Statistics and Computing (2000) **10**, 147–165

Statistical approaches in quantitative positron emission tomography

RICHARD M. LEAHY and JINYI QI

Signal and Image Processing Institute, Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564, USA
leahy@sipi.usc.edu

Received March 1998 and accepted March 1999

Positron emission tomography is a medical imaging modality for producing 3D images of the spatial distribution of biochemical tracers within the human body. The images are reconstructed from data formed through detection of radiation resulting from the emission of positrons from radioisotopes tagged onto the tracer of interest. These measurements are approximate line integrals from which the image can be reconstructed using analytical inversion formulae. However these direct methods do not allow accurate modeling either of the detector system or of the inherent statistical fluctuations in the data. Here we review recent progress in developing statistical approaches to image estimation that can overcome these limitations. We describe the various components of the physical model and review different formulations of the inverse problem. The wide range of numerical procedures for solving these problems are then reviewed. Finally, we describe recent work aimed at quantifying the quality of the resulting images, both in terms of classical measures of estimator bias and variance, and also using measures that are of more direct clinical relevance.

Keywords: positron emission tomography, computed tomography, image reconstruction, maximum likelihood estimation, Bayesian imaging

1. Introduction

The physical basis for positron emission tomography (PET) lies in the fact that a positron produced by a radioactive nucleus travels a very short distance and then annihilates with an electron to form a pair of high energy (511 keV) photons. The pair of photons travel in opposite directions along a straight line path. Detection of the positions at which the photon pair intersect a ring of detectors, Fig. 1, allows us to approximately determine the locus of a line containing the positron emitter. The total number of photon pairs measured by a pair of detectors will be proportional to the total number of positron emissions along the line joining the detectors. If positron emitters are spatially distributed with density $f(\boldsymbol{x})$ at location $\boldsymbol{x}$, the number of detected events between a pair of detectors is an approximate line integral of $f(\boldsymbol{x})$.

By tagging different molecules with positron emitters, PET can be used to reconstruct images of the spatial distribution of a wide range of biochemical probes. Typical applications of PET include glucose metabolism studies for cancer detection and cardiac imaging, imaging of blood flow and volume, and studies of neurochemistry using a range of positron labeled neuro-receptors and transmitters. Tracer kinetics can also be studied using PET by acquiring dynamic data sets. An excellent review of the current state of the art of PET instrumentation and applications can be found in Cherry and Phelps (1996). Here we address only those aspects of system design and calibration that directly impact on the design and performance of statistically based image estimators. However, we agree with the view expressed by Ollinger and Fessler (1997), that much of the literature dealing with statistically based image reconstruction uses an over-simplified model. Here we will attempt to provide a balance between this and an overly detailed description of the technical aspects of PET system design.

A PET scanner requires one or more rings of photon detectors coupled to a timing circuit that detects coincident photon pairs by checking that both photons arrive at the detectors within a few nanoseconds of each other. A unique aspect of PET, as compared to most other tomographic systems, is that the complete ring of detectors surrounding the subject allows simultaneous acquisition of a complete data set so that no rotation of the detector system is required. A schematic view of two modern PET

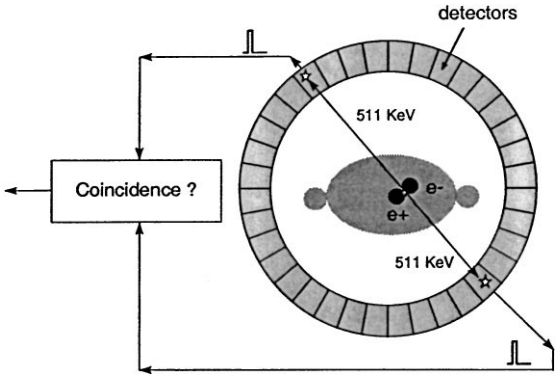

Fig. 1. *The physical basis for PET: annihilation of a positron and an electron produces a pair of 511 keV photons that are detected by a pair of scintillation detectors*

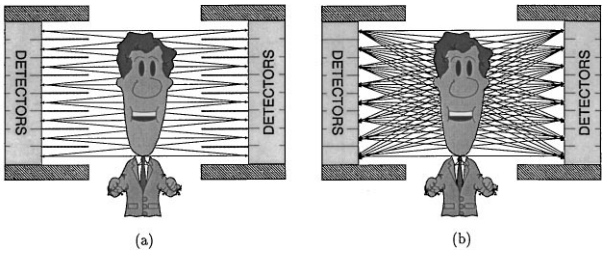

Fig. 2. *Schematic of an axial cross section through (a) a 2D and (b) a 3D PET scanner. The septa in the 2D scanner stop out-of-plane photons while the 3D scanner detects these events as additional data*

scanners is shown in Fig. 2. Multiple rings of detectors surround the patient with rings of dense material, or "septa", separating each ring. These septa stop photons traveling between rings so that coincidence events are collected only between pairs of detectors in a single ring. We will refer to this configuration as a 2D scanner since the data are separable and the image can be reconstructed as a series of 2D sections.[1] In contrast, the 3D scanners have no septa so that coincidence photons can be detected between planes. This results in a factor of 4 to 7 increase in the total number of photons detected and hence increases the signal to noise ratio. In this case the reconstruction problem is not separable and must be treated directly in 3D.

In most scanners, the detectors consist of a combination of scintillators and photomultiplier tubes (PMTs). Scintillators used in PET include bismuth germinate (BGO), sodium iodide (NaI) and lutetium oxyorthosilicate (LSO). These convert the high energy 511 keV photons into a large number of low energy photons which are then collected and amplified by the PMTs. Typically 64 scintillation detectors will be coupled to four PMTs as shown in Fig. 3. The output signal from all four PMTs is used to determine in which of the 64 crystals the 511 keV photon was absorbed. Arranging the detector blocks in circular fashion produces a ring scanner; additional rings of blocks are added to increase the axial field of view of the scanner.

With this basic picture of a PET system, we can now turn to the issues of data modeling and image reconstruction.

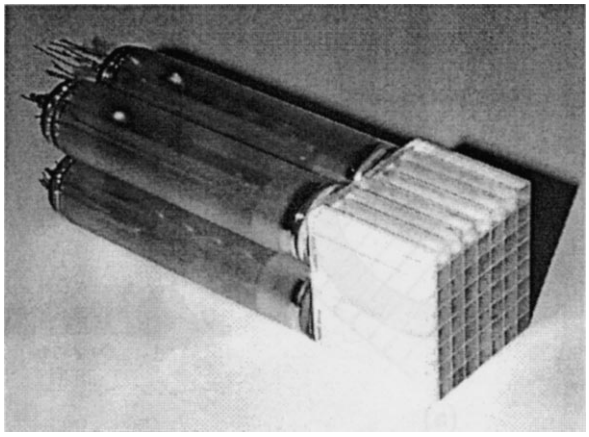

Fig. 3. *Photograph of a block-detector: an 8 by 8 array of BGO crystals are coupled to four larger photomultiplier tubes (PMTs). The light output from each crystal is shared between the PMTS. The resulting output signals from the PMTs are used to decide from which crystal this light originated*

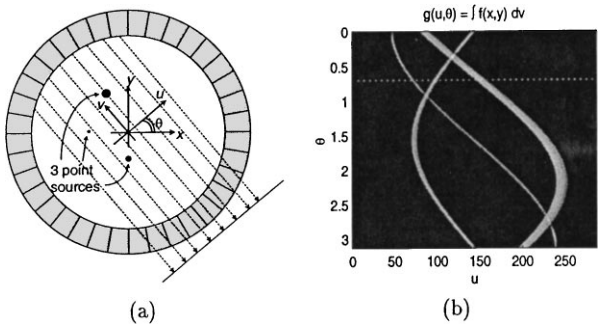

Fig. 4. *Coincidence detection between parallel pairs of detectors in (a) corresponds to one line of the Radon transform of the source distribution in (b). A point source in (a) maps to a sinusoid in Radon transform space (b)*

Consider first the 2D arrangement. In the absence of an attenuating medium and assuming perfect detectors, the total number of coincidences between any detector pair is, approximately, a line integral through the 2D source distribution. The set of line integrals of a 2D function form its Radon transform, thus the coincidence data are readily sorted into a Radon transform or sinogram format as illustrated in Fig. 4. Once the sinogram data have been collected, the source distribution can be reconstructed using the standard filtered backprojection (FBP) algorithm which is a numerical method for inverting the Radon transform of a 2D function (Shepp and Logan 1974).

In the 3D case, the presence of oblique line-integral paths between planes make the analytic method less straightforward. Not only is there a huge increase in the amount of data, but also the limited axial extent of the scanner results in missing data in the oblique sinograms. One solution to this problem is to use an analytic 3D reconstruction method in combination with a reprojection procedure to fill in the missing data (Kinahan and Rogers 1989). An alternative approach to 3D reconstruction is to "rebin" the data into equivalent 2D sinograms and

apply 2D reconstruction algorithms to the result. The cruder forms of rebinning lead to substantial resolution loss. However, recent Fourier rebinning methods achieve impressive speed up in computation with little loss in performance (Defrise *et al.* 1997).

While the analytic approaches result in fast reconstruction algorithms, accuracy of the reconstructed images is limited by the approximations implicit in the line integral model on which the reconstruction formulae are based. In contrast, the statistical methods that we will review here can adopt arbitrarily accurate models for the mapping between the source volume and the sinograms. A second limitation of the analytic approaches is that they do not take account of the statistical variability inherent in photon limited coincidence detection. The resulting noise in the reconstructions is controlled, at the expense of resolution, by varying the cut-off frequency of a linear filter applied to the sinogram. Since the noise is signal dependent, this type of filtering is not particularly effective at achieving an optimal bias-variance trade-off. Again, the statistical approaches allow explicit modeling of statistical noise associated with photon limited detection.

The combination of improved modeling of the detection process and improved handling of statistical noise when using statistically based methods offers the possibility for enhanced performance of PET with both high count data (where model accuracy limits resolution) and low count data (where statistical noise limits resolution). In its simplest form the imaging problem can be cast as one of parameter estimation, where the data are Poisson random variables with mean equal to a linear transformation of the parameters. This formulation is complicated, as we will describe here, by the impact of additional noise and correction terms. To give an idea of the scale of the problem, a single 3D scan from the latest generation of PET systems could produce 10^7–10^8 sinogram elements with 10^6 image elements to be estimated.

We have organized the paper as follows. We first develop a model for the PET data based on the physics of coincidence detection. We then review various formulations of the inverse problem that derive from the Poisson model for the coincidence data. Here we also address the issue of ill-posedness and review the various forms of regularization used to overcome it. We then turn our attention to the wide range of numerical methods for optimizing the chosen cost function. We also describe some of the recent work in evaluating estimator performance using basic properties such as bias and variance, and also using task specific evaluation. We conclude with some examples.

2. Data modeling

2.1. *The coincidence model*

For the purposes of this review, we will assume that the image is represented using a finite set of basis functions. While there has been some interest in alternative basis elements, (e.g. smooth spherically symmetric "blobs" (Matej and Lewitt 1996)), almost all researchers currently use a cubic voxel basis function. Each voxel is an indicator function on a cubic region centered at one of the image sampling points in a regular 2D or 3D lattice. The image value at each voxel is proportional to the total number of positron-emitting nuclei contained in the volume spanned by the voxel. A single index will be used to represent the lexicographically ordered elements of the image, $\boldsymbol{f} = \{f_j,\ j = 1, \ldots, N\}$. Similarly, the elements of the measured sinograms will be represented in lexicographically ordered form as $\boldsymbol{y} = \{y_i,\ i = 1, \ldots, M\}$. To give an idea of the size of the problem, we have listed some of the basic parameters for a 3D whole body PET scanner in Table 1.

Since the data are inherently discrete and the detection process approximately linear, the mapping between the source image and the expected value of the true coincidence data can be represented by a forward projection matrix, $\boldsymbol{P} \in \mathbb{R}^{M \times N}$. The elements, p_{ij}, contain the probabilities of detecting an emission from voxel site j at detector pair i. As we will see below, the measured data are corrupted by additive random coincidences, $\boldsymbol{r}$, and scattered coincidences, $\boldsymbol{s}$, so that the basic model for the mean of the data is:

$$\bar{\boldsymbol{y}} = E[\boldsymbol{y}] = \boldsymbol{P}\boldsymbol{f} + \boldsymbol{r} + \boldsymbol{s} \tag{1}$$

The emission of positrons from a large number of radioactive nuclei is well known to follow a Poisson distribution. Provided that the detection of each photon pair by the system is independent and can be modeled as a Bernoulli process, then the sinogram data are a collection of Poisson random variables. The independence in detection does not strictly hold since all PET scanners are limited in the rate at which they can count – a

Table 1. *Typical problem dimensions for reconstruction of images from the Siemens/CTI EXACT HR+ body scanner operating in 3D mode. Use of sparse structures and symmetry reduce the size of the projection matrix to manageable proportions*

Ring diameter, mm	826	Object size, mm	$577 \times 577 \times 155$
Detectors per ring	576	Object size, voxels	$128 \times 128 \times 63$
Number of rings	32	Voxel size, mm	$4.5 \times 4.5 \times 2.4$
Angles per sinogram	144	Full size of $\boldsymbol{P}$	10^{13}
Rays per angle	288	Storage size of $\boldsymbol{P}_{\text{geom}}$	42 Mbytes
Number of sinograms	239	Storage size of $\boldsymbol{P}_{\text{det.blur}}$	0.5 Mbytes
Projections per sinogram	41,472	Storage size of $\boldsymbol{P}_{\text{attn}}$ and $\boldsymbol{P}_{\text{det.sens}}$	40 Mbytes
Total projection rays	10^7	Total storage size of $\boldsymbol{P}$	82.5 Mbytes

restriction reflected in the so-called "dead-time" calibration factor, which is a measure of the fraction of time that the scanner is unable to record new events because it is processing photons that have already arrived. Here we will assume that the count rates are sufficiently low that the system is operating in the linear range and the data in (1) can be well modeled as Poisson with mean $\bar{\boldsymbol{y}}$. As we will see later, in most PET scanners the data are actually pre-corrected for random coincidences so that they are no longer Poisson. We now consider each of the three terms on the right hand side of (1) and describe how they can be handled within a statistical formulation of the inverse problem.

2.2. *True coincidences*

In an effort to develop an accurate and computationally efficient representation of the projection data we have developed the following factored representation:

$$\boldsymbol{P} = \boldsymbol{P}_{\text{det.sens}}\boldsymbol{P}_{\text{det.blur}}\boldsymbol{P}_{\text{attn}}\boldsymbol{P}_{\text{geom}}\boldsymbol{P}_{\text{positron}} \tag{2}$$

This specific form has been used only in our own work, e.g. Qi *et al.* (1998a). Factorizations that included some of these terms were earlier proposed by Kearfott (1985) and Baker (1991). In other cases, implicit factorizations are often used; for example, when attenuation and detector normalization corrections are applied separately from the forward or backward projection procedures (Kaufman 1987) or when a Gaussian blurring of the computed sinogram is used to match the approximate response of the detectors.

$\boldsymbol{P}_{\text{positron}}$: the emitted positron travels a small distance before annihilating with an electron. The distance is dependent on the specific isotope and the density of the surrounding tissue. In water, the common PET isotopes have full-width-at-half-maximum (FWHM) ranges between 0.1 mm for ^{18}F and 0.5 mm for ^{15}O. The range distributions are long tailed so that although these factors are negligible for ^{18}F studies, they are one of the primary factors limiting resolution in ^{15}O studies. We can include positron range in the PET data model by making $\boldsymbol{P}_{\text{positron}}$ a local image blurring operator that is applied to the true source distribution (Terstegge *et al.* 1996). If we assume that the density inside the patient is that of water, then these factors would be shift invariant. More sophisticated modeling with shift-variant blurs would involve the use of an attenuation map (see Section 2.3) to determine spatially variant positron range distributions. In most work to date these factors have not been included in the model, although there have been attempts to de-convolve these factors from either the sinogram data or the reconstructed image (Haber, Derenzo and Uber 1990).

$\boldsymbol{P}_{\text{geom}}$: is a matrix that contains the geometrical mapping between the source and data. The $(i, j)^{\text{th}}$ element is equal to the probability that a photon pair produced in voxel j reaches the front faces of the detector pair i. While the conventional model for this is based on computing the intersection of a tube joining the detector pair with each voxel, a more correct model is based on the solid angle subtended at each of the detectors by each voxel (Hoffman *et al.* 1982). We can account for finite voxel size by numerically integrating the solid angles over the volume of each voxel.

The dominant cost in almost all iterative PET reconstruction algorithms is that involved in forward and backward projection. Using the factored form above, the most expensive part of this operation is multiplication by $\boldsymbol{P}_{\text{geom}}$ or its transpose. Consequently, it is very important that this matrix be represented effectively to minimize storage and computation costs. Although $\boldsymbol{P}_{\text{geom}}$ is extremely large, it is also very sparse with a high degree of symmetry. The sparseness arises from the small fraction of voxels that can produce coincidences at each detector pair. In 2D there is also a total of an 8-fold rotation and reflection symmetry in the geometric projection matrix for a circular ring of detectors; in 3D there are additional symmetries for coincidences between detector rings (Johnson *et al.* 1995, Chen, Lee and Cho 1991, Qi *et al.* 1998a). Further savings in storage costs and computation time can be realized by storing only non-zero elements and using run-length coding. The reductions that can be achieved are illustrated in Table 1.

$\boldsymbol{P}_{\text{attn}}$: the geometric term above will determine the number of photons reaching the detectors in the absence of an attenuating medium. In fact, the body prevents a substantial fraction of photons reaching the detectors, primarily through Compton scattering of one or both photons (Barrett and Swindell 1981). It is straightforward to show that the probability of attenuation is approximately constant for all photon pairs that would otherwise impinge on a given detector pair. Thus the attenuation factor can be represented by a diagonal matrix containing the survival probabilities. Accurate attenuation factors are crucial for obtaining high quality PET images and we return to this topic in Section 2.3.

$\boldsymbol{P}_{\text{det.blur}}$: Once photons arrive at the detectors, the detection process is complicated by a number of factors which we have lumped into the matrix $\boldsymbol{P}_{\text{det.blur}}$. This matrix acts as a local blurring function applied to the sinogram formed by multiplying the source distribution by the positron range and geometric projection matrices. The blurring occurs for three primary reasons: (i) the photons are not exactly co-linear; (ii) photons may be scattered from one crystal to another resulting in a mis-positioning of the detected photon; (iii) the crystal surface is not always orthogonal to the direction of arrival of the photon so that a photon may penetrate through one or more crystals before being stopped. These factors are illustrated in Fig. 5. In principle the non-colinearity effect should have been combined with the geometric projection matrix, but we have found that it can be included in the blurring factors without noticeable loss in model accuracy. Exact computation of the three factors is not practical. Instead we have used Monte Carlo simulations in which we track large numbers of photon pairs through a simplified model of the detectors (Mumcuoglu *et al.* 1996). As with the geometric projection matrix, there is a great deal of symmetry in these blur factors.

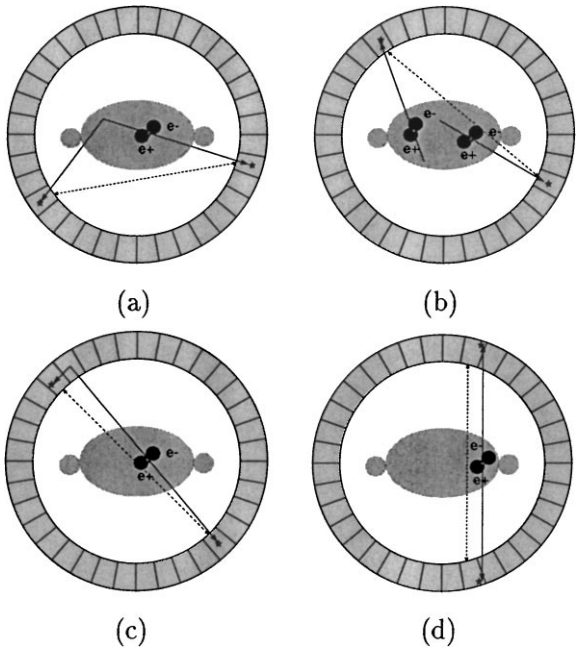

Fig. 5. *Figure shows various factors that complicate the basic PET model: (a) scatter within the body (b) random coincidences (c) inter-crystal scatter (d) crystal penetration*

Furthermore the blurring only extends over a small area of the sinogram so that storage and computation costs associated with these factors are small. By factoring these effects out of the geometric projection matrix we achieve a reduction by a factor of approximately three in the geometric projection matrix size and comparable savings in reconstruction time.

$\boldsymbol{P}_{\text{det.sens}}$: Once a photon pair reaches a detector it may not be detected since no detector is 100% efficient. $\boldsymbol{P}_{\text{det.sens}}$ is a diagonal matrix that contains the detection efficiency of each detector pair. The terms that contribute to these factors include the intrinsic sensitivities of the individual crystals, the relative position of the crystals within a detector block, and geometric factors related to the distance of the detector pair from the center of the field of view (Casey, Gadagkar and Newport 1995). These factors are all measured through calibration procedures and are typically provided to the user in a form that can be used to directly generate the diagonal $\boldsymbol{P}_{\text{det.sens}}$ matrix. An additional complicating factor is that of system dead-time. This is an approximate measure of the fraction of counts lost due to the detectors being unable to detect new photons while the system is occupied with events that were previously detected. The dead-time correction factor is usually estimated from the singles rate. It should be noted that the dead-time effect causes non-linear behavior at high count rates and methods to account for this are still needed.

2.3. *Attenuation effects*

A substantial fraction of the photons emitted through positron-electron annihilation do not directly exit the body. Rather they undergo Compton scattering (Barrett and Swindell 1981) in which the energy of the photon is reduced and its direction is altered. If this photon is later detected, then a scattered event is recorded as discussed below. Whether or not the scattered photon is detected, there is a net loss of counts along the original path on which the photon pair was traveling. It is straightforward to show that the probability of attenuation along any straight line path is independent of the location along the path that the original annihilation occurs. The survival probability for a photon pair is equal to $\exp\{-\int \mu(\boldsymbol{x})\, dl\}$, where $\mu(\boldsymbol{x})$ is the linear attenuation coefficient at $\boldsymbol{x}$ and the integral is taken along the straight line path.

In most instances the attenuation factor is found using an external transmission source. These sources are usually either a ring of positron emitting material that surrounds the patient, or a rotating rod of activity. Photons traveling along straight line paths through the patient are attenuated according to the same probability as emitted photons within the body traveling on the same paths. A simple estimate of the probability of survival can be computed as the ratio of the number of photon pairs detected with the patient present (the transmission scan) to the number detected in the absence of the patient or attenuating medium (the blank scan).

This simple division method of computing survival probabilities produces high variance and biased estimates. Errors are particularly large when the number of detected transmission counts is low. Alternatively, the transmission data can be used to reconstruct an image of the linear attenuation coefficients via statistical methods very similar to those reviewed below. As with emission data, the transmission scans are photon limited and contain scattered and random coincidences. Formulation of the transmission reconstruction problem follows in a similar manner to the emission methods described here, with the primary difference being that the mean of the data contain the line integrals of the attenuation image in exponential form. Rather than pursue this issue further here, we refer the interested reader to Lange, Bahn and Little (1987), Fessler *et al.* (1997), Mumcuoglu *et al.* (1994).

2.4. *Scattered and random coincidences*

The true coincidence data are corrupted by two forms of additive noise. These are the scatter and randoms components, with means $\boldsymbol{s}$ and $\boldsymbol{r}$ respectively, in (1). Scattered events refer to coincidence detection after one or both of the photons has undergone Compton scattering. Clearly, after scattering, we can no longer assume that positron emission occurred along the path joining the two detectors.

Scattered photons have lower energy than their unscattered 511 keV counterparts. Consequently some of the scattered events can be removed by rejecting events for which the energy detected by the PMTs does not exceed some reasonable threshold. Unfortunately, setting of this threshold sufficiently high to reject most of the scattered radiation will also result in rejection of a large fraction of unscattered photons due to the poor

energy resolution of BGO detectors. For the most commonly used BGO detectors with the standard energy thresholds, 2D PET studies typically have a scattered to true coincidence ratio of about 10%, while in 3D studies the fraction often exceeds 30%. Typically scatter contributions to the data tend to be fairly uniform. They are often simply ignored in qualitative studies since they result in an approximately constant offset in the image when using linear estimators. When non-linear methods are used, or when accurate quantitation is required, these factors must be modeled.

Given the distribution of the source image and an image of the linear attenuation coefficient, an accurate scatter profile can be computed using the Klein-Nishina formula for Compton scatter (Barrett and Swindell 1981). Since the scatter profiles are smooth, it is possible to compute them with reasonable computational load from a low resolution, preliminary reconstruction of the emission source. Once this is estimated, the scatter contribution can be viewed as a known offset in the mean of the data in (1) rather than as an explicit function of the data that must be re-computed with each new estimate of the image. Model based scatter methods are described in Ollinger, Johns and Burney (1992), Watson, Newport and Casey (1995), Mumcuoglu, Leahy and Cherry (1996). In the following we will assume that the scatter component in the data has been estimated using one of these methods.

Random coincidences (henceforth called "randoms"), as mentioned above, are caused by the detection of two independent photons within the coincidence timing window. The randoms contribution to the data is a function of the length of this timing window and of the source activity. By simply delaying the timing window by a fixed amount, one can obtain data which consist of purely randoms and with the same mean number of counts as for the non-delayed window. Thus on most scanners, a randoms corrected data set is collected in which two timing windows are used, one to collect true and random coincidences, the second to collect randoms only. The difference of these two is the corrected data. While this does correct the data, in mean, for the randoms, the resulting data has increased variance due to the subtraction of two Poisson processes. This has important implications for the data model as discussed below.

3. Formulating the inverse problem

3.1. *Likelihood functions*

The great majority of publications employing statistical PET models assume the data $\boldsymbol{y}$ is Poisson with mean $\bar{\boldsymbol{y}}$ and distribution

$$p(\boldsymbol{y} \mid \boldsymbol{f}) = \prod_{i=1}^{M} \frac{\bar{y}_i^{y_i} e^{-\bar{y}_i}}{y_i!} \tag{3}$$

The corresponding log-likelihood, after dropping constants, is

$$L(\boldsymbol{y} \mid \boldsymbol{f}) = \sum_{i=1}^{M} y_i \log \bar{y}_i - \bar{y}_i \tag{4}$$

The mean $\bar{\boldsymbol{y}}$ is related to the image through the affine transform (1). In most cases, the effects of scatter and randoms, whether present or subtracted from the data, are simply ignored and the data are assumed to follow this Poisson model with mean $\boldsymbol{y} = \boldsymbol{Pf}$.

In an effort to reduce computation costs and numerical problems associated with the logarithm that occurs in the log-likelihood function (Bouman and Sauer 1996) suggests using a quadratic approximation to the data:

$$L(\boldsymbol{y} \mid \boldsymbol{f}) = -\frac{1}{2} \sum_{i=1}^{M} \frac{(\bar{y}_i - y_i)^2}{\hat{\sigma}_i^2} \tag{5}$$

where $\hat{\sigma}$ is an estimate of the variance of each measurement. For Poisson data the variance should equal the mean $\bar{y}_i$ and is often approximated using the observed data y_i. This approximation can be improved using a forward projection of an early estimate of the image when using iterative reconstruction algorithms.

The Poisson model above is appropriate only when the data have not been corrected for randoms, and when the randoms and scatter components are explicitly included in the model. When operated in standard mode, PET scanners pre-correct for randoms by computing the difference between coincidences collected using a standard timing window and those in a delayed timing window. These data are Poisson processes with means $\bar{\boldsymbol{y}} = E[\boldsymbol{y}] = \boldsymbol{Pf} + \boldsymbol{r} + \boldsymbol{s}$ and $\boldsymbol{r}$, respectively. The precorrected data $\boldsymbol{y}$ has mean $\boldsymbol{Pf} + \boldsymbol{s}$ and variance $\boldsymbol{Pf} + \boldsymbol{2r} + \boldsymbol{s}$ so that a Poisson model does not reflect the true variance. The true distribution has a numerically intractable form and an approximation should be used (Yavuz and Fessler 1998). One possibility is to modify the quadratic approximation in (5) using an increased variance as proposed in Fessler (1994). A better approximation is the shifted-Poisson model in which the first two moments of the corrected data are matched by assuming that $\boldsymbol{y} + \boldsymbol{2r}$ is Poisson with mean $\boldsymbol{Pf} + \boldsymbol{2r} + \boldsymbol{s}$. This results in the modified log likelihood (Yavuz and Fessler 1998):

$$L(\boldsymbol{y} \mid \boldsymbol{f}) = \sum_{i=1}^{M} (y_i + 2r_i) \log ((\boldsymbol{Pf})_i + 2r_i + s_i) - ((\boldsymbol{Pf})_i + 2r_i + s_i) \tag{6}$$

In closing this section we note that likelihood functions that model the increase in variance due to randoms subtraction require estimates of the mean of this randoms process. These must, in turn, be estimated from the measurements and calibration data (Fessler 1994, Qi *et al.* 1998b).

3.2. *Priors*

Direct maximum likelihood (ML) estimates of PET images exhibit high variance due to ill-conditioning. Some form of regularization is required to produce acceptable images. Often this is accomplished simply by starting with a smooth initial estimate and terminating an ML search before convergence. Here we consider explicit regularization procedures in which a prior distribution is introduced through a Bayesian reformulation of the

problem to resolve the ill-conditioning in the likelihood function. Some authors prefer to present these regularization procedures as penalized ML methods but the differences are largely semantic, except in the case where the penalty functions are explicit functions of the data, e.g. Fessler and Rogers (1996).

Bayesian methods can address the ill-posedness inherent in PET image estimation through the introduction of random field models for the unknown image. In an attempt to capture the locally structured properties of images, researchers in emission tomography, and many other applications of image processing, have adopted Gibbs distributions as suitable priors. The Markovian properties of these distributions make them theoretically attractive as a formalism for describing empirical local image properties, as well as computationally appealing since the local nature of their associated energy functions result in computationally efficient update strategies.

Let $S = \{1, 2, \ldots, N\}$ denote the ordered set of image voxel indices. The Gibbs distribution is defined on a neighborhood system which associates a set of sites $W_j \subset S$ with each site j. The neighborhood system must satisfy the property that $i \in W_j$ iff $j \in W_i$. The sites in W_j are typically the collection of voxels closest, up to some maximum Euclidean distance, to site j. The Gibbs distribution has the general form

$$p(\boldsymbol{f} \mid \beta) = \frac{1}{Z} e^{-\beta U(\boldsymbol{f})} \tag{7}$$

where $U(\boldsymbol{f})$ is the Gibbs energy function defined as a sum of potentials, each of which is a function of a subset or clique $c_k \subset S$. The cliques for a particular neighborhood system must satisfy the condition that each pair of sites in each clique c_k are mutual neighbors.

The form of Gibbs distributions most commonly used in image processing are those for which the energy function $U(\boldsymbol{f})$ contains potentials defined only on pair-wise cliques of neighboring voxels:

$$U(\boldsymbol{f}) = \sum_{j=1}^{N} \sum_{k \in W_j, k > j} \phi_{jk}(f_j - f_k) \tag{8}$$

For a 3D problem, the neighbors of an internal voxel would be the nearest 6 voxels for a 1st order model, or the nearest 26 voxels for a 2nd order model (with appropriate modifications for the boundaries of the lattice).

The potential functions $\phi_{jk}(f_j - f_k)$ are chosen to attempt to reflect two conflicting image properties: (i) images are locally smooth, (ii) except where they are not! For example, in PET images we might expect to see smooth variations in tracer uptake within a specific organ or type of tissue, and abrupt changes as we move between different organs or tissue types. A wide range of functions have been studied in the literature that attempt to produce local smoothing while not removing or blurring true boundaries or edges in the image. All have the basic property that they are monotonic non-decreasing functions of the absolute intensity difference $|(f_j - f_k)|$. Taking the square of this function leads to a Gauss-Markov prior which produces smooth images with very low probability of sharp transitions in intensity. In an attempt to increase the probability of these sharp transitions, Bouman and Sauer (1996) propose using the generalized p-Gaussian model where $\phi(f) = |f|^p$, $1 < p < 2$. An alternative function with similar behavior, that derives from the literature on robust estimation, is the Huber prior in which the continuous function $\phi(f)$ switches from quadratic to linear at a user specified transition point (Huber 1981, Qi *et al.* 1998a). The function $\phi(f) = \log\cosh(cf)$, where c is a user specified parameter, has similar behavior except that the function has the advantage of continuous derivatives, but with higher computation costs (Green 1990). All of these examples produce convex energy functions.

In an attempt to produce even sharper intensity transitions, several highly non convex functions have also been proposed. For example, Geman and McClure (1985), who were also the first to specifically use Gibbs distributions in emission tomography, proposed the function $\phi(f) = \frac{f^2}{f^2+\delta^2}$. This and other non-convex potentials have the property that the derivative of the function decreases once the intensity difference exceeds some threshold. The limiting case of this approach is found in the weak membrane model which is quadratically increasing up to some threshold and then remains constant beyond this (Gindi *et al.* 1993).

Higher order neighborhoods are able to capture more complex correlation structure than the simple pair-wise models (Chan, Herman and Levitan 1995). Unfortunately, the problem of choosing and justifying such a model becomes increasingly difficult with the size of the neighborhood. One example where a higher order neighborhood has been used to nice effect is the thin plate spline model of Lee, Rangarajan and Gindi (1995). This model uses a discrete approximation to the bending energy of a thin-plate as the Gibbs energy function. Since the thin plate energy involves second order derivatives, higher order cliques must be used in the model.

Rather than implicitly modeling image boundaries as in the examples above, the compound MRFs include a second coupled random field, defined on a dual lattice, that explicitly represents the boundaries in the image. The dual lattice points are placed between each pair of sites in the image lattice and are set to unity if there is an image boundary between that pair of voxels, otherwise they are set to zero. In this case, the number of unknown parameters to be estimated is doubled. The boundary is usually estimated from the current estimate of the image at each iteration. In this way, the prior can explicitly model edges in the image, and introduce additional potential terms to encourage the formation of connected boundaries (Geman and Geman 1984). A wide range of compound MRFs have been studied in the PET literature, e.g. Lee, Rangarajan and Gindi (1995), Johnson *et al.* (1991), Leahy and Yan (1991).

One of the primary attractions of the compound MRFs is that they serve as a natural framework for incorporating anatomical information into the reconstruction process (Leahy and Yan 1991, Gindi *et al.* 1991, 1993). Since different anatomical

structures have different physiological functions, we can expect to see differences in tracer uptake between structures. This general observation is borne out in high resolution autoradiographic images in which functional images also clearly reveal the morphology of the underlying structures (Gindi *et al.* 1993). Because the anatomical modalities, such MR and CT, have superior resolution to PET, fairly accurate estimates of anatomical boundaries can be formed. These can then be used to influence the formation of the boundary process in the PET image estimation procedure.

3.3. *The posterior density*

The likelihood function and image prior are combined through Bayes rule to produce the posterior density

$$p(\boldsymbol{f} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{y})}. \tag{9}$$

Bayesian formulations of the PET problem are usually reduced to computing a maximum *a posteriori* (MAP) estimate of the image as the maximizer of the posterior. Taking the log of the posterior and dropping constants, we have the basic form of the MAP objective function

$$\Phi(\boldsymbol{f}, \boldsymbol{y}) = L(\boldsymbol{y} \mid \boldsymbol{f}) - \beta U(\boldsymbol{f}). \tag{10}$$

The log likelihood functions in Section 3.1 are concave so that $\Phi(\boldsymbol{f},\boldsymbol{y})$ will also be concave if the Gibbs energy is convex. In this case, a global maximum can be found using standard nonlinear optimization methods.

If MRFs with non-convex potential functions are used then local maxima will exist for the MAP objective function. Global search techniques such as simulated annealing are typically impractical for PET reconstruction because of the number of voxels in a typical image. Furthermore, the non-local property of the forward and backward projection results in a fully coupled posterior, i.e. the posterior density does not share the local neighborhood system of the prior. In practice, local searches for a stationary point of the objective function are typically used to find a local maximum of $\Phi(\boldsymbol{f}, \boldsymbol{y})$. In addition to the difficulty in computing MAP estimates for non-convex priors, it should be noted that the solutions are also inherently discontinuous with respect to the data. This can result in high variance estimates. When using compound MRFs, the reconstruction problem is further complicated by the need to estimate the additional set of discrete variables that represent image boundaries.

There are a much broader class of Bayesian estimators that could be developed to minimize an expected loss computed over the posterior density. These could be developed to optimize the reconstruction algorithm for specific well-defined tasks. One example of a Bayesian formulation that does not involve the computation of a MAP estimate was described by Bowsher *et al.* (1996). A hierarchical MRF model for emission tomography is described in which the image consists of regions of activity, each of which has an intensity that is assumed to vary smoothly about its mean. The procedure described involves estimating the support of each of these regions and their mean intensities. Several other researchers, e.g. Green (1996), Weir (1997), Higdon *et al.* (1997), have proposed alternative Markov Chain Monte Carlo methods for image estimation. These methods are attractive since they are better suited to more complex priors and the use of hyperprior densities on unknown parameters, however the high computational burden is a major obstacle preventing their widespread use.

4. Image estimation

4.1. *ML estimators*

Since there is no closed form solution of the ML problem for the Poisson likelihood, solutions are computed iteratively. Iterative estimation schemes in PET have their basis in the row-action method or ART (algebraic reconstruction techniques) developed during the 1970s (Censor 1983). ART solves a set of linear equations by successive projection of the image estimate onto the hyperplanes defined by each equation. The approach is attractive for the sparse matrix structures encountered in PET but has no statistically optimal properties. In cases where the data are consistent, ART will converge to a solution of the equations $\boldsymbol{y} = \boldsymbol{Pf}$. In the inconsistent case, the iterations will not converge to a single solution and the properties of the image at a particular stopping point will be dependent on the sequence in which the data are ordered. Many variations on the ART theme can be found in the literature (Censor 1983) but we will restrict attention here to estimators that are based on explicit statistical data models.

Rockmore and Macovski (1976) published an early paper on ML methods for emission tomography, but it was the work of Shepp and Vardi (1982) and Lange and Carson (1984), who applied the EM algorithm of Dempster, Laird and Rubin (1977) to the problem, that lead to the current interest in ML approaches for PET. It is interesting to note that one of the major attractions of this method to the nuclear medicine community was that the EM method produces an elegant closed-form update equation reminiscent of the earlier ART methods.

The EM algorithm is based on the introduction of a set of complete but unobservable data, $\boldsymbol{w}$, that relates the incomplete observed data $\boldsymbol{y}$ to the image $\boldsymbol{f}$. The algorithm alternates between computing the conditional mean of the complete data log likelihood function, $\ln p\ (\boldsymbol{w} \mid \boldsymbol{f})$, from $\boldsymbol{y}$ and the current image estimate $\boldsymbol{f}^{(k)}$, and then maximizing this quantity with respect to the image:

$$\begin{aligned} &\text{E-step:} \quad Q\big(\boldsymbol{f} \mid \boldsymbol{f}^{(k)}\big) = E\big[\ln L(\boldsymbol{w} \mid \boldsymbol{f}) \mid \boldsymbol{y}; \boldsymbol{f}^{(k)}\big] \\ &\text{M-step:} \quad\quad \boldsymbol{f}^{k+1} = \arg\max_{\boldsymbol{f}} Q\big(\boldsymbol{f} \mid \boldsymbol{f}^{(k)}\big) \end{aligned} \tag{11}$$

For the PET reconstruction problem the complete data is chosen as $\boldsymbol{w} = \{\{w_{ij}\}_{j=1}^{N}\}_{i=1}^{M}$ with each w_{ij} denoting the emissions from voxel j being detected by detector pair i (Lange and Carson 1984, Vardi, Shepp and Kaufman 1985). In this model, randoms and scatter are ignored but modifications to deal with these as

additive factors are straightforward. The final EM algorithm has the form:

E-step: $Q(\boldsymbol{f} \mid \boldsymbol{f}^{(k)})$

$$= \sum_j \left(f_j^{(k)} \sum_i \frac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k)}} \log(p_{ij} f_j) - f_j \sum_i p_{ij} \right) \tag{12}$$

M-step:

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{\sum_i p_{ij}} \sum_i \frac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k)}}$$

Two problems were widely noted with this algorithm: it is slow to converge and the images have high variance. The variance problem is inherent in the ill-conditioning of the Fisher information matrix. In practice it is controlled in EM implementations using either stopping rules (Veklerov and Llacer 1987, Coakley 1991, Johnson 1994a) or post-smoothing of the reconstruction (Llacer *et al.* 1993, Silverman *et al.* 1990). An alternative approach to avoiding instability is to use Grenander's method of sieves (Grenander 1981). The basic idea is to maximize the likelihood over a constrained subspace and then relax the constraint by allowing the subspace to grow with the sample size. This usually produces consistent estimates provided that the sieve grows sufficiently slowly with sample size. Snyder and Miller (1985) have successfully applied this approach to PET using a Gaussian convolution-kernel sieve.

Many researchers, (Lewitt and Muehllehner 1986, Kaufman 1987, Rajeevan, Rajgopal and Krishna 1992), have studied methods for speeding up the EM algorithm by re-writing the EM update equation (12) as:

$$f_j^{(k+1)} = f_j^{(k)} + f_j^{(k)} \frac{1}{\sum_i p_{ij}} \frac{\partial L(\boldsymbol{y} \mid \boldsymbol{f}^{(k)})}{\partial f_j}. \tag{13}$$

Re-written in this way, EM looks like a special case of gradient ascent and some degree of speed-up can be realized using over-relaxation or line-search methods. More substantial gains are achieved by returning to standard gradient ascent methods, and in particular pre-conditioned conjugate gradient searches (Kaufman 1993).

One distinct attraction of the original EM algorithm is that the updates impose a natural non-negativity constraint. This is not shared by the gradient-based methods and imposition of a non-negativity condition on these methods requires careful handling (Kaufman 1993). An alternative to gradient based searches is to use iterated coordinate ascent (ICA) methods in which the voxels are updated sequentially, thus making imposition of the non-negativity constraint trivial. ICA also leads to dramatic speed up in convergence rate in comparison to the EM algorithm (Sauer and Bouman 1993). We will return to gradient based and ICA approaches in our discussion of regularized methods.

The ordered subsets EM (OSEM) algorithm (Hudson and Larkin 1994) is a modification of the EM algorithm in which each update uses only a subset of the data. Let $\{S_i\}_{i=1}^p$ be a disjoint partition of the integer interval $[1, M] = \bigcup_{i=1}^p S_i$. Let k denote the index for a complete cycle and i the index for a sub-iteration, and define $f^{(k,0)} = f^{(k-1)}$, $f^{(k,p)} = f^{(k)}$. Then the update equation for OSEM is given by

$$f_j^{(k,i)} = \frac{f_j^{(k,i-1)}}{\sum_{i \in S_i} p_{ij}} \sum_{i \in S_i} \frac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k,i-1)}},$$
$$\text{for } j = 1, \ldots, N, \quad i = 1, \ldots, p. \tag{14}$$

Typically, each subset will consist of a group of projections, with the number of subsets equal to an integer fraction of the total number of projections. "Subset balance" is recommended in (Hudson and Larkin 1994), i.e. the subsets should be chosen so that an emission from each pixel has equal probability of being detected in each of the subsets. In practice this can be difficult to achieve due to spatially varying attenuation and detector sensitivities. The grouping of projections within the subsets will alter both the convergence rate and the sequence of images generated. To avoid directional artifacts, the projections are usually chosen to have maximum separation in angle in each subset. Earlier examples of iterating over subsets of the data for ML estimation in emission tomography can be found in Hebert, Leahy and Singh (1990), Holte *et al.* (1990). OSEM produces remarkable improvements in convergence rates in the early iterations, although subsequent iterations over the entire data is required for ultimate convergence. Byrne (1997) has reported the empirical observation that OSEM seems to enter a limit cycle condition when the number of subsets remains greater than one, but currently there has been no proof that OSEM will always exhibit such behavior.

As with the original EM algorithm, OSEM produces high variance at large iteration numbers. This is typically controlled using either early termination or post smoothing of the image. Although OSEM does not converge in general, it is currently the most widely used iterative method for statistically based reconstruction in emission tomography. This is primarily due to the obvious improvements in image quality over standard filtered backprojection methods coupled with the relatively low computational cost involved in using this method.

A more general treatment of block iterative algorithms, of which OSEM is a special case, is given in Byrne (1997). Convergence properties for a number of block iterative methods are investigated and their limit cycle behavior for inconsistent data described. Another interesting variant on OSEM is the row-action maximum likelihood algorithm (RAMLA) of Browne and De Pierro (1996). This is similar to OSEM, but is shown to converge to a true ML solution under certain conditions.

While the EM and OSEM methods were originally derived for the pure Poisson model, modifications for the case of an offset due to randoms and scatter or increase in variance due to randoms subtraction have also been developed (Lange and Carson 1984, Politte and Snyder 1991, Qi *et al.* 1998b).

4.2. *Bayesian methods and other forms of regularization*

The EM algorithm can be directly extended to include prior terms by using the generalized EM (GEM) method (Dempster, Laird and Rubin 1977, Hebert and Leahy 1989). The treatment of the complete data remains the same as for ML-EM, so that the E-step given in (12) does not change. With the addition of the prior, the M-step must now maximize the log posterior given the complete data, $\boldsymbol{w}$, i.e.

M-step:

$$\boldsymbol{f}^{k+1} = \arg\max_{\boldsymbol{f}} \sum_j \left(f_j^{(k)} e_j\left(\boldsymbol{f}^{(k)}\right) \times \log\left(f_j\right) - f_j \sum_i p_{ij} \right) - \beta U(\boldsymbol{f}) \tag{15}$$

where $e_j(\boldsymbol{f}^{(k)}) = \sum_i p_{ij} y_i / (\sum_l p_{il} f_l^{(k)})$. In the absence of a non-negativity constraint, the necessary condition can be obtained by differentiating and setting the result to zero:

$$\frac{f_j^{(k)} e_j\left(\boldsymbol{f}^{(k)}\right)}{f_j} - \sum_i p_{ij} - \beta \frac{\partial}{\partial f_j} U(\boldsymbol{f}) = 0, \quad j = 1, \ldots, N \tag{16}$$

For the case where a non-negativity constraint is used, (16) can be replaced by the appropriate Kuhn-Tucker conditions. Direct solutions of (16) exist only for priors in which the voxels are statistically independent, such as the gamma prior (Lange, Bahn and Little 1987, Wang and Gindi 1997). For the case where voxels are not independent, a gradient search can be applied to the optimization subproblem in (16) (Hebert and Leahy 1989).

Green (1990) proposed a "one-step-late" (OSL) algorithm to solve (16). The partial derivatives of $U(\boldsymbol{f})$ are evaluated at the current estimate $f^{(k)}$, resulting in the simple update equation

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{\sum_i p_{ij} + \beta \frac{\partial}{\partial f_j} U(\boldsymbol{f})|_{\boldsymbol{f}=\boldsymbol{f}^{(k)}}} e_j\left(\boldsymbol{f}^{(k)}\right). \tag{17}$$

This procedure only converges to a MAP solution under restricted conditions, although Lange (1990) described a modified version of OSL with stronger convergence properties. De Pierro (1995) used an alternative functional substitution method to solve the problem. The objective function is approximated locally by a separable function of the current estimate $f^{(k)}$, so that the M-step involves only a one dimensional maximization that can be solved using either an analytic or Newton-Raphson method. This technique has provable convergence with an appropriate choice of the approximating separable function.

The GEM algorithm is readily modified for use with compound MRFs. In that case a set of binary line site variables must also be estimated. These variables are appended to the set of parameters to be estimated in the M-step. Since these variables are binary, standard gradient optimization methods cannot be applied. Instead, mean field annealing methods have been used (Leahy and Yan 1991, Bilbro *et al.* 1992, Gindi *et al.* 1991).

An alternative to the GEM algorithm is the space-alternating generalized EM (SAGE) method (Fessler and Hero 1995, McLachlan and Krishnan 1997). Unlike the EM algorithms which update image voxels simultaneously, SAGE updates image voxels sequentially using a sequence of small "hidden" data spaces. Because the sequential update decouples the M-step, the maximization can often be performed analytically. Hidden data spaces that are less informative than those used for ordinary EM are able to accelerate the convergence rate yet maintain the desirable monotonicity properties of EM algorithms.

Attempts to produce faster convergence have involved returning to more generic optimization techniques based either on gradient or coordinate-wise ascent. The gradient ascent methods (Kaufman 1987, Mumcuoglu, Leahy and Cherry 1996, Fessler and Ficaro 1996) employing preconditioners and conjugate gradients can give very fast convergence. They are also easily extended to include estimation of line processes using mean field annealing methods (Mumcuoglu *et al.* 1994, Bilbro *et al.* 1992).

A major problem in using gradient-based methods is the incorporation of the non-negativity constraint. Attempts to address this problem include using restricted line searches (Kaufman 1987), bent line searches (Kaufman 1987), penalty function methods (Mumcuoglu *et al.* 1994) and active set approaches (Kaufman 1993, Mumcuoglu, Leahy and Cherry 1996).

The non-negativity constraint is more easily dealt with using coordinate-wise updates. While there are a number of variations on this basic theme (Fessler 1994, Bouman and Sauer 1996, Sauer and Bouman 1993), the essence of these methods is to update each voxel in turn so as to maximize the objective function with respect to that voxel. Given the current estimate $\boldsymbol{f}^{(k)}$, the update for the jth voxel is

$$f_j^{(k+1)} = \arg\max_{x \geq 0} [L(\boldsymbol{y} \mid \boldsymbol{f}) - \beta U(\boldsymbol{f})]_{\boldsymbol{f}=(f_1^{(k+1)}, f_2^{(k+1)}, \ldots, f_{j-1}^{(k+1)}, x, f_{j+1}^{(k)}, \ldots, f_N^{(k)})}. \tag{18}$$

To solve the 1D maximization problem, polynomial approximations of the log likelihood function can be used to reduce the update step to closed form (Fessler 1995). The Newton-Raphson method can be used for the general case (Bouman and Sauer 1996). These methods can easily incorporate the non-negativity constraint and achieve similar convergence rates to preconditioned conjugate gradient (PCG) methods. If an ICA algorithm updates voxels in a raster-scan fashion, then the algorithm will exhibit a faster convergence rate in the scan direction than in the orthogonal direction. To avoid this problem, it is preferable to update the image voxels using either four different raster-scan orderings or a random ordering (Fessler 1994).

The computational cost of statistical reconstruction methods are heavily dependent on the specific implementation as well as the number of iterations required and the number of forward and backward projection operations per iteration. The ICA methods

access one voxel per iteration so that it is important that the projection matrix be stored in a voxel-driven format. However, we have found that it is easier to achieve efficient storage of the projection matrix described in Section 2 using a ray-driven format, which is more suitable for gradient based methods. Partitioning the updates among multiple processors is also more straightforward for the gradient based methods than ICA. However, both ICA and gradient based methods produce substantially faster convergence than the EM algorithm and its variants, and have the advantage over the OSEM method that, when used to compute a MAP estimate, they are stable at higher iterations so that selection of the stopping point of the algorithm is not critical.

4.3. *Parameter selection*

A key problem in the use of regularized or Bayesian methods is the selection of the regularization parameters, or equivalently, the hyperparameters of the prior. MAP estimates of the image $\boldsymbol{f}$ computed from (10) are clearly functions of β, which controls the relative influence of the prior and that of the likelihood. If β is too large, the prior will tend to have an over-smoothing effect on the solution. Conversely, if it is too small, the MAP estimate may be unstable, reducing to the ML solution as β goes to zero.

Data-driven selection of the hyperparameter is often performed in an ad hoc fashion through visual inspection of the resulting images. Making the parameter user selectable is analogous to the case in filtered backprojection image reconstruction where the user selects a filter cut-off frequency to choose image resolution and hence effects a subjective trade-off between bias and variance. Objective measures of image quality can also serve as the basis for user selection of the hyperparameter. When combined with the method for uniform resolution discussed below (Fessler and Rogers 1996), one can build an object independent table relating the prior parameter to the spatial resolution of the resulting image. The parameter can then be selected for the desired spatial resolution by the user. Alternatively, the hyperparameter can be selected to maximize the local contrast to noise ratio as described in Qi and Leahy (1998).

We can separate other approaches for choosing β in to two broad classes: (i) treating β as a regularization parameter and applying techniques such as generalized cross validation and the L-curve; and (ii) estimation theoretic approaches such as maximum likelihood.

The generalized cross-validation (GCV) method (Craven and Wahba 1979) has been applied in Bayesian image restoration and reconstruction (Johnson 1994b). Several difficulties are associated with this method: the GCV function is often very flat and its minimum is difficult to locate numerically (Varah 1983). Also the method may fail to select the correct hyperparameter when measurement noise is highly correlated (Wahba 1990). For problems of large dimensionality, this method may be impractical due to the amount of computation required. A Monte Carlo approximation of GCV was proposed to eliminate this drawback (Girard 1995). The L-curve method is based on the empirical observation that the corner of the L-curve corresponds to a good choice of β in terms of other validation measures (Hansen 1992). The L-curve has similar performance to GCV for uncorrelated measurement errors but, under certain restrictions, also works for correlated errors (Hansen 1992). Unfortunately the corner of the L-curve is difficult to find without multiple evaluations of the MAP solution for different hyperparameter values so that the computational cost is high.

As an alternative to the regularization based methods, the hyperparameter can be selected using ML estimation. The problem can be viewed in an incomplete/complete data framework in which the image is the (unobserved) complete data and the PET coincidence data are the (observed) incomplete data. Hyperparameter selection is therefore a natural candidate for the EM algorithm although the high dimensionality of the densities involved makes a true EM approach impractical. Markov Chain Monte Carlo methods (Besag *et al.* 1995) have been proposed for overcoming the intractability of ML parameter estimation (Geman and McClure 1987, Zhang, Modestino and Langan 1994, Saquib, Bouman and Sauer 1998, Geyer and Thompson 1992), but the computation costs remain high. Zhou, Leahy and Qi (1997) developed a ML approach based on mean field approximation. In this method, separable approximations of the prior and posterior densities are used to compute an approximate ML solution with lower computational cost than the MCMC methods. Other estimation methods, such as the generalized maximum likelihood approaches (Besag 1986, Lakshmanan and Derin 1989) and the method of moments (Geman and McClure 1987, Manbeck 1990), have also been studied. These methods have lower computational cost, but they do not share the desirable properties of true ML estimation and often exhibit poor performance.

5. Examining estimator performance

5.1. *Resolution*

Shift invariant linear imaging systems are often characterized by their point spread function (PSF). Noiseless data from a point source will produce an image of the PSF so that measurement of the full-width-at-half-maximum (FWHM) of the PSF is a measure of system resolution. This measure is useful for images reconstructed using filtered backprojection since the reconstruction procedure is linear. For non-linear estimators, PSFs are spatially variant and object dependent. Therefore, the PSF can only be examined locally and with a specific object. The *local impulse response* (Stamos *et al.* 1988, Fessler and Rogers 1996) or the effective local Gaussian resolution (Liow and Strother 1993) have been used to quantify the resolution properties of the statistical reconstructions.

It has been shown by Fessler and Rogers (1996) that the use of a shift invariant regularizing function or prior in (10) will produce spatially variant resolution. This is due to the decreasing influence of the prior as the noise variance in the data decreases; since the noise is spatially variant, so will be the influence of the prior and hence so will the resolution. These authors propose a

data dependent quadratic penalty function from which a nearly uniform local impulse response can be obtained. As mentioned in Section 4.3, this spatially-invariant property enables the selection of a "hyperparameter" based on the desired image resolution. In this case, the spatially varying weighting gives rise to a predictable, and shift invariant, resolution analogous to the FWHM resolution reported for the linear reconstruction methods. Of course, in this case, (10) cannot be viewed as a true posterior density since the prior term is data dependent.

5.2. *Estimator bias and variance*

Closed-form expressions of estimator bias and variance are easily derived for linear reconstruction methods (Barrett 1990). Variance properties of linear FBP reconstruction have been studied intensively, e.g. Alpert *et al.* (1982), Huesman (1984), Palmer *et al.* (1985), Carson *et al.* (1993), Maitra and O'Sullivan (1998). Derivation for the case of non-linear estimators is substantially more difficult. Monte Carlo studies can always be used to study the performance of any of the estimators and algorithms discussed above and has been widely used to explore bias-variance trade-offs in iterative algorithms, e.g. Carson *et al.* (1994). More recently there has been a great deal of progress in the development of approximate analytic expressions for estimator bias and variance which make it practical to explore algorithm behavior in a far more efficient manner.

An important advance was made by Barrett, Wilson and Tsui (1994) who derived approximate recursive formulae for computing the first and second order statistics for the ML EM algorithm as a function of iteration. The results agree well with Monte Carlo studies for the lower iterations at which the EM algorithm is usually terminated, and also at higher iterations when a large number of photon pairs are detected (Wilson, Tsui and Barrett 1994). Wang and Gindi (1997) made a further advance by extending this analysis to the subset of the GEM algorithms for MAP estimation in which a closed form update step is used. These results were able to accurately approximate the first and second order statistics of the GEM algorithm for two special cases: (i) the case in which the prior is an independent gamma distribution on the voxel intensities: and (ii) the OSL algorithm for multivariate Gaussian priors.

The main limitation of these methods is that explicit update equations are required. A larger class of algorithms for PET estimation have implicitly defined solutions which require numerical optimization at each iteration. To derive the statistics of these estimators, (Fessler 1996) studied the behavior at fixed points of the iterations. The objective functions must satisfy certain differentiability criteria, and have a unique, stable fixed point which can be found as the point where the partial derivatives are zero. As a result, inequality constraints and stopping rules are precluded. Fortunately, the non-negativity constraints in PET image reconstruction have little effect on the nonzero voxel locations, so the mean and variance approximations for an unconstrained estimator may agree closely with the actual performance of an estimator implemented with non-negativity constraints (Fessler 1996). Comparison of these formulae with results of Monte Carlo studies showed generally good agreement except in regions of very low activity. Qi and Leahy (1999b) have developed a method for compensating for the effect of the non-negativity constraint in variance.

Recently, we have extended this approach by deriving simplified approximate expressions for the resolution and covariance of MAP reconstructions (Qi and Leahy 1998). These results are fast to compute requiring a combination of backward projection and Fourier transforms. These closed form expressions avoid the numerical matrix inversion required in Fessler (1996) and hence reveal more clearly the relationship between the hyperparameter and the variance and resolution. The methods in Fessler (1996) and Qi and Leahy (1998) can also be used in "plug-in" form to compute surprisingly accurate estimates of the estimator variance using a single data set.

Hero, Fessler and Usman (1996) proposed an alternative to exploring estimator bias-variance tradeoffs by using the uniform Cramer-Rao bound. They introduce a delta-sigma plane, which is indexed by the norm of the estimator bias gradient and the variance of the estimator. The norm of the bias gradient is related to the maximum variation in the estimator bias function over a neighborhood of the parameter space. The larger the norm of the bias gradient, the larger the uncompensatable bias the estimator will have. Using a uniform Cramer-Rao bound on the estimator variance, a delta-sigma tradeoff curve can be generated, which defines an unachievable region on the delta-sigma plane for a specified statistical model. This delta-sigma tradeoff curve can then be used to compare different statistical models and also to assess estimator performance.

In concluding this section on classical performance measures for image estimates, a note of caution is appropriate. The limited resolution of PET systems and the finite size of the voxels give rise to "partial volume" effects. A user-selected region of interest over which the uptake is to be quantified will often include partial volume voxels which are not entirely contained in the anatomical region of interest simply because of the finite size of the voxels. This effect is exacerbated by the finite resolution of the PET images, so that activity that is actually confined to a specific anatomical structure will spread into neighboring regions. These partial volume effects produce biases in the estimated activity that can far exceed errors that result from the limitations of the reconstruction method itself. Methods for partial volume correction when computing regional activity are currently being developed, e.g. Meltzer *et al.* (1990), but are, as yet, not in widespread use.

5.3. *Task specific evaluation*

There are two distinct applications of clinical PET. The first of these is to provide quantitative measures of the uptake of the tracer in a volume of interest specified by the user. These quantitative measures can be used in monitoring disease progression and in pharmacokinetic studies. The second major application of PET is in the detection of small cancerous lesions in the

body indicating the presence of primary cancers or metastatic disease in the patient (Strauss and Conti 1991). It is important in developing reconstruction methods, and even more so when evaluating them, that the ultimate applications are kept in mind. While estimator bias and variance are clearly relevant to the task of accurate quantitation, they do not directly reflect on the algorithm's performance for lesion detection. Here we will briefly discuss some of the approaches used for algorithm evaluation that are applicable to lesion detection.

The gold standard for measuring lesion detectability is the ROC (receiver operating characteristic) study (Gilland *et al.* 1992, Llacer 1993). A study comparing false positive vs. false negative rates for human observers for the task of lesion detection in images reconstructed using two or more different methods indicates which is superior for this task. These tests require access to data in which the presence or absence of lesions is independently verified. In practice, real clinical studies of this type are virtually impossible to find in the numbers necessary to establish statistically significant differences between different reconstruction methods. Instead these studies can be performed by introducing artificial (computer generated) lesions into otherwise normal scans (Llacer 1993, Farquhar 1998) and again evaluating two or more algorithms using human observers.

While computer generated lesions can be used to produce realistic images for ROC studies, the need for human observers makes these studies extremely time consuming and limits the range of parameters that can realistically be explored. Using computers as "observers" is a potential solution to the problem. There is now a substantial body of literature dealing with the development of computer observers that reflect human observer performance in lesion detection (Yao and Barrett 1992, Abbey and Barrett 1996). As these techniques mature they can be used to compare algorithm performance through computer generated ROC curves (e.g. King, de Vries and Soares 1997, Chan, Leahy and Cherry 1997).

In addition to simple visual inspection of PET images, semi-quantitative analysis is often performed as an aid in deciding on the presence or absence of a lesion. These measures include standardized uptake ratios (SURs) and ratios of lesion to background (Adler *et al.* 1993, Lowe *et al.* 1994, Duhaylongsod *et al.* 1995). As with the ROC studies, appropriate clinical data is scarce. Instead studies that reflect the performance that may be achieved using these measures can be performed by computing contrast recovery coefficients (CRC) vs. background noise variance using lesions in simulated or phantom data (Liow and Strother 1991, Kinahan *et al.* 1995).

6. Examples

We conclude this review with a few recent results. The first study, Fig. 6, shows coronal sections of a 3D glucose metabolism image of a normal volunteer collected using the tracer ^{18}FDG (Cherry and Phelps 1996). The data was collected in 3D mode using a whole body CTI ECAT HR+ scanner. To produce the image, data

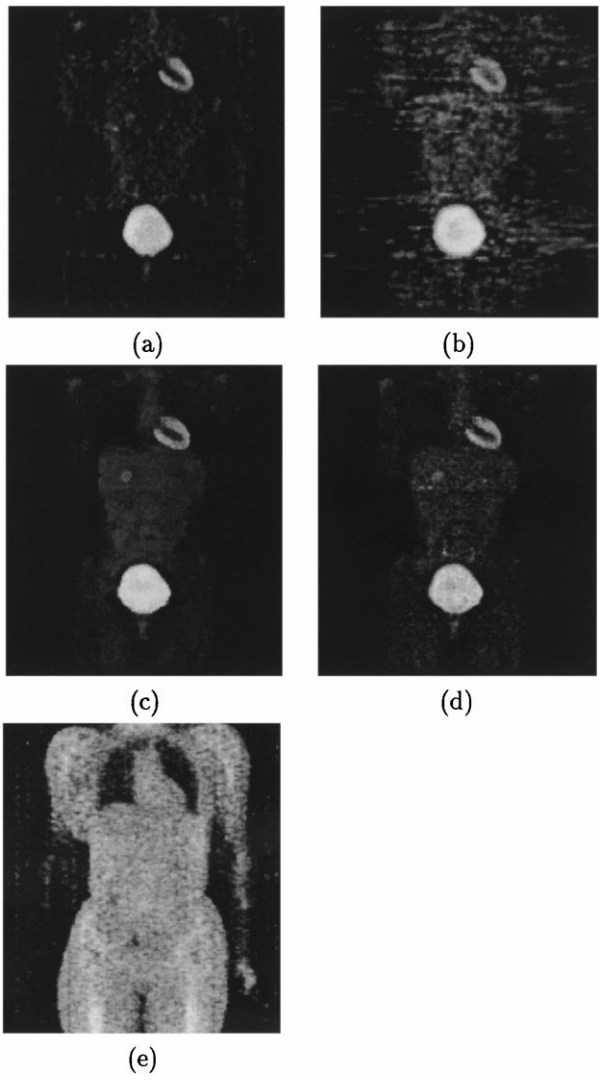

Fig. 6. *Example of a whole body ^{18}FDG scan: (a) FBP without attenuation correction; (b) FBP with attenuation correction; (c) MAP; (d) OSEM; (e) MAP reconstruction of transmission image.*

was collected for several different bed positions of the subject within the scanner. The 3D images were then reconstructed for each bed position. The images shown represent a single section through the subject along the axis of the scanner. In each of the emission images, the highest uptake can clearly be seen in the heart, in which the glucose is metabolized, and the bladder, in which the unmetabolized tracer accumulates. These results show the lower noise property of the MAP (Fig. 6(c)) and OSEM (Fig. 6(d)) images in comparison to the FBP study (Fig. 6(b)). The lower uptake of the tracer in muscle tissue and the abdominal organs is far more clearly delineated in the MAP and OSEM images than the FBP. Also shown in this figure is an attenuation image (Fig. 6(e)) that was reconstructed from a transmission scan of the subject–this image was used to compute the attenuation correction factors as described in Qi *et al.* (1998b). The attenuation image reveals the lower attenuation in the lungs and higher attenuation in bone. We note that the FBP images are

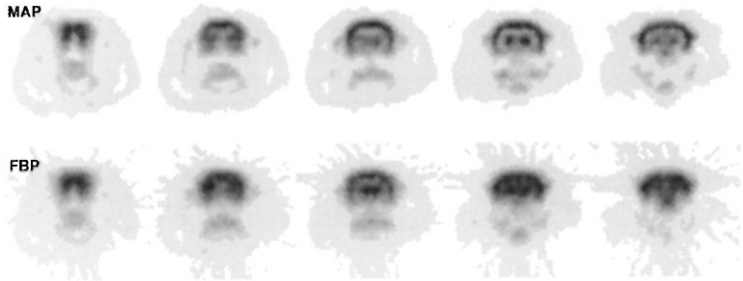

Fig. 7. *Example of a rat brain scan using the high resolution microPET scanner. Shown in this figure are coronal slices through the rat brain reconstructed using FBP with a ramp reconstruction filter for maximum resolution (top row) and MAP (bottom row). Note the higher tracer uptake in the cortical and subcortical grey matter that is clear in the MAP images but not in the FBP.*

far less noisy if they are not attenuation corrected as shown in Fig. 6(a), but in this case the images are not quantitative.

The above example shows improvements in image quality that can be realized using statistical methods when the number of detected coincidence events per voxel is low. Shown in Fig. 7 are images reconstructed from the microPET small animal scanner (Cherry *et al.* 1997). This is a small high resolution 3D scanner that utilizes the scintillator LSO in its detectors. In this example, the number of counts per image voxel was far higher so that it is the accuracy of the system model, rather than noise, that determines the ultimate image resolution. Shown here are coronal sections through an ^{18}FDG study of a rat brain reconstructed using FBP and MAP. The FBP images were reconstructed for maximum resolution yet are unable to match that of the MAP images which clearly reveal the higher uptake of ^{18}FDG in the cortex and subcortical nuclei than the intervening white matter.

Comparison of different reconstruction algorithms can be performed using theoretical analysis, Monte Carlo simulation studies, and experimental studies of phantom and human data, as described in Section 5. In Fig. 8 we show the results of a phantom study performed using simulated lesions placed in a simple thorax phantom. The lesions had higher activity than the background structures – a detailed description of the experiment can be found in Qi *et al.* (1998b). Shown in the figure are plots of the contrast recovery coefficient (CRC) (i.e. the measured lesion to background contrast in the reconstructed image normalized by the true constrast in the phantom) vs. the estimated background noise standard deviation. As higher degrees of smoothing are used, the background noise is reduced but so is the contrast in the lesion. At a particular noise level, a higher contrast is indicative of a higher probability of detecting a lesion. Shown in the figure are the results for four different sizes of lesion computed for various degrees of smoothing. These results show that both OSEM and MAP consistently produce superior contrast recovery than FBP at matched noise levels, and that the MAP method produces a small additional improvement over OSEM.

Our final example shows a result obtained using the theoretical approximations described in Qi and Leahy (1998, 1999). In Fig. 9 we show the variance of each voxel computed using the theoretical approximation and a Monte Carlo method.

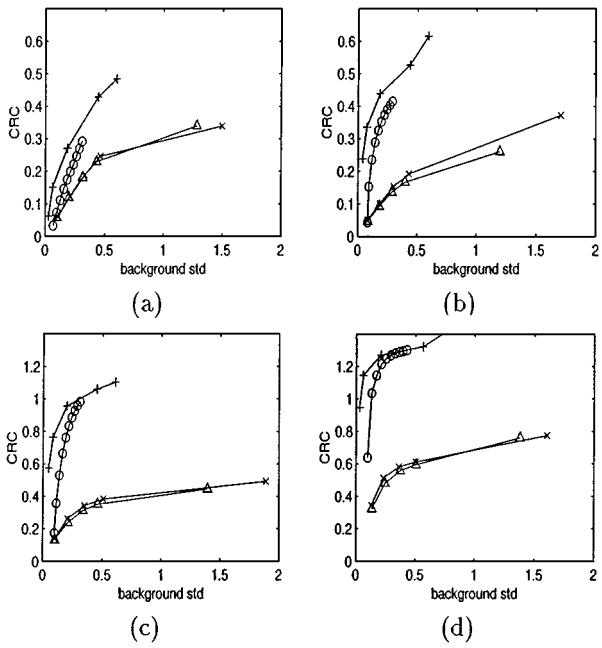

Fig. 8. *Comparison of CRCs versus background standard deviation for four different sized lesions in a thorax phantom reconstructed using MAP* ('+')*, OSEM* ('$\circ$')*, and FBP* ('$\times$' *and* '$\triangle$') *– the two different FBP results are for two different voxel sizes. Lesion size: (a) 0.2 ml lesion, (b) 0.45 ml lesion, (c) 1.0 ml lesion, and (d) 1.9 ml lesion*

The data represent a simulation of the CTI ECAT HR+ scanner operating in 2D mode. An average of 200,000 counts per data set were generated for 8,000 Monte Carlo runs. From each of these a 128×128 voxel image was reconstructed using the MAP method in Qi *et al.* (1998b). The variances were computed using the Monte Carlo method and also using the method in Qi and Leahy (1999). There is generally good agreement between the theoretical and Monte Carlo variance estimates. The advantage of the theoretical method over the Monte Carlo is that the computation cost for the former is similar to that for a single reconstruction. Furthermore, these approximations can be used in a "plug-in" mode to estimate voxel-wise variance from a single data set.

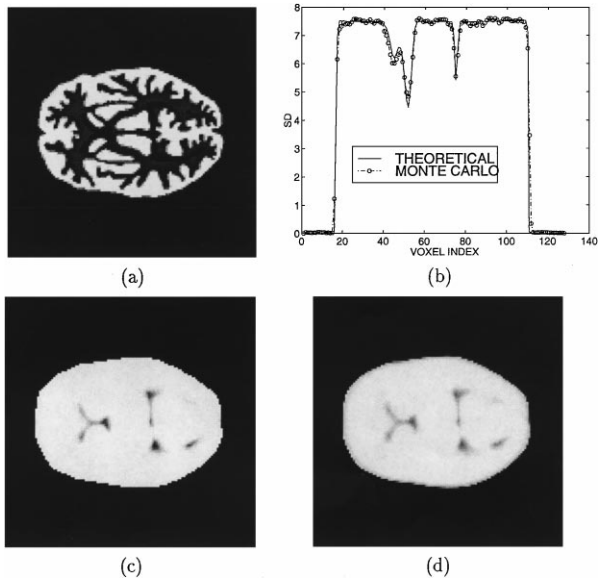

Fig. 9. *Example of variance computation for a 2D brain phantom. (a) The brain phantom. (b) Comparison of profiles of variance images– solid line denotes theoretical results and "×" Monte Carlo result. (c) variance image computed using theoretical approximation, (d) variance image from 8000 Monte Carlo reconstructions*

7. Conclusions

We have attempted to provide a brief introduction to issues involved in computing images from PET data and the methods that have been developed to solve this problem. While we have discussed various approaches to evaluating algorithm performance, we have not addressed the issue of relative performance of different algorithms. It is clear from the substantial literature on statistically based PET reconstruction algorithms that virtually any implementation of an ML or MAP estimator will produce generally superior performance to the standard filtered backprojection protocols that are used in most clinical PET facilities. The differences between the various ML and MAP implementations are probably less striking, but nonetheless important if, for instance, they impact on the specificity and sensitivity of PET in early cancer detection.

The two major objections to the use of iterative statistically based methods for PET image reconstruction that have often been raised are that the computational cost is too high and that the behavior of these nonlinear methods is not well understood. With recent advances in low cost, high performance computers, the first of these objections is no longer a significant obstacle to the adoption of statistically based methods in clinical PET. Recent advances in producing accurate measures of image bias and variance have done much to answer the second objection. As PET instrumentation matures, it appears reasonable to expect that these approaches will be adopted as the preferred method for image estimation.

Acknowledgments

This work was supported by the National Cancer Institute under Grant No. RO1 CA59794. The authors would like to thank Mark Berman of Macquarie University, Jeff Fessler of the University of Michigan, and the anonymous reviewers for their comments on an earlier version of the manuscript. We would also like to thank Simon Cherry and Arion Chatziioannou for providing the data shown in Fig. 7.

Note

1. This picture is rather simplified since 2D systems do allow detection of events between *adjacent* rings. These are used to reconstruct additional transaxial images, so that the thickness of each plane is half of the axial extent of a single detector ring and the number of reconstructed planes in a 2D scanner is usually $2P - 1$ where P is the number of detector rings. This arrangement is illustrated in Fig. 2(a).

References

Abbey C.K. and Barrett H.H. 1996. Observer signal-to-noise ratios for the ML-EM algorithm. In: Proc. of SPIE, Vol. 2712, pp. 47–58.

Adler L.P. Crowe J.P., Al-Kaisi N.K., and Sunshine J.L. 1993. Evaluation of breast masses and axillary lymph nodes with [F-18] 2-deoxy-2-fluoro-d-glucose PET. Radiology 187(3): 743–750.

Alpert N.M., Chesler D.A., Correia J.A., Ackerman R.H., Chang J.Y., Finklestein S., adn G.L. Brownell S.M.D., and Taveras J.M. 1982. Estimation of the local statistical noise in emission computed tomography. IEEE Transactions on Medical Imaging MI-1: 142–146.

Baker J. 1991. Spatially variant tomographic imaging: estimation, identification and optimization. PhD thesis, Lawrence Berkeley Laboratory, University of California.

Barrett H.H. 1990. Objective assessment of image quality: Effects of quantum noise and object variability. Journal of the Optical Society of America A 7(7): 1266–1278.

Barrett H. and Swindell W. 1981. Radiological Imaging. Academic Press, Vol. 1.

Barrett H.H., Wilson D.W., and Tsui B.M.W. 1994. Noise properties of the EM algorithm: I. theory. Physics in Medicine and Biology 39: 833–846.

Besag J. 1986. On the statistical analysis of dirty pictures. J. Royal Statist. Soc. B 48: 259–302.

Besag J., Green P., Higdon D., and Mengersen K. 1995. Bayesian computation and stochastic systems. Statistical Science 10: 3–66.

Bilbro G.L., Snyder W.E., Garnier S.J., and Gault J. 1992. Mean field annealing: A formalism for constructing GNC-like algorithms. IEEE Transactions on Neural Networks 3(1): 131–138.

Bouman C. and Sauer K. 1996. A unified approach to statistical tomography using coordinate descent optimization. IEEE Transactions on Image Processing 5(3): 480–492.

Bowsher J.E., Johnson V.E., Turkington T.G., Jaszczak R.J., Floyd C.E., and Coleman R.E. 1996. Bayesian reconstruction and use of

anatomical a priori information for emission tomography. IEEE Transactions on Medical Imaging 15(5): 673–686.

Browne J. and De Pierro A.R. 1996. A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography. IEEE Transactions on Medical Imaging 15: 687–699.

Byrne C. 1997. Convergent block-iterative algorithms for image reconstruction from inconsistent data. IEEE Transactions on Image Processing 6: 1296–1304.

Carson R.E., Yan Y., Daube-Witherspoon M.E., Freedman N., Bacharach S.L., and Herscovitch P. 1993. An approximation formula for the variance of PET region-of-interest values. IEEE Transactions on Medical Imaging 12(2): 240–250.

Carson R., Yan Y., Chodkowski B., Yap T., and Daube-Witherspoon M. 1994. Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm. IEEE Transactions on Medical Imaging 13(3): 526–537.

Casey M., Gadagkar H., and Newport D. 1995. A component based method for normalization in volume PET. In: 3rd International Meeting on fully 3D Reconstruction in Nuclear Medicine and Radiology, Aix-les-Bains, pp. 67–71.

Censor Y. 1983. Finite series-expansion reconstruction methods. IEEE Proceed. 71(3): 409–418.

Chan M., Herman G.T., and Levitan E. Bayesian image reconstruction using a high-order interacting MRF model. 1995. In: Proceedings of the 8th International Conference on Image Analysis and Processing, San Remo, Italy, Springer, pp. 609–614.

Chan M.T., Leahy R., and Cherry S. 1997. Comparing lesion detection performances of pet image reconstruction algorithms: A case study. IEEE Transactions on Nuclear Science 44(4): 1558–1563.

Chen C., Lee S., and Cho Z. 1991. Parallelization of the EM algorithm for 3D PET image reconstruction. IEEE Transactions on Medical Imaging 10: 513–522.

Cherry S.R. and Phelps M.E. 1996. Imaging brain function with positron emission tomography. In: Toga A.W. and Mazziotta J.C. (Eds.), Brain Mapping: The Methods. Academic Press. ch. 8.

Cherry S.R., Shao Y., Siegel S., Silverman R.W., Meadors K., Young J., Jones W.F., Newport D., Moyers C., Mumcuoglu E., Andreaco M., Paulus M., Binkley D., Nutt R., and Phelps M.E. 1997. MicroPET: A high resolution PET scanner for imaging small animals. IEEE Transactions on Nuclear Science 44(3): 1161–1166.

Coakley K.J. 1991. A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra. IEEE Transactions on Nuclear Science 38(1): 9–15.

Craven P. and Wahba G. 1979. Smoothing noisy data with spline function. Numerische Mathematik 31: 377–403.

De Pierro A.R. 1995. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. IEEE Transactions on Medical Imaging 14: 132–137.

Defrise M., Kinahan P.E., Townsend D.W., Michel C., Sibomana M., and Newport D.F. 1997. Exact and approximate rebinning algorithms for 3-D PET data. IEEE Transactions on Medical Imaging 16(2): 145–158.

Dempster A., Laird N., and Rubin D. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of Royal Statistical Society, Series B 39(1): 1–38.

Duhaylongsod F.G., Lowe V.J., Patz E.F., Vaughn A.L., Coleman R.E., and Wolfe W.G. 1995. Detection of primary and recurrent lung cancer by means of F-18 fluorodeoxyglucose positron emission tomography (FDG PET). The Journal of Thoracic and Cardiovascular Surgery 110(1): 130–140.

Farquhar T.H. 1998. Improved Lesion Detection in Whole Body PET. PhD thesis, University of California, Los Angeles.

Fessler J.A. 1994. Penalized weighted least squares image reconstruction for PET. IEEE Transactions on Medical Imaging 13: 290–300.

Fessler J. 1995. Hybrid polynomial objective functions for tomographic image reconstruction from transmission scans. IEEE Transactions on Image Processing 4(10): 1439–1450.

Fessler J. 1996. Mean and variance of implicitely defined biased estimators (such as penalized maximum likelihood): Applications to tomography. IEEE Transactions on Image Processing 5(3): 493–506.

Fessler J.A. and Ficaro E.P. 1996. Fully 3D PET image reconstruction using a Fourier preconditioned conjugate-gradient algorithm. In: Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference, Vol. 3, Anaheim, CA, pp. 1599–1562.

Fessler J.A., Ficaro E.P., Clinthorne N.H., and Lange K. 1997. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. IEEE Transactions on Medical Imaging 16(2): 166–175.

Fessler J.A. and Hero A.O. 1995. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. IEEE Transactions on Image Processing 4: 1417–1429.

Fessler J.A. and Rogers W.L. 1996. Spatial resolution properties of penalized-likelihood image reconstruction: Spatial-invariant tomographs. IEEE Transactions on Image Processing 9(5): 1346–1358.

Geman S. and Geman D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6(6): 721–741.

Geman S. and McClure D. 1985. Bayesian image analysis: An application to single photon emission tomography. In: Proc. of Statistical Computing Section of the American Statistical Association, pp. 12–18.

Geman S. and McClure D. 1987. Statistical methods for tomographic image reconstruction. In: Proceedings of the 46th Session of the International Statistical Institute, Bulletin of the ISI, Vol. 52, pp. 4–20.

Geyer C.J. and Thompson E.A. 1992. Constrained monte carlo maximum likelihood for dependent data. Journal of Royal Statistical Society B 54: 657–699.

Gilland D.R., Tsui B.M.W., Metz C.E., Jaszczak R.J., and Perry J.R. 1992. An evaluation of maximum likelihood-expectation maximization reconstruction for SPECT by ROC analysis. The Journal of Nuclear Medicine 33: 451–457.

Gindi G., Lee M., Rangarajan A., and Zubal I. 1991. Bayesian reconstruction of functional images using registered anatomical images as priors. In: Information Processing in Medical Imaging, XIIth IPMI International Conference, Wye, UK, pp. 121–131.

Gindi G., Lee M., Rangarajan A., and Zubal I.G. 1993. Bayesian reconstruction of functional images using anatomical information as priors. IEEE Transactions on Medical Imaging 12(4): 670–680.

Girard D. 1995. The fast Monte Carlo cross-validation and c_l procedures: Comments, new results and application to image recovery problems. Computational Statistics 10(3): 205–232.

Green P. 1990. Bayesian reconstructions from emission tomography data using a modified EM algorithm. IEEE Transactions on Medical Imaging 9(1): 84–93.

Green P. 1996. MCMC in image analysis. In: Gilks W., Richardson S., and Spiegelhalter D. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, ch. 21, pp. 381–399.

Grenander U. 1981. Abstract Inference. New York, John Wiley and Sons.

Haber S., Derenzo S., and Uber D. 1990. Application of mathematical removal of positron range blurring in positron emission tomography. IEEE Transactions on Nuclear Science 37(3): 1293–1299.

Hansen P.C. 1992. Analysis of discrete ill-posed problems by means of the l-curve. SIAM Review 34: 561–580.

Hebert T. and Leahy R. 1989. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. IEEE Transactions on Medical Imaging 8(2): 194–202.

Hebert T., Leahy R., and Singh M. 1990. 3D ML reconstruction for a prototype SPECT system. J. Opt. Soc. Amer. Part A, Optics and Image Science 7(7): 1305–1313.

Hero A.O., Fessler J.A., and Usman M. 1996. Exploring estimator bias-variance tradeoffs using the uniform CR bound. IEEE Transactions on Signal Processing 44: 2026–2041.

Higdon D.M., Bowsher J.E., Johnson V.E., Turkington T.G., Gilland D.R., and Jaszczak R.J. 1997. Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. IEEE Transactions on Medical Imaging 16: 516–526.

Hoffman E., Huang S., Plummer D., and Phelps M.P. 1982. Quantitation in positron emission computed tomography: Effect of nonuniform resolution. Journal of Computer Assisted Tomography 6(5): 987–999.

Holte S., Schmidlin P., Linden A., and Rosenqvist G. 1990. Iterative image reconstruction for positron emission tomography: A study of convergence and quantitation problems. IEEE Transactions on Nuclear Science 37(2): 629–635.

Huber P.J. 1981. Robust Statistics. New York, John Wiley and Sons.

Hudson H.M. and Larkin R.S. 1994. Accelerated image reconstruction using ordered subsets of projection data. IEEE Transactions on Medical Imaging 13(4): 601–609.

Huesman R. 1984. A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography. Physics in Medicine and Biology 29: 543–552.

Johnson V. 1994a. A note on stopping rules in EM-ML reconstructions of ECT images. IEEE Transactions on Medical Imaging 13(3): 569–571.

Johnson V.E. 1994b. A model for segmentation and analysis of noisy images. Journal of American Statistical Association 89(425): 230–241.

Johnson V., Wong W., Hu X., and Chen C. 1991. Bayesian restoration of PET images using Gibbs priors. In: Colchester A. and Hawkes D. (Eds.), Information Processing in Medical Imaging. Wiley-Liss, pp. 15–28.

Johnson C., Yan Y., Carson R., Martino R., and Daube-Witherspoon M. 1995. A system for the 3D reconstruction of retracted-septa PET data using the EM algorithm. IEEE Transactions on Nuclear Science 42(4): 1223–1227.

Kaufman L. 1987. Implementing and accelerating the EM algorithm for positron emission tomography. IEEE Transactions on Medical Imaging 6(1): 37–51.

Kaufman L. 1993. Maximum likelihood, least squares, and penalized least squares for PET. IEEE Transactions on Medical Imaging 12(2): 200–214.

Kearfott K.J. 1985. Practical considerations, comment in "A statistical model for positron emission tomography". Journal of American Statistical Association 80: 26–28.

Kinahan P.E., Matej S, Karp J.S., Herman G.T., and Lewitt R.M. 1995. A comparison of transform and iterative reconstruction techniques for a volume-imaging PET scanner with a large axial acceptance angle. IEEE Transactions on Nuclear Science 42: 2281–2287.

Kinahan P. and Rogers J. 1989. Analytic 3D image reconstruction using all dectected events. IEEE Transactions on Nuclear Science 36: 964–968.

King M., de Vries D., and Soares E. 1997. Comparison of the channelised hotelling and human observers for lesion detection in hepatic SPECT imaging SPIE 3036: 14–20.

Lakshmanan S. and Derin H. 1989. Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-11: 799–813.

Lange K. 1990. Convergence of EM image reconstruction algorithms with Gibbs smoothing. IEEE Transactions on Medical Imaging 9(4): 439–446. Correction: 10(2), June 1991, pp. 228.

Lange K., Bahn M., and Little R. 1987. A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. IEEE Transactions on Medical Imaging 6(2): 106–114.

Lange K. and Carson R. 1984. EM reconstruction algorithms for emission and transmission tomography. Journal of Computer Assisted Tomography 8(2): 306–316.

Leahy R. and Yan X. 1991. Incorporation of anatomical MR data for improved functional imaging with PET. In: Colchester A. and Hawkes D. (Eds.), Information Processing in Medical Imaging. Wiley-Liss, pp. 105–120.

Lee S.J., Rangarajan A., and Gindi G. 1995. Bayesian image reconstruction in SPECT using higher oreder mechanical models as priors. IEEE Transactions on Medical Imaging 14(4): 669–680.

Lewitt R. and Muehllehner G. 1986. Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation. IEEE Transactions on Medical Imaging 5(1): 16–22.

Liow J.S. and Strother S.C. 1991. Practical tradeoffs between noise, quantitation, and number of iterations for maximum likelihood-based reconstructions. IEEE Transactions on Medical Imaging 10(4): 563–571.

Liow J.S. and Strother S.C. 1993. The convergence of object dependent resolution in maximum likelihood based tomographic image reconstruction. Physics in Medicine and Biology 38: 55–70.

Llacer J. 1993. Results of a clinical receiver operating characteristic study comparing filtered backprojection and maximum likelihood estimator images in FDG PET studies. The Journal of Nuclear Medicine 34(7): 1198–1203.

Llacer J., Veklerov E., Coakley K., Hoffman E., and Nunez J. 1993. Statistical analysis of maximum likelihood estimator images of human brain fdg studies. IEEE Transactions on Medical Imaging 12(2): 215–231.

Lowe V.J., Hoffman J.M., DeLong D.M., Patz E.F., and Coleman R.E. 1994. Semiquantitative and visual analysos of FDG-PET images

in pulmonary abnormalities. The Journal of Nuclear Medicine 35(11): 1771–1776.

Maitra R. and O'Sullivan F. 1998. Variability assessment in PET and related generalized deconvolution methods. Journal of American Statistical Association 93(444): 1340–1355.

Manbeck K. 1990. Bayesian Statistical Methods Applied to Emission Tomography with Physical Phantom and Patient Data. PhD thesis, Brown University.

Matej S. and Lewitt R. 1996. Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. IEEE Transactions on Medical Imaging 15(1): 68–78.

McLachlan G. and Krishnan T. 1997. The EM Algorithm and Extensions: John Wiley & Sons, Inc., ch. 5.12.

Meltzer C.C., Leal J.P., Mayberg H.S., Jr. H.N.W., and Frost J.J. 1990. Correction of PET data for partial volume effects in human cerebral cortex by MR imaging. Journal of Computer Assisted Tomography 14(4): 561–570.

Mumcuoglu E., Leahy R., and Cherry S. 1996. Bayesian reconstruction of PET images: Methodology and performance analysis. Physics in Medicine and Biology 41: 1777–1807.

Mumcuoglu E., Leahy R., Cherry S., and Hoffman E. 1996. Accurate geometric and physical response modeling for statistical image reconstruction in high resolution PET. In: Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim, CA, pp. 1569–1573.

Mumcuoglu E., Leahy R., Cherry S., and Zhou Z. 1994. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. IEEE Transactions on Medical Imaging 13(4): 687–701.

Ollinger J. and Fessler J. 1997. Positron Emission Tomography. IEEE Signal Processing Magazine 14(1): 43–55.

Ollinger J., Johns G., and Burney M. 1992. Model-based scatter correction in three dimensions. In: Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference, Vol. 2, Orlando, Florida, pp. 1249–1251.

Palmer M.R., Bergstrom M., Beddoes M.P., and Plate B.D. 1985. Effects of detector wobble motion on image noise in positron emission tomography. IEEE Transactions on Medical Imaging MI-4: 58–62.

Politte D., and Snyder D. 1991. Corrections for accidental coincidences and attenuation in maximum likelihood image reconstruction for positron emission tomography. IEEE Transactions on Medical Imaging 10(1): 82–89.

Qi J., and Leahy R.M. 1999a. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. IEEE Transactions on Medical Imaging 18(4): 293–305.

Qi J. and Leahy R.M. 1999b. Fast computation of the covariance of map reconstructions of pet images. In: Proceedings of SPIE 3661: 344–355.

Qi J., Leahy R.M., Cherry S.R., Chatziioannou A., and Farquhar T.H. 1998a. High resolution 3D bayesian image reconstruction using the microPET small animal scanner. Physics in Medicine and Biology 43(4): 1001–1013.

Qi J., Leahy R.M., Hsu C., Farquhar T.H., and Cherry S.R. 1998b. Fully 3D bayesian image reconstruction for ECAT EXACT HR+. IEEE Transactions on Nuclear Science 45(3): 1096–1103.

Rajeevan N., Rajgopal K., and Krishna G. 1992. Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography. IEEE Transactions on Medical Imaging 11(1): 9–20.

Rockmore A., and Macovski A. 1976. A maximum likelihood approach to emission image reconstruction from projections. IEEE Transactions on Nuclear Science NS-23: 1428–1432.

Saquib S., Bouman C., and Sauer K. 1998. Ml parameter estimation for markov random fields, with applications to bayesian tomography. IEEE Transactions on Image Processing 7: 1029–1044.

Sauer K., and Bouman C. 1993. A local update strategy for iterative reconstruction from projections. IEEE Transactions on Signal Processing 41(2): 534–548.

Shepp L. and Logan B. 1974. The Fourier reconstruction of a head section. IEEE Transactions on Nuclear Science NS-21: 21–33.

Shepp L. and Vardi Y. 1982. Maximum likelihood reconstruction for emission tomography. IEEE Transactions on Medical Imaging 1(2): 113–122.

Silverman B., Jones M., Wilson J., and Nychka D. 1990. A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. Journal of Royal Statistical Society, Ser. B 52(2): 271–324.

Snyder D. and Miller M. 1985. The use of sieves to stabilize images produced with the EM algorithm for emission tomography. IEEE Transactions on Nuclear Science NS-32(5): 3864–3872.

Stamos J.A., Rogers W.L., Clinthorne N.H., and Koral K.F. 1988. Object-dependent performance comparison of two iterative reconstruction algorithms. IEEE Transactions on Nuclear Science 35(1): 611–614.

Strauss L. and Conti P. 1991. The application of PET in clinical oncology. Journal of Nuclear Medicine 32: 623–648.

Terstegge A., Weber S., Herzog H., Muller-Gartner H.W., and Hailling H. 1996. High resolution and better quantification by tube of response modelling in 3D PET reconstruction. In: Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim, CA, pp. 1603–1607.

Varah J. 1983. Pitfalls in the numerical solution of ill-posed problems. SIAM J. Sci. Comput. 4: 164–176.

Vardi Y., Shepp L., and Kaufman L. 1985. A statistical model for positron emission tomography. Journal of the American Statistical Association 80(389): 8–37.

Veklerov E. and Llacer J. 1987. Stopping rule for the MLE algorithm based on statistical hypothesis testing. IEEE Transactions on Medical Imaging 6(4): 313–319.

Wahba G. 1990. Spline Models for Observational Data. Cbms-Nsf Regional Conference Series in Applied Mathematics, 59. SIAM.

Wang W. and Gindi G. 1997. Noise analysis of MAP-EM algorithms for emission tomography. Physics in Medicine and Biology 42: 2215–2232.

Watson C.C., Newport D., and Casey M.E. 1995. A single scatter simulation technique for scatter correction in 3D PET. In: International Meeting on Fully Three Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Aix-les-Bain, France, pp. 255–268.

Weir I. 1997. Fully Bayesian reconstructions from single-photon emission computed tomography data. Journal of the American Statistical Association 92(437): 49–60.

Wilson D.W., Tsui B.M.W., and Barrett H.H. 1994. Noise properties of the EM algorithm: II. Monte Carlo simulations. Physics in Medicine and Biology 39: 847–872.

Yao J. and Barrett H.H. 1992. Predicting human performance by a channelized Hotelling model. In: SPIE Mathematical Methods in Medical Imaging, Vol. 1768, pp. 161–168.

Yavuz M. and Fessler J.A. 1998. Statistical image reconstruction methods for randoms-precorrected pet scans. Medical Image Analysis 2(4): 369–378.

Zhang J., W. Modestino J., and A. Langan D. 1994. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. IEEE Transactions on Image Processing 3: 404–420.

Zhou Z., Leahy R., and Qi J. 1997. Approximate maximum-likelihood hyperparameter estimation for gibbs-priors. IEEE Transactions on Image Processing 6: 844–861.