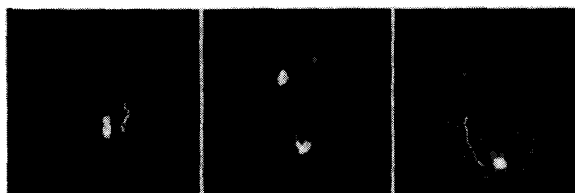
times during the six 5-min intervals of the independent measurements. In addition, we cannot pool all the measurements because of different levels of activity in different patients, different structures chosen as ROI's, and different tomograph sensitivities at different planes (odd versus even planes in a multiplane tomograph). A preliminary analysis on a plane-by-plane and patient-by-patient basis has been carried out, disregarding the possible problem of temporal changes in activity. Then, a more sophisticated method based on analysis of variance has been used. The results of the two methods are consistent. We shall show an example of the first method, and report on the latter method in more detail since it takes into account all the variations that we encounter.

The analysis method used is based on the "components of variance model" (see [39], for example). It assumes that the measured ROI activity data can be tabulated as a two-dimensional matrix with elements x_{ij}. One matrix is generated for each method of reconstruction and for each of major ROI divisions: gray and white matter. For each matrix, index i corresponds to an independent instance of the measurement. Thus, $1 \leq i \leq 6$ for the human data. Index j ranges over all the different patients and image planes that enter into the study. The model assumes that x_{ij} can be represented as the sum of three independent, normal variables:

$$x_{ij} = u_i + v_j + w_{ij}. \tag{27}$$

The variable u_i depends only on the variability of measured activity in the ROI as a function of the time. For a large number of ROI's from different planes and patients, we can consider it a normal random variable with a mean μ_a and standard deviation σ_a. Variable v_j depends on the particular plane and patient. That component of activity will have different values, and if we introduce a sample large enough, one can define a mean and standard deviation for that variable as μ_b and σ_b, respectively. Finally, w_{ij} is the true brain activity whose standard deviation σ_c from some mean μ_c we want to evaluate. It is caused only by the statistical fluctuations in the activity measurements.

From the data of a matrix x_{ij}, there is a staightforward procedure to calculate the standard deviations σ_a, σ_b, and σ_c [39]. The corresponding means are of no interest. The resulting variances and standard deviations can be shown in a table.

6) Pixel-by-Pixel Standard Deviation Analysis: Continuing the preliminary work by Llacer and Bajamonde in [33], we have obtained two-dimensional histograms of the number of

pixels that have a certain standard deviation from their mean versus number of counts in that pixel. The procedure is the following.

For a set of independent reconstructions of a particular image (24realizations for simulated data, six realizations for each plane of each patient for real data), a mean image is generated. Then, a two-dimensional histogram is generated in which the abscissa corresponds to number of counts in a pixel of the mean image, and the ordinate corresponds to the standard deviation over the ensemble of reconstructions for that pixel with respect to the same pixel in the mean image. The gray scale corresponds to the logarithm of the number of pixels in each bin of the two-dimensional histogram. The logarithm is taken in order to be able to see the large range of number of pixels in the different histogram bins in a single display. The two-dimensional histogram shows the relationship between average counts in a region of the image and the standard deviation that can be expected in the reconstruction of that region, or "expected error."

B. Results of Measurements with Simulated Data (Matrix Exactly Known)

1) Resolution Measurements and Postfiltering Kernel Size: Fig. 5(a) shows profiles taken vertically through the mean image of the 24 independent reconstructions obtained from the phantom of Fig. 1, and from the mean source image. FBP results with the Butterworth filter of Fig. 2, MLE–CV with postfiltering by $\sigma = 0.75$, and sieves results (300 its), $\sigma = 1.25$ are shown. The profiles correspond to structures approximately in the center of the left half of the phantom. It is shown that ringing is well controlled in all cases. Fig. 5(b) shows the corresponding cut through the edge strength image obtained from the same mean images as above for the three methods of reconstruction. For the majority of structures, edge strength is very similar for all the reconstructions, although the outer edges are sharper for the MLE–CV than for the other methods. The reason for this effect is not understood at present. FBP results with the Shepp–Logan filter of Fig. 2 show significantly lower edge strength than those for the Butterworth filter.

2) ROI Bias and Standard Deviation Results: Fig. 6(a) shows the average over the 24 reconstructions of mean intensity value over the high-activity ROI's ($r1$, $r2$, and $r3$) for a variety of reconstruction methods and for the original source images. The line across the graph corresponds to the true known mean value. Fig. 6(b) shows similar results for the low-activity ROI's ($r4$, $r5$, and $r6$). The standard deviations of the mean intensity values have also been plotted by error bars. For each method of reconstruction, averaging the mean intensity values over regions and adding the corresponding variances yields the results of Table I.

Three initial comments will be made here on the results.

a) The MLE–CV unfiltered results and the unfiltered results at 300 iterations ($\sigma = 0$ points in Fig. 6(a) and (b) and in Table I) are the important bias results for the MLE methods. All filtered results (including FBP) may include some edge effects which bring the regional results down for the high-intensity ROI's and up for the low-intensity ROI's, unless the
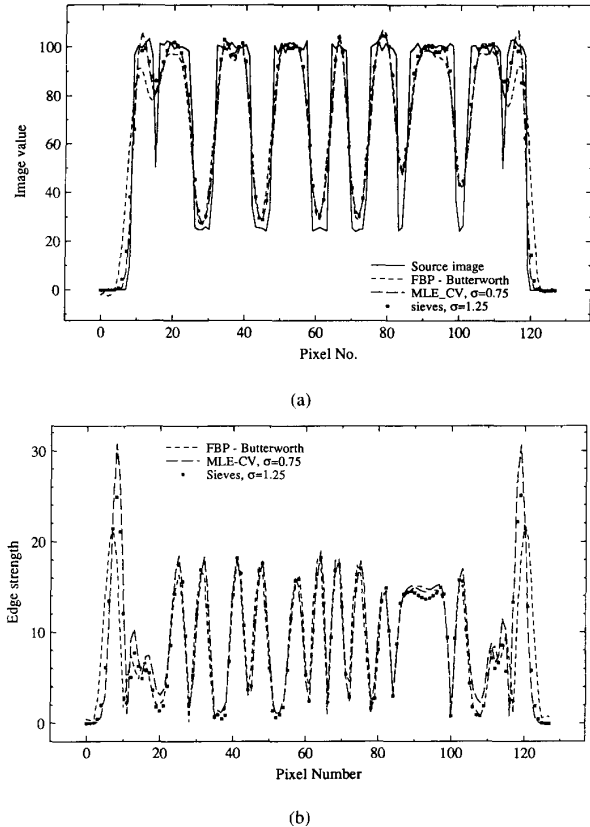

(a)

(b)

Fig. 5. (a) Vertical profile of a cut through the source and reconstructions of the phantom of Fig. 1, approximately at the middle of the left half. The mean image over 24 independent data sets was first obtained and the profiles taken on that mean image. Good control of the edge artifact is exhibited in all the reconstructions. (b) Edge strength for a cut corresponding to the same location of (a). It shows a good match of edge strength, except at the edges, indicating similar resolution for the three methods of reconstruction.

TABLE I
GLOBAL MEAN AND STANDARD DEVIATIONS FOR
THE DIFFERENT KINDS OF SIMULATED IMAGES

Image Type	High intensity ROIs (counts)			Low intensity ROIs (counts)		
	mean	bias	stadev	mean	bias	stadev
Source	238.79	0.0	1.83	59.69	0.0	1.07
FBP - Butter.	235.81	-2.98	5.67	59.78	0.09	6.77
FBP - Shepp-L.	234.44	-4.35	5.54	60.46	0.77	6.49
MLE - σ=0.0	244.00	5.21	6.77	58.19	-1.50	4.78
MLE - σ=0.75	242.21	3.42	5.98	59.01	-0.68	4.47
Sieves, σ=0.0	239.12	0.33	9.33	59.13	-0.56	7.46
Sieves, σ=1.25	236.43	-2.36	5.69	60.84	1.15	5.57

size of the ROI's is made small compared to the size of image features. Since linear filtering cannot change the mean value of a region except by inclusion of edges, it is clear that the bias of a reconstruction has to be estimated in its unfiltered form. b) The MLE–CV, $\sigma = 0$ results show some bias (2.1%and 2.5%, high and low ROI's, respectively), while at 300iterations (sieves, $\sigma = 0$ data), the bias has dropped considerably (0.14% and 1%). It appears, then, that the MLE carried to a large number of iterations will lead to an unbiased estimation. The standard deviation of the results at 300 iterations is large,
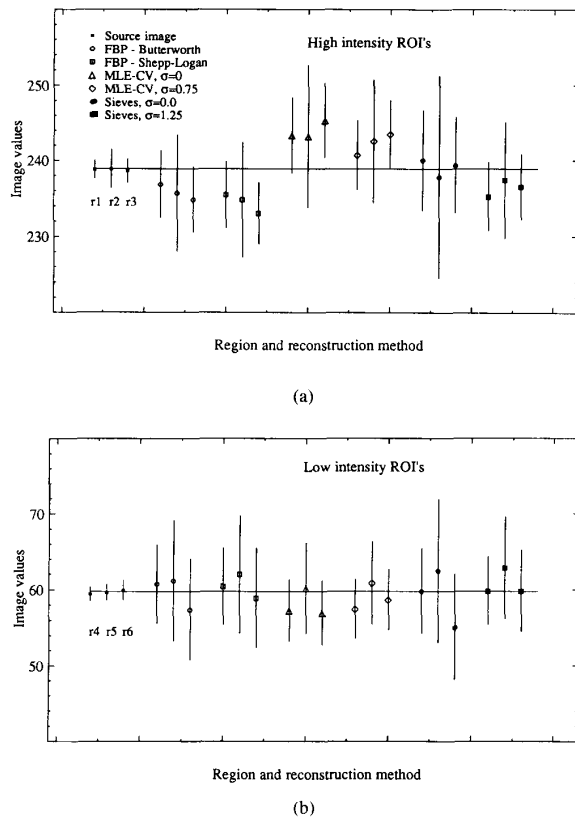
(a)

(b)

Fig. 6.   Mean and standard deviation results of measuring the average activity in ROI's of the 24 independent phantom reconstructions for a variety of reconstruction methods. (a) is for high-intensity ROI's, (b) is for low-intensity ROI's.

however, and in any one single measurement, the value of having an unbiased estimator may be offset by having a large variability.   c) The size of the filter kernel determines the standard deviation of a measurement. For equal resolution, the MLE–CV and sieves show standard deviations which are similar to FBP results in the high-intensity ROI's. In the regions of low intensity, the MLE–CV shows a lowered standard deviation with respect to FBP, approximately by a factor of 0.66. The method of sieves results in a lowering by approximately 0.83.

3) Pixel-by-Pixel Expected Error Histograms: Fig. 7 shows pixel-by-pixel expected error histograms for the ensemble of 24 independent reconstructions of the simulation. We show FBP with the Butterworth filter (top), with the MLE–CV, $\sigma = 0.75$ (center) and sieves, $\sigma = 1.25$ (bottom). Each histogram shows three main regions: a) the rightmost "cloud" of pixels, corresponding to the regions of high activity in the images (100%); b) a second "cloud" to the left, corresponding to the background regions inside the phantom (25%); and c) farthest to the left, points corresponding to the background outside the phantom (0%). Selecting the MLE filters to yield approximately the same resolution as the FBP reconstruction has resulted in high-count regions in the three histograms that indicate a similar expected pixel-by-pixel error. The
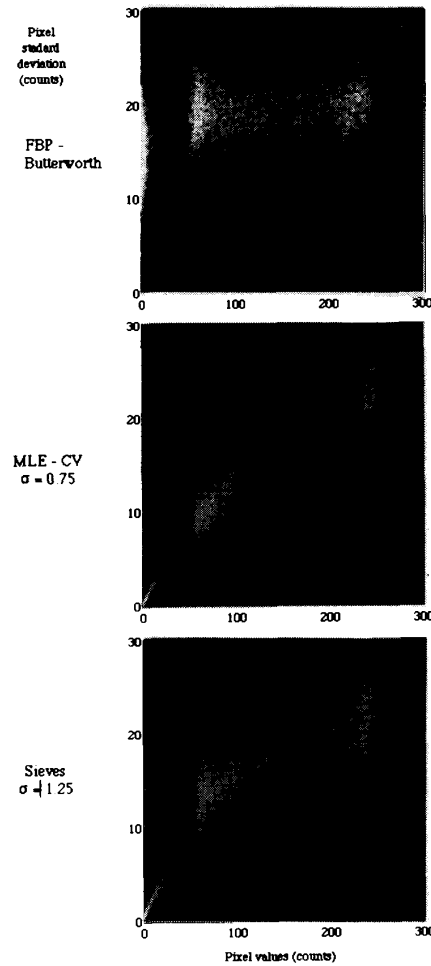
Fig. 7.   Pixel-by-pixel standard deviation versus pixel value histograms (scatter plots) for the mean image of the 24 independent phantom reconstructions for three methods of reconstruction. The gray scale corresponds to the logarithm of the number of pixels that fall in one bin of the histogram. To allow a reasonable printed reproduction, the gray scale has a substantial pedestal and is compressed.

background inside region, however, is different for the three cases: the FBP results indicate a similar expected error to that of the high-count region, the MLE–CV shows a substantially lower expected error, while the sieves show an intermediate error. Finally, external background has an expected error going towards zero for the MLE reconstructions, while it remains high for the FBP reconstructions.

The even distribution of noise in different regions in linear methods of tomographic reconstruction has been explained by Barrett and Swindell [40] and it can be expected for any linear filter. The MLE reconstructions exhibit a noise behavior that is local, i.e., it depends on the average number of counts in a particular region. There appears to be no analytic explanation for this behavior at the time of this writing. The pixel-to-pixel expected errors of the three methods shown are, evidently, the underlying cause for the ROI standard deviation values of Table I.

FBP - Butterworth

MLE-CV, simple matrix

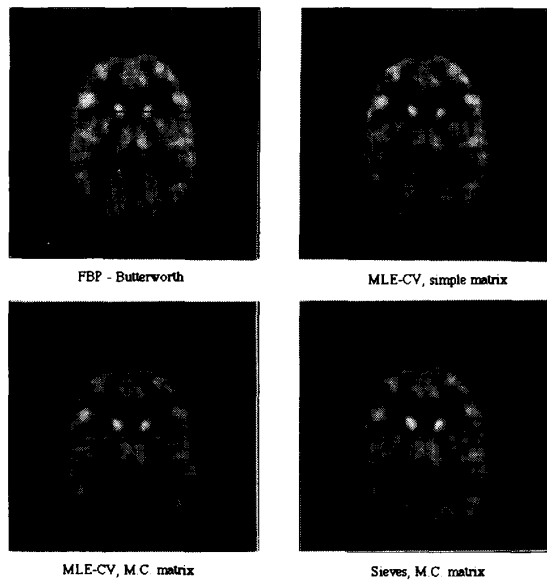MLE-CV, M.C. matrix

Sieves, M.C. matrix

Fig. 8. Mean images obtained from six independent data sets of a midbrain plane of one of the subjects for four different methods of reconstruction. The black line on the top left image corresponds to the cut used for analyzing resolution. Normalized to maximum brightness in each image.

C. Results of Measurements with Patient Data

1) Resolution Measurements: Following the procedure established for the simulation data, we show here that the choice of filters established in the simulations also results in resolutions for the MLE–CV and sieves that are at least equal to that of FBP–Butterworth reconstructions. Fig. 8 shows the mean images from the six independent data sets for subject $S1$, at a midbrain plane, for four different methods of reconstruction. The reconstructions with the MC matrix exhibit a slight degree of "fuzziness" when compared to the MLE–CV image obtained with the simple matrix. As shown in Appendix B, there is some noise in the matrix elements which we believe causes the imperfections noted.

Fig. 9(a) shows a cut through the line crossing the FBP image of Fig. 8. The four reconstructions were normalized to 100% intensity at the leftmost bright area near the skull, just above the line. The two central bright features show approximately 102% intensity for the FBP and MLE–CV, simple matrix, while the results with the MC matrix are higher by approximately 16%. This is due to a bias that results from using the MC matrix together with the standard corrections vector from the CTI-831 tomograph, as discussed in Appendix A. Fig. 9(b) shows the edge strength corresponding to the same line as Fig. 9(a). Since edge strength is proportional to the intensity of a feature [37], [38], the edge strengths for the two central features in the MC results are also higher by at least the same amount as the features themselves. Then, for most of the features that are fully crossed by the cut line, all the MLE–CV and sieves results show equal or higher edge strength than FBP, while features that are crossed by the cut line at "skirts" show equal or lower edge strength. There is more noise in the mean images for real data than there was

for the simulations (approximately 8.4M counts in the set of data versus 33M, respectively), and therefore, the edge strength results for real data are less consistent than in the simulation. We can conclude, however, that the MLE or sieve results are of similar sharpness as the FBP results.

2) ROI Bias and Standard Deviation Results: Table II(a) and (b) presents a sample of ROI data obtained for one plane of one of the subjects, for white and gray matter brain regions, respectively. They correspond, specifically, to the regions of the middle image of Fig. 4 for patient $S4$. For each of the six independent data sets, they show the measured means over the ROI for the different reconstruction methods. The last two lines of the tables contain the means over the corresponding columns and the standard deviation for the six measurements. There is a total of 12 of those tables pairs, and in order to simplify the presentation, we have pooled all the white matter data and the gray matter data into the global averages shown in Table III. All the MLE and sieve statistical results for real data have been evaluated with filtered images for consistency in comparisons with FBP, which is, of course, filtered.

It is found that the FBP method and the MLE–CV using the simple matrix show very good agreement in the mean values of both white and gray matter regions. Both methods of reconstruction assume perfect detectors with a uniform response across the width of a tube joining a pair of detectors. As explained in Appendix A, the gain corrections applied to the reconstruction process as part of the multiplicative vector c (Section II-A) are applied to both reconstructions similarly, and both can be expected to have a small similar
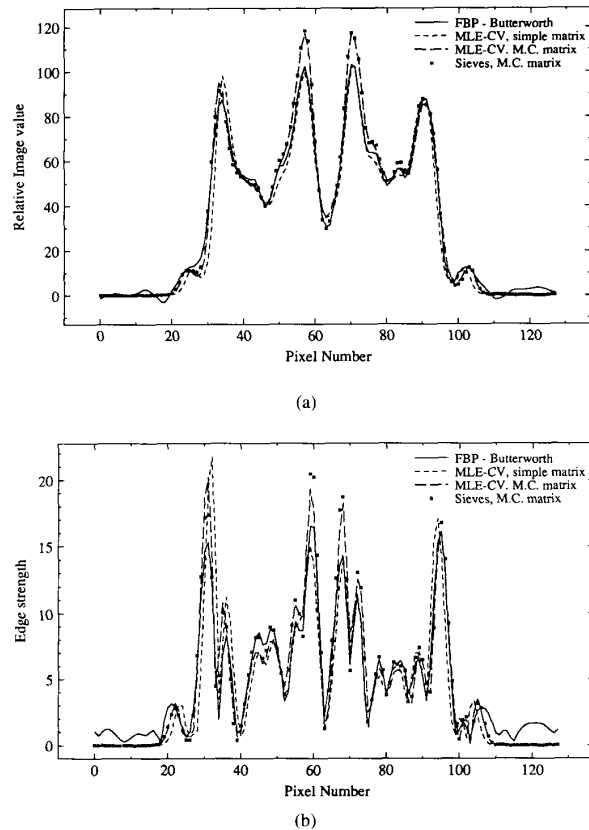
TABLE II
(a) Sample of Gray Matter ROI Data for Six Independent Data Sets, Average Over ROI (Counts) (b) Sample of White Matter ROI Data for Six Independent Data Sets, Average OverROI (Counts)

Data Set No.	FBP - Butterworth	MLE-CV simple matrix	MLE-CV M.C. matrix	Sieves M.C. matrix
1	2106.7	2101.3	1960.5	1950.6
2	2203.6	2199.6	2059.3	2048.1
3	2097.8	2115.8	1972.7	1952.9
4	2108.0	2156.3	1973.9	1961.5
5	2114.0	2135.9	1978.3	1963.8
6	2054.0	2062.2	1916.3	1907.6
mean	2114.0	2128.5	1976.8	1964.1
stadev	48.9	47.28	46.39	45.95

Data Set No.	FBP - Butterworth	MLE-CV simple matrix	MLE-CV M.C. matrix	Sieves M.C. matrix
1	846.8	864.5	837.7	837.2
2	860.6	872.33	851.1	855.9
3	812.6	952.3	822.4	830.9
4	706.5	757	727.9	729.4
5	768.2	788.2	790.5	791.2
6	702.5	740.2	730.9	724.0
mean	782.86	812.42	793.4	794.7
stadev	68.58	57.9	53.5	56.8

TABLE III
Global Mean and StandardDeviation from
Components of Varance Model (Counts)

Reconstruction Method	Grey matter ROI's		White matter ROI's	
	mean	stadev	mean	stadev
FBP - Butterworth	2283.3	67.5	918.8	61.3
MLE-CV, simple matrix	2313.8	67.0	947.3	43.2
MLE-CV, M.C. matrix	2163.9	63.5	966.2	41.1
Sieves, M.C. matrix	2141.8	62.5	927.6	47.9

(a)

(b)

Fig. 9. (a) Cut through the images of Fig. 8 at the level indicated by the black line. (b) Cut through the corresponding edge strength images.

Fig. 10. Pixel-by-pixel standard deviation versus pixel value histograms for the mean image of six independent data sets for the images of Fig. 8.

bias, which we cannot measure. The results obtained with the full MC matrix are substantially different from the FBP results, particularly for the gray matter regions. This is expected by the fact that the MC matrix includes the true response of the detector system, which is also included in the vector *c*. Thus, geometric corrections are applied twice, resulting in the observed bias. It has not been found possible, *a posteriori,* to modify the corrections vector to remove its contribution to the bias. For the clinical use of an MC matrix, corrections will have to be measured differently (Appendix A).

For the ROI standard deviation measurements, Table III shows the results of analyzing the measured data by the method of components of the variance. The value σ_c of the model described in Section V A-5) is the one we are interested in since it corresponds to the statistical effects of reconstruction methods. The values of σ_a are small (approximately 20 and 10 for gray and white matter, respectively), indicating that the effects due to temporal variation in the radioisotope activity in the patient are small. The values of σ_b are large and reflect the variability between patients and planes, but are of no concern here. The global values of σ_c, shown in Table III in the "stadev" columns, are in basic agreement with those found in the phantom measurements: the standard errors in measuring ROI intensity in gray matter are very similar for all the reconstruction methods studied, with the best values being
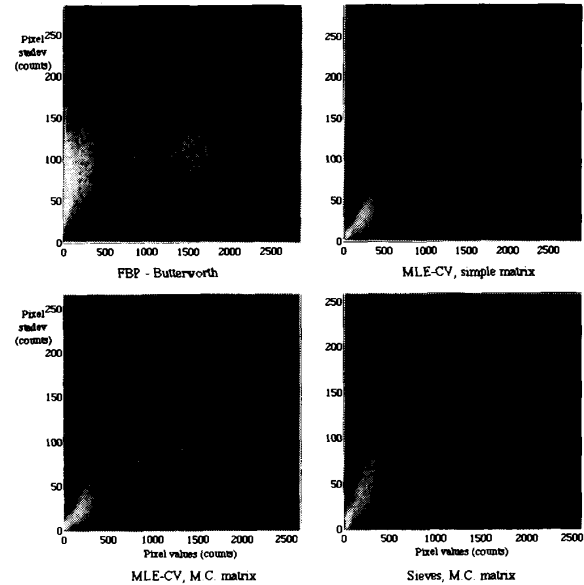
obtained with the MC matrix (MLE–CV and sieves). For white matter, the MLE methods show a substantial improvement in expected error, with the MLE–CV MC being the best and sieves MC being an intermediate with respect to FBP.

3) Pixel-to-Pixel Standard Deviation Results: Fig. 10 shows two-dimensional expected error histograms for patient data, similar to those of Fig. 7. We show FBP–Butterworth, MLE–CV ($\sigma = 0.75$) with the simple matrix, with the full MC matrix, and the sieve method, also with the MC matrix. The histograms shown correspond to the ensemble of six reconstructions for the same patient and plane as in Fig. 8. The results obtained in all cases are very similar to those of the simulations, and the comments made in Section V-B.3) also apply here. One should add the observation that the histograms for MLE–CV with the simple matrix and those for the MC matrix are nearly identical.

4) Effect of Background Subtraction Methods: The observation that the cross-validation stopping rule stops later for real data sets in which the background has been presubtracted raises the question of whether images reconstructed from data sets with separate random background data yield better images than those with presubtracted background. In order to answer this question, the data sets for subject $S1$ have been reprocessed by subtracting the random background (setting negative numbers to zero) and reconstructing accordingly. The results obtained show that regional bias and pixel-to-pixel standard deviation at the new cross-validation stopping point are practically indistinguishable from the results obtained originally if a somewhat larger filter kernel is used. The only conclusion that one can extract from this experiment is that, in cases with relatively low background (on the order of 5% or less), reconstruction from data in which the background data were obtained separately from the scan data allows for a

more "elegantly" defined stopping point, but the results are not significantly different from reconstructions with presubtracted background. This may not be the case with data sets in which the background fraction is substantially higher.

5) Effect of Speeding Up the Reconstruction by the Successive Substitutions Method: The reconstructions reported above have been carried out by using the MLE iterative formulas based on the EM algorithm, (13) and (14). The data sets for subject $S1$ have also been processed by the faster iterative formulas derived from the successive substitutions (SS) algorithm, (20) and (21), with the simple matrix. Although in simulation experiments a speed-up parameter of 3.0 could be used with no apparent instability, we have found that, with real data, we should stay below $n = 2.0$ for reliably stable solutions. This results in reconstructions of the $S1$ data sets that are visually indistinguishable from those of the EM algorithm. ROI mean value results differ from the EM case usually in the third digit, while the expected error histograms are not significantly different from those of the EM. The SS results have been obtained in approximately one half the number of iterations needed by the EM method.

VI. DISCUSSION AND CONCLUSIONS

In this section, we will summarize the findings of Section V, compare our results with those of other workers, and present the conclusions from our study.

From simulated data (matrix known exactly), we have observed the following.

1) The MLE procedure appears to lead to an unbiased estimator for high iteration numbers.

2) Stopping according to the cross-validation rule leads to some bias (2–2.5% for the example shown).

3) The expected error in a single measurement of an ROI average is substantially higher for the unbiased estimator than for the MLE–CV case, both for high- and low-intensity ROI's.

4) Filtering the unbiased estimator (leading to the method of sieves) or the MLE–CV reduces the expected error in a single measurement, but depending on the size of the ROI, the underlying anatomical structures, and the size of the filter kernel, it may lead to bias.

5) For equal resolution, the filtered MLE–CV and sieves pixel-by-pixel standard deviation and the corresponding standard deviation in the estimation of an ROI mean are very similar to those of the FBP methods in regions of high intensity. For regions of low intensity, the filtered MLE–CV shows substantially lower standard deviation than the FBP, by a factor of approximately 0.66, while the sieves method is lower by approximately 0.83.

For real data, the following is observed.

1) The FBP results can be expected to have some relatively small unspecified bias due to the way the correction measurements are carried out.

2) In MLE reconstructions, bias will be dependent on the characteristics of the transition matrix and of the quality of the corrections vector. For the filtered MLE–CV method using our simple matrix, global ROI means show small differences with respect to the FBP data (at most 3% in white matter). The use

of an MC matrix along with a corrections vector tailored to the FBP method of reconstruction leads to substantial differences with the FBP results (up to 6–7%).

3) The pixel-by-pixel and corresponding ROI standard deviations in gray matter for the MLE–CV simple matrix reconstructions are similar to the FBP results. The use of the MC matrix for both the MLE–CV and sieves reconstructions yields significantly smaller standard deviation values.

4) The standard deviation results in white matter follow the same pattern as the simulated results. The improvement in ROI standard deviation in the MLE–CV simple matrix case is by a factor of 0.7, dropping to 0.67 for the MLE–CV MC matrix. The improvement for the sieves and the modified MC is approximately 0.77.

5) A biased MLE reconstruction by having used an inappropriate corrections vector c is an effect that is decoupled from the ROI or pixel-by-pixel standard deviation results. The latter are affected by the accuracy of the matrix and the statistical characteristics of the data which, in turn affect the stopping point of the MLE–CV method.

For the quantitation of ROI uptake in regions of high uptake, our results coincide with those of Rosenqvist *et al.* [30], who found that there is no significant advantage in using their highly modified MLE algorithm with respect to FBP. Our results indicate a possible exception when using an MC matrix well matched to the tomograph. The results of Holte *et al.* [32], Llacer and Bajamonde [33], Liow and Strother [34], and Herman and Odhner [24] are all verified by our work. We find our results difficult to reconcile with those of Politte and Snyder [9]. For time-of-flight (TOF) PET, at equal resolution, those authors have found a very strong improvement in expected error with the sieves method with respect to their linear method of reconstruction. The improvement is for both gray and white matter regions in a simulation. The most important difference between their work and ours is in the nature of the problem. The TOF–PET problem is considerably better posed mathematically than the standard PET [41]. The spectrum of eigenvalues for TOF–PET is considerably more uniform than that of the standard PET. It is not clear, though, how this difference could account for the differences in reconstruction statistics between their results and ours.

In conclusion, it appears that our work, together with the work of the above referenced workers, clarifies the issue of what are the characteristics of MLE images that can make that method of reconstruction superior to FBP: when properly handled, MLE images provide measurably better quantitation in regions of low intensity in the presence of regions of high intensity in the same image. One would also expect that those characteristics would lead to better lesion detectability in regions of low uptake. A clinical ROC study is currently being conducted to test that idea. It is felt that a positive outcome of such studies would help establish the MLE methods as the preferred methodology for PET reconstructions.

In choosing between the MLE–CV and a sieves method, one is faced with a tradeoff between the shorter computation times for the MLE–CV (even when reconstructing split data sets concurrently), a basic slight bias and a low standard

deviation in ROI measurements, and longer computation times, a basically unbiased estimator, and higher standard deviation for the sieves method. It should be mentioned that the method of sieves used in the current work is only for the case of the sieve and resolution kernels having the same size ($\sigma_s = \sigma_r$). In earlier work, Veklerov and Llacer [26] examined the range of values for the two kernels leading to feasible images. It was determined that the condition $\sigma_s \approx \sigma_r$ was necessary in order to obtain images at the upper end of the feasibility range, with good control of the edge artifact [3], just as the MLE–CV method does. For that reason, we consider our current choice of parameters for the sieve method appropriate.

With the availability of cost-effective computing power, we feel that the MLE method of reconstruction, particularly the MLE–CV, is now ready to be tested in the PET clinic. We will start with the "simple" matrix and progress to using MC matrices for the direct and the cross planes of a multiplane tomograph after sufficient preparation and validation work. The principal author of this paper has made available a package of programs in the C language for the generation of the simple matrix, simulation, and reconstruction of simulated and real PET data.

VII. APPENDIX A
RELATIONSHIP BETWEEN TOMOGRAPH CALIBRATIONS AND TRANSITION MATRIX GENERATION

The process of tomograph calibration, which is carried out frequently in a PET installation, is done, in concept, according to the following procedure. A flat source that can rotate about the center of the tomograph, as seen in Fig. 11, is placed at some angle ($\theta = 0$ for the example in the figure) with respect to the x and y coordinate frame shown. The response of pairs of detectors that form tubes perpendicular to the source is recorded for a fixed period of time. Then, the source is successively rotated so that all the detector pairs are recorded. This procedure would be impractical, and only a relatively small number of equiangular source positions is chosen, with appropriate geometrical corrections carried out on the measured data [8]. Using the example of the $\theta = 0$ angle of Fig. 10, we observe that the results of a measurement m_j on the jth tube consist of the product of two components:

$$m_j = I_j g_j \tag{28}$$

where I_j is the integral of the response of the jth tube to all the pixels in the source that are within the field of view of the corresponding detector pair, assuming uniform detector crystals and photomultiplier tubes, and g_j is a factor on the order of 1.0 that corresponds to the actual relative sensitivity of the particular pair of detectors involved. Assuming no absorption, the multiplicative correction factors c_j used in the reconstruction are calculated by taking the mean $\bar{m}$ over all the measured m_j's and defining

$$c_j = \bar{m}/m_j. \tag{29}$$

Let us now consider the set of transition matrix elements f_{ji} that correspond to the tubes and pixels of the above example.
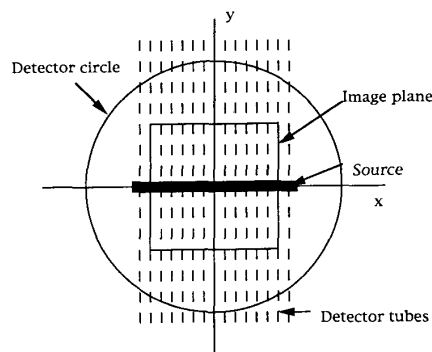

Fig. 11. Schematic description of the method used to obtain detector gain corrections in a tomograph.

When the corrections factors c_j are used to correct the matrix elements as in (1), complete rows of the matrix are corrected, including not only the set of pixels at $y = 0$, but also sets of pixels for those same tubes corresponding to source positions at $y \neq 0$. The correction would be good if the measurements m_j were independent of the y coordinate of the flat source. A linear or nonlinear reconstruction method that assumes perfect detectors (FBP or MLE with our "simple" matrix) can be expected to have some bias due to that error in the calibration. Initial simulations, described in Appendix B, show that these effects are not negligible.

If a transition matrix could be calculated perfectly by taking into consideration the actual geometry, detector response, detector Compton scattering, positron range, etc., the matrix elements f_{ji} would already include the correct values I_j, and only the elements g_j should enter into the correction factors c_j. In the specific case of the real data reconstructions described in this paper, the available correction factors from the CTI-831 tomograph contain the values I_j, and application of the corrections together with an MC calculated matrix can be expected to lead to substantial bias, as has been observed.

For reconstruction by FBP, the scan data resulting from a patient measurement are partially rebinned, following a prescription detailed in [8], for the purpose of compensating for gamma-ray crystal penetration in peripheral tubes. This rebinning should also be done when reconstructing with our simple matrix. This would result, however, in changing the Poisson nature of the data, and we have chosen not to carry it out. The effect is to obtain a slightly distorted, smaller image. When using the MC matrix, rebinning is not necessary.

VIII. APPENDIX B
MC MATRIX CALCULATION

Reference [7] discusses the generation of an MC matrix for PET and its application to reconstruction of real data from the ECAT-III tomograph. For the purpose of providing a more accurate matrix than the "simple" matrix used in the present paper, we have adapted the methodology described in [7] to the CTI-831 tomograph. The process is carried out in three parts. 1) A Monte Carlo method is used to obtain detector responses for positron emission sources located in a series
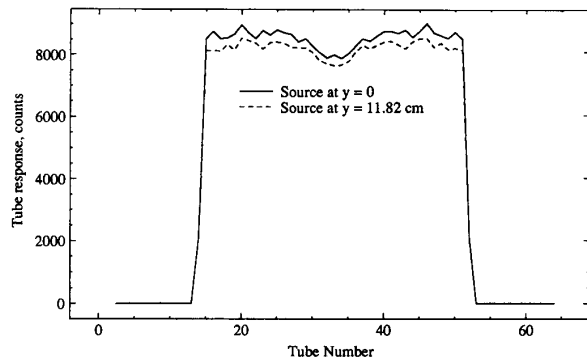
Fig. 12. Monte Carlo calculated response of the detector system of Fig. 11 to a flat source at $y = 0$ and at the $y = 11.82$ cm edge of the image plane. The lack of smoothness indicates noise in the process of obtaining the matrix.

of points forming a raster in a quadrant of the image plane, only for tubes that are perpendicular to the raster lines and for those corresponding to the next projection angle. 2) The obtained responses are fitted to a piecewise linear function, convolved with a Gaussian. The parametrized responses are smoothed versions of the discrete response results of the first part. 3) Finally, the matrix is calculated by suitable rotations of the available parametrized functions. For the MC part of the matrix generation, we have used 50 000 gamma pairs from a single source point directed into a fixed solid angle that will always cover the detector element pairs, independent of the position of the emitting point source. Positron range for ^{18}F (in FDG) was used in the calculations, although the magnitude of this effect is negligible for the low-energy positrons emitted by that radioisotope. After the complete generation, some tests have been carried out to establish the correctness of the matrix. Fig. 12 shows the calculated response of the detector pairs of Fig. 11 to a flat source along the x axis for both the cases of $y = 0$ and $y = 11.82$ cm (at the top of the image plane). Except for some irregularities that we believe are due to noise in the MC phase of the calculations and, possibly, to errors in the fitting functions process, the shape and magnitude of the response variations for the $y = 0$ case are almost identical to the responses for the same situation presented by Hoffman et al. in [8, Fig. 2] from tomograph measurements. From the curve at $y = 11.82$ cm, we see that one can expect a y dependence in the measurements of system response, as discussed in Appendix A.

ACKNOWLEDGMENT

REFERENCES

[1] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imaging,* vol. MI-1, pp. 113–121, June 1982.
[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.,* vol. B39, pp. 1–37, 1977.

[3] D. L. Snyder and M. I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. Nucl. Sci.,* vol. NS-32, no. 5, 1985.
[4] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 228–238, 1987.
[5] F. B. Hildebrand, *Introduction to Numerical Analysis,* 2nd ed. New York: McGraw-Hill, 1974.
[6] E. J. Hoffman, S. C. Huang, M. E. Phelps, and D. E. Kuhl, "Quantitation in positron emission computed tomography: IV—Effect of accidental coincidences," *J. Comput.-Assisted Tomography,* vol. 5, pp. 391–400, 1981.
[7] E. Veklerov, J. Llacer, and E. J. Hoffman, "MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule," *IEEE Trans. Nucl. Sci.,* vol. 35, no. 1, pp. 603–607, 1988.
[8] E. J. Hoffman, T. M. Guerrero, G. Germano, W. M. Digby, and M. Dahlbom, "PET system calibrations and corrections for quantitative and spatially accurate images," *IEEE Trans. Nucl. Sci.,* vol. 36, no. 1, pp. 1108–1112, 1989.
[9] D. G. Politte and D. S. Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography," *IEEE Trans. Med. Imaging,* vol. 10, pp. 82–89, Mar. 1991.
[10] L. A. Shepp and R. J. Vanderbei, "New insights into emission tomography via linear programming," in *Medical Images: Formation, Handling and Evaluation,* A. E. Tod-Pokropek and MA. Viergever, eds. Heidelberg: NATO ASI Series F, Springer, 1991, pp. 65–86.
[11] T. J. Hebert and R. Leahy, "Fast methods for including attenuation in the EM algorithm," *IEEE Trans. Nucl. Sci.,* vol. 37, no. 2, pp. 754–758, 1990.
[12] B. W. Silverman, M. C. Jones, J. D. Wilson, and C. W. Nychka, "A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography," *J. Roy. Statist. Soc.,* part 1, series B, pp. 277–324, 1990.
[13] R. M. Lewitt and G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimator," *IEEE Trans. Med. Imaging,* vol. MI-5, pp. 16–22, Mar. 1986.
[14] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 37–51, Mar. 1987.
[15] E. Tanaka, N. Nohara, T. Tomitani, and M. Yamamoto, "Utilization of non-negativity constraints in reconstruction of emission tomograms," in *Information Processing in Medical Imaging,* S. L. Bacharach, ed. Dordrecht: Nijhoff, 1986, pp. 379–393.
[16] N. Rajeevan, K. Rajgopal, and G. Krishna, "Vector extrapolated fast maximum likelihood estimation algorithms for emission tomography," *IEEE Trans. Med. Imaging,* vol. 11, pp. 9–20, Mar. 1992.
[17] J. Llacer, E. Veklerov, and J. Nunez, "Statistically based image reconstructions for emission tomography," *Int. J. Imaging Syst. Technol.,* vol. 1, pp. 132–148, 1989.
[18] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules in iterative image reconstruction," *IEEE Trans. Med. Imaging,* vol. 8, pp. 186–193, June 1989.
[19] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 313–319, Dec. 1987.
[20] J. Llacer, "On the validity of hypothesis testing for feasibility of image reconstructions," *IEEE Trans. Med. Imaging,* vol. 9, pp. 226–230, June 1990.
[21] J. Llacer, E. Veklerov, and J. Nunez, "The concept of causality (feasibility) in image reconstruction," in *Medical Images: Formation, Handling and Evaluation,* A. E. Tod-Pokropek and MA. Viergever, eds. Heidelberg: NATO ASI Series F, Springer, 1991, pp. 363–378.
[22] M. Defrise, "Possible criteria for choosing the number of iterations in some iterative reconstruction methods," in *Mathematics and Computer Science in Medical Imaging,* M. Viergever and A. Todd-Pokropek, eds. Heidelberg: NATO ASI Series F, Springer, 1988, pp. 293–303.
[23] H. Fleming, "Equivalence of regularization and truncated iteration in the solution of ill-posed image reconstruction problems," *Linear Algebra Appl.,* vol. 130, pp. 133–150, 1990.
[24] G. T. Herman and D. Odhner, "Performance evaluation of an iterative image reconstruction algorithm for positron emission tomography," *IEEE Trans. Med. Imaging,* vol. 10, pp. 336–346, Sept. 1991.
[25] J. Llacer, E. Veklerov, and J. Nunez, "Stopping rules, Bayesian reconstructions and sieves," in *Information Processing in Medical Imaging,* D. A. Ortendahl and J. Llacer, eds. New York: Wiley-Liss, 1991, pp. 81–93.

[26] E. Veklerov and J. Llacer, "The feasibility of images reconstructed with the method of sieves," *IEEE Trans. Nucl. Sci.,* vol. 37, no. 2, pp. 835–841, 1990.

[27] J. J. Koendrink, "The structure of images," *Biol. Cybern.,* vol. 50, pp. 363–370, 1984.

[28] K. J. Coakley, "A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra," *IEEE Trans. Nucl. Sci.,* vol. 38, no. 1, pp. 9–15, 1991.

[29] K. J. Coakley and J. Llacer, "The use of cross-validation as a stopping rule in emission tomography image reconstruction," *SPIE Proc., Image Phys.: Med. Imaging V,* vol. 1443, pp. 226–233, 1991.

[30] G. Rosenqvist, M. Dahlbom, L. Eriksson, C. Bohm, and G. Blomqvist, "Quantitation of PET data with the EM reconstruction technique," *IEEE Trans. Nucl. Sci.,* vol. 36, no. 1, pp. 1113–1116, 1989.

[31] E. Tanaka, "A fast reconstruction algorithm for stationary positron emission tomography based on a modified EM algorithm," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 98–104, June 1987.

[32] S. Holte, P. Schmidlin, A. Linden, G. Rosenqvist, and L. Eriksson, "Iterative image reconstruction for positron emission tomography: A study of convergence and quantitation problems," *IEEE Trans. Nucl. Sci.,* vol. 37, no. 2, pp. 629–635, 1990.

[33] J. Llacer and A. C. Bajamonde, "Characteristics of feasible images obtained from real PET data by MLE, Bayesian and sieve methods," *SPIE Proc., Digital Image Synthesis and Inverse Opt.,* vol. 1351, pp. 300–312, 1990.

[34] J. S. Liow and S. C. Strother, "Practical tradeoffs between noise, quantization and number of iterations for maximum likelihood-based reconstructions," *IEEE Trans. Med. Imaging,* vol. 10, pp. 563–571, June 1991.

[35] R. H. Huesman, G. T. Gullberg, W. L. Greenberg, and T. F. Budinger, *Donner Algorithms for Reconstruction Tomography,* LBL Publ. 214.

[36] B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink, and M. A. Viergever, "Scale space: Its natural operators and differential invariants," in *Information Processing in Medical Imaging, Lecture Notes in Computer Science, Vol. 511.* Berlin: Springer, 1991, pp. 239–255.

[37] B. M. ter Haar Romeny and L. M. J. Florack, "A multi-scale geometric model of human vision," in *Perception of Visual Information,* Hendee and Wells, eds. Berlin: Springer, 1991.

[38] J. Llacer, B. M. ter Haar Romeny, L. M. J. Florack, and M. A. Viergever, "The representation of medical images by visual response functions," Invited Paper to be published in *IEEE Eng. Med. Biol. Mag.,* Rep. LBL-32546, Lawrence Berkeley Lab., Berkeley, CA.

[39] P. G. Hoel, *Introduction to Mathematical Statistics,* 2nd ed. New York: Wiley, 1960, Ch. 12.

[40] H. H. Barrett and W. Swindell, *Radiological Imaging.* New York: Academic, 1981.

[41] J. Llacer, "Tomographic image reconstruction by eigenvector decomposition: Its limitations and areas of applicability," *IEEE Trans. Med. Imaging,* vol. MI-1, pp. 34–42, Mar. 1982.