Invited Paper 21.1
46th Session of the ISI

Reprinted from Bull. Intl. Statist. Inst., Vol. LII(4), 1987, pp. 5-21.

STATISTICAL METHODS FOR TOMOGRAPHIC IMAGE RECONSTRUCTION

Stuart Geman and Donald E. McClure
Division of Applied Mathematics
Brown University, Providence, Rhode Island, U.S.A.

1. Introduction

Interest in statistical approaches to reconstruction problems in emission computed tomography was greatly enhanced by the work of Shepp and Vardi (1982) on the use of maximum likelihood (ML) methods. There are earlier instances of suggestions to regard the reconstruction problem as a statistical estimation problem; however, the demonstration of the versatility of the approach as well as the specification of algorithms *that work* were advanced substantially by Shepp and Vardi's work.

The image reconstruction problem, viewed as an estimation problem, is inherently nonparametric: one seeks an estimate of a function of general form on a continuous domain. As such, it is widely recognized that the estimates need to be regularized or smoothed, especially in "small sample" implementations. Various approaches to regularization have been suggested, including penalized ML, the method of sieves, and Bayesian methods. In Geman and McClure (1985), we proposed that *a priori* spatial information be built into a statistical reconstruction algorithm, in a Bayesian approach, by quantifying spatial constraints in the form of a Gibbs prior distribution. In this paper we will expand on our earlier description and present recent work on parameter estimation for the Gibbs priors, which leads to completely data-driven algorithms.

This application to single photon emission computed tomography (SPECT) follows a general Bayesian paradigm for problems in image processing and vision laid out in Geman and Geman (1984) and Grenander (1984).

1. Following the general procedure, we shall describe in §2 and §3 the *deformations* that transform the object X that we wish to reconstruct into the data Y that we can observe. The deformation is embodied in a probability distribution $\Pi(Y \mid X)$ reflecting the physics of the observed phenomenon, the characteristics of the sensor used, etc. Alone, $\Pi(Y \mid X)$ is the basis for ML reconstructions.
2. The prior information about the unknown object X is then prescribed in the form of a *Gibbs prior distribution* $\Pi(X)$ (§4). In this particular application, the prior is designed to express *spatial constraints*, such as "isotope concentrations within subregions of common tissue type and common metabolic activity are fairly homogeneous."
3. The prior distribution and the deformation mechanism let us solve, by Bayes formula, for the *posterior distribution* $\Pi(X \mid Y)$ (§5).
4. With the posterior distribution in hand, we can base reconstruction algorithms on the statistical principle of *minimum risk*. In §5 we define procedures for the MAP and MMSE reconstructions.
5. The special association of the Gibbs prior with a statistical mechanical system translates into Monte Carlo computational methods, which mimic the dynamics of the physical system. *Stochastic relaxation* (§5) is a technique for sampling from the posterior distribution $\Pi(X \mid Y)$.

In §6 we describe two methods for parameter estimation for a *natural parameter* of the family of Gibbs priors. Finally, we give examples of the reconstruction and parameter estimation methods.

This paper is intended as an introduction, with emphasis on the *statistical* perspective. A more complete discussion of physical, computational, and mathematical issues will be provided in a following paper.

2. Single Photon Emission Tomography

Emission tomography is used to determine the distribution of a pharmaceutical in a part of the body such as the brain, liver, or heart. Depending upon the pharmaceutical used, this concentration can be taken as a measure of local blood flow (perfusion) and/or local metabolic activity. Glucose, for example, is taken up by neuronal cells in proportion to metabolic activity, and the latter generally mirrors recent electrical activity. Thus, areas of the brain most used in performing a cognitive or motor task will demonstrate a relatively increased uptake of glucose immediately following the task. For the heart, pharmaceuticals can be chosen whose uptake reflects local perfusion. The concentration of these pharmaceuticals can thereby be used to assess the adequacy of blood flow to the different parts of the heart.

In SPECT, pharmaceutical concentration is estimated by detecting *photon emissions* from an injected or inhaled dose of the pharmaceutical that has been chemically combined with a radioactive isotope. This combined agent is called a radiopharmaceutical. The goal of SPECT is to determine radiopharmaceutical concentration (equivalently, isotope concentration or density) as a function of position in a region of the body. Detectors with collimators are strategically placed around the region of interest, and these are able to count photons emitted by radioactive decay of the isotope. A detector will capture those photons which escape attenuation and whose trajectories carry them down the bore of the collimator.

The determination from photon counts of isotope concentration as a function of position is referred to as *reconstruction*.

Let $X(s)$ denote the concentration of the radiopharmaceutical at the point $s = (x, y)$ in the domain Ω of interest. We shall take Ω to be a bounded two-dimensional region, though for the models and methods we will describe there are no essential changes when Ω is three-dimensional.

We assume that the detectors are arranged in a linear array, at equally spaced lateral sampling intervals, and that the detector array can be positioned at any orientation θ relative to the x-axis. (See Figure 1.) We assume the detectors are of so-called *parallel bore* type, meaning that they detect only those photons in a small interval $[\theta - \Delta\theta/2, \theta + \Delta\theta/2]$ when the array has orientation θ. Let L denote the total number of detectors in the array and let $\Delta\sigma$ denote the spacing between detectors.

The physical effects incorporated in the model are the *spatial Poisson process* that describes the sites of the radioactive decays from which photons emanate and the *process of photon attenuation* by which photons are annihilated and their energy is absorbed by matter through which their trajectories pass. Attenuation is accurately described by a linear attenuation function $\mu(s)$ on Ω. The function μ is assumed to be known; values of μ for bone, muscle, etc. and for various photon energies are known *a priori* or could be measured by transmission tomographic methods. Attenuation is a memoryless process and we can thus deduce the functional form of the probability that a photon survives to reach the detector array. When a photon trajectory has direction θ and it emanates from site $s = (x, y)$ in Ω, then

$$P(\text{photon survival}) = \exp\{-\int_{\mathcal{L}(x,y)} \mu(\xi, \eta) dl\},$$

where the line integral is taken over the segment $\mathcal{L}(x, y)$ from (x, y) to the detector and dl is differential arc length.

For our sampling design, we shall position the detector array at n equally spaced angles θ_k for duration T time units at each angle. Then at each angle, we observe the random variables

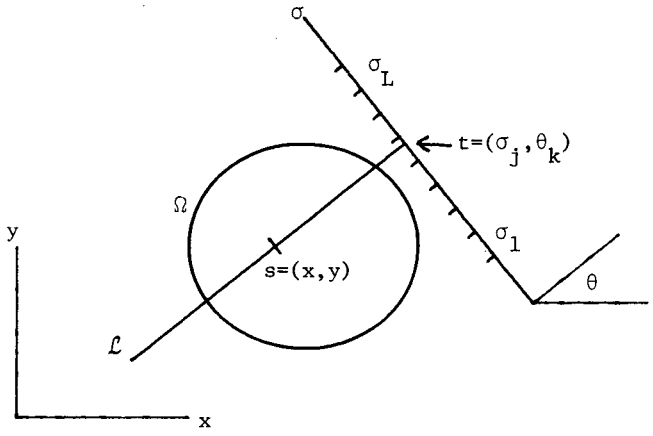

Figure 1.

$Y(t)$, for $t \in D_k = \{(\sigma_j, \theta_k), j = 1, \ldots, L\}$ that give the numbers of photons reaching the respective detectors during the sampling interval. Assuming that (i) photons are generated by a spatially nonhomogeneous Poisson process with intensity $X(s)$ per time unit, and (ii) the orientations θ of photon trajectories are uniformly distributed on $[0, 2\pi)$, we can show that $Y(t)$, for $t \in D = \cup_{k=1}^{n} D_k$, is itself a Poisson process with a nonhomogeneous intensity function described in terms of the attenuated Radon transform (ART) of X. The ART of X is defined as

$$(R_{\mu,T}X)(\sigma,\theta) = \int_{\mathcal{L}} TX(x,y) \exp(-\int_{\mathcal{L}(x,y)} \mu(\xi,\eta)dl')dl$$

where $\mathcal{L}$ is the line with orientation θ, through point σ of the detector array, $\mathcal{L}(x,y)$ is the segment of $\mathcal{L}$ starting at point (x,y) in Ω, and dl and dl' are differential arc length in the two line integrals. The intensity function of Y is then given by

$$EY(t) = \int_{\theta_k - \Delta\theta/2}^{\theta_k + \Delta\theta/2} \int_{\sigma_j - \Delta\sigma/2}^{\sigma_j + \Delta\sigma/2} (R_{\mu,T}X)(\sigma,\theta) d\sigma d\theta,$$

where $t = (\sigma_j, \theta_k)$. The important feature of this representation is that the intensity function of Y is the result of applying a positive linear integral operator $\mathcal{A}_T$ to X:

$$EY = \mathcal{A}_T X. \tag{2.1}$$

The model includes the predominant physical effects. Other potentially significant effects, such as photon scattering and background radiation, are assumed for now to be negligible. Further, we have not included effects from the sensor, such as imperfect collimation, blurring, and noise. We note, however, that the reconstruction methods described below, since they are based on the generally applicable principles of maximum likelihood and Bayes optimality, are adaptable to models incorporating additional physical and sensor effects. Mertus (1987) has made extensions for scattering and collimation errors.

3. Maximum Likelihood and EM

A variety of reconstruction algorithms for emission tomography are described by Budinger et. al. (1979). The algorithms that are traditionally used are based on ideas of extracting a signal in the presence of noise and related methods of linear filtering.

More recently, interest has been heightened in the use of algorithms that use fuller information of the mathematical model sketched above, along with the ML principle. Shepp and Vardi (1982) laid the mathematical foundations and developed effective algorithms based on EM (Dempster, Laird and Rubin (1977)) for implementing ML reconstructions in positron emission tomography (PET). A penetrating description, written from a statistician's perspective, is given in Vardi, Shepp and Kaufman (1985). (In PET, photon attenuation does not enter the model relating isotope concentration to the observables.) McClure and Accomando (1984) have developed the foundations for applying ML to SPECT reconstructions and have implemented EM algorithms on a variety of computer systems. Independently, Miller, Snyder and Miller (1985) have made similar extensions of ML and EM for SPECT.

By exploiting properties of the Poisson process, it can be shown that the observables $Y(t)$ are mutually independent and Poisson distributed; the likelihood function is then easily obtained from (2.1). To carry out a ML reconstruction, we first discretize the domain Ω into *pixels* parameterized by discrete points s in a square lattice S. Now $\{X(s)\}_{s\in S}$ represents a piecewise constant approximation of the isotope concentration on the continuous domain. When Ω is discretized, then equation (2.1) takes the form

$$EY = A_T X,$$

where A_T is a matrix, $A_T = \{A(t,s)\}_{t\in D, s\in S}$; commonly, the order of A_T is extremely large and it may not have full column rank. Now for a given X, the Poisson probability function of Y is

$$\Pi(Y \mid X) = \prod_{t\in D} \frac{[(A_T X)(t)]^{Y(t)}}{Y(t)!} \exp\{-(A_T X)(t)\} \tag{3.1}$$

where our notation is making convenient abuse of the distinction between a random variable and its value.

The log-likelihood function is

$$\ln L(x) = \sum_{t\in D} \{-\ln(Y(t)!) + Y(t)\ln[(A_T X)(t)] - (A_T X)(t)\}. \tag{3.2}$$

The necessary conditions for maximizing $\ln L(X)$ obtained by setting derivatives to zero do not yield explicit solutions for a maximizing X. Nonetheless, $-\ln L(X)$ is globally convex, and the ML optimization problem conveniently adapts to the EM method. In general, $-\ln L(X)$ is not *strictly* convex; this is an identifiability issue related to the column rank of A_T. Conditions for *strict* convexity are discussed by Accomando (1984).

The EM algorithm becomes an explicit iterative reconstruction procedure. We initialize the iteration with $\{X^{(0)}(s)\}_{s\in S}$ and update $X^{(r)}$ by the formula

$$X^{(r+1)} = \{[A_T'(Y \oslash A_T X^{(r)})] \oslash A_T' \mathbf{1}\} \otimes X^{(r)}, \tag{3.3}$$

where $\mathbf{1}$ is the vector whose components are identically one, $\oslash$ denotes component-by-component division, and $\otimes$ denotes component-by-component multiplication. At each step, the iteration requires two (large) matrix multiplications. The sequence of iterates converges to an X^* that maximizes $\ln L(X)$. Consistency results that depend on the sampling design and on the discretization of Ω can be proved.

Figure 2C in §5 shows an example of a ML reconstruction for a simulation experiment. The true isotope density used for the simulated data is depicted in Panel A of Figure 2. The noisy appearance of the ML reconstruction is not atypical, even though the sample size is rather large in this experiment for estimating the 32×32 discrete image. The high degree of local irregularity occurs because ML builds in *no* spatial information, e.g. about relative locations of pixels in the grid. Snyder and Miller (1985), recognizing the inherent nonparametric nature of the reconstruction problem, have suggested using Grenander's method of sieves (Grenander (1981)) to regularize the ML estimates. Accomando (1984) also uses sieves to study consistency questions.

4. Gibbs Prior Distribution

We suggest a Bayesian formulation for incorporating prior spatial constraints into the reconstructions. We shall construct a prior distribution on X that captures simple prior expectations about the qualitative nature of the isotope density. Mainly, we wish to exploit the anticipated smoothness of X. Neighboring locations will typically have similar intensity levels. But we must also accommodate sharp changes in concentration, which might occur across an arterial wall or across a boundary between two tissue types.

In the spirit of nonparametric estimation, we might construct the prior on a suitable space of functions $X : \Omega \to \mathrm{R}$. It is more convenient, however, to do the construction on the discrete domain S introduced in §3. The prior, therefore, is on the array $X = \{X(s)\}_{s \in S}$. The range of values of $X(s)$ will be confined to a compact interval, usually $[0, 255]$, and might be further restricted to only the integer values in the interval. As a further convenience, we will restrict ourselves to priors with *Gibbs representation*

$$\Pi(X) = \frac{1}{Z} \exp\{-U(X)\} \tag{4.1}$$

where Z is the normalizing constant, $Z = \int \exp\{-U(X)\} dX$, and $U : \mathrm{R}^S \to \mathrm{R}$ is known as the "energy". As it stands, the Gibbs representation is only mildly restrictive since U is arbitrary. However, we shall restrict U to involve only "nearest neighbor" interactions among the components of X.

We employ the Gibbs representation because it is easier to design an energy function with desired properties (such as localization of interactions, Markovian restrictions on conditional distributions, ...) than it is to construct a distribution Π directly. We will design U so that the expected configurations have *low energy* as they do in a real physical system. The expected configurations are those for which typical neighboring sites $s, t \in S$ have similar intensities $X(s), X(t)$. This is a *local* constraint and it is conveniently captured by a locally composed energy function U,

$$U(X) = \sum_{[s,t]} \beta \phi(X(s) - X(t)) + \sum_{<s,t>} \frac{\beta}{\sqrt{2}} \phi(X(s) - X(t)). \tag{4.2}$$

Here we use $[s, t]$ to indicate that s and t are nearest horizontal or vertical neighbors in the lattice S, and $< s, t >$ to denote diagonal neighbors. The constant β is positive and the function $\phi(\xi)$ is even and minimized at $\xi = 0$. Thus U is minimized by configurations of constant intensity. Under the Gibbs distribution (4.1) the more likely isotope densities are those with small site-to-site variation in intensity.

This definition of ϕ and U induces a graph on S in which each pixel site s is linked to its eight nearest neighbors in the square lattice. The distribution Π then determines a Markov random field with this neighborhood structure.

To achieve the desired properties for the more likely isotope densities, the exact form of ϕ is probably not important, but its qualitative features can make a difference. We have experimented with ϕ's that are increasing in ξ for $\xi \geq 0$. An obvious choice is $\phi(\xi) = \xi^2$, but then under $\Pi(X)$,

large intensity gradients, as would be associated with certain natural boundaries, are exceedingly unlikely. Instead, we use functions of the form

$$\phi(\xi) = \frac{-1}{1+(\xi/\delta)^2} \tag{4.3}$$

where δ, like β, is a constant to be fixed later.

There are two free parameters in the specification of U: δ is easily interpreted as a scale parameter on the range of values of $X(s)$ and β controls the "strength" of the interactions between a pixel and its neighbors. It is a *natural parameter* of the exponential family (4.1), and admits meaningful statistical and physical interpretations. From the physical viewpoint, β is the reciprocal of temperature for the statistical mechanical system defined by (4.1). From the statistical viewpoint, it will be seen as a "smoothing parameter" controling the tradeoff for our reconstructions between the influence of the observables and the influence of the prior constraints.

Levitan and Herman(1987) have recently proposed the use of Gaussian priors in a Bayesian formulation. Liang and Hart (1987) also suggest the use of Gaussian priors, as well as others, deduced by max-ent arguments from prior constraints on low-order moments of X. Our earlier experiments with the *quadratic* energy function indicated that the resulting Bayesian algorithms oversmoothed real boundaries where the difference $(X(s) - X(t))$ should be allowed to be large. The finite asymptotic behavior of our ϕ-function was designed to mitigate this oversmoothing.

5. Posterior Distribution and Bayes Optimal Reconstructions

From (3.1) and (4.1) the *posterior distribution* on X is

$$\Pi(X \mid Y) = \frac{1}{Z(Y)} \exp\{-U(X) + \sum_{t \in D} [Y(t) \ln[(A_T X)(t)] - (A_T X)(t)]\} \tag{5.1}$$

where $Z(Y)$ is a normalizing constant that depends on Y.

We have developed algorithms for two Bayes opptimal reconstructions of X—the minimum-mean-squared-error (MMSE) estimator

$$X^* = E(X \mid Y) \tag{5.2}$$

and the maximum-a-posteriori (MAP) estimator, which maximizes the value of $\Pi(X \mid Y)$ or equivalently minimizes the posterior energy

$$U(X) - \sum_{t \in D} [Y(t) \ln[(A_T X)(t)] - (A_T X)(t)]. \tag{5.3}$$

The algorithm for each of these reconstructions is built around a technique for simulating computationally the dynamics of a statistical mechanical system with energy given by (5.3). Details of the generic algorithm, a variant of the Metropolis algorithm (Metropolis et. al. (1953)) known as *stochastic relaxation* (SR), are given in Geman and Geman (1984); the idea is sketched below.

Notice in (5.3) that we have the usual equivalence between Bayesian MAP estimation and so-called penalized ML. ML maximizes

$$\sum_{t \in D} [Y(t) \ln[(A_T X)(t)] - (A_T X)(t)],$$

wheras MAP estimation includes the "penalty term" $-U(X)$, which penalizes lack of smoothness. One advantage, we believe, of the Bayesian viewpoint is that it suggests mechanisms for *estimating*

the required degree of smoothness, which amounts to estimating the pivotal parameter β in the Gibbs prior. We focus on this estimation problem in the next section.

MMSE Algorithm. The computational method is iterative. We initialize $X = X^{(0)}$. In practice, we choose a "good" initialization such as the EM reconstruction, but easy theory says that convergence is independent of the initialization. We visit each site s in the pixel array, successively in any order, and replace $X(s)$ by a value sampled from the conditional distribution on $X(s)$, under (5.1) and conditioning on all $X(t), t \neq s$; this is the essence of stochastic relaxation (SR). The iterates $X^{(\tau)}$ form a Markov chain with equilibrium distribution (5.1). The ergodicity of the chain guarantees that an ergodic average of $\{X^{(\tau)}\}_{\tau=0}^{\infty}$ will converge to X^* a.s. In practice, we compute N iterates and average the final M, with choices such as $N = 25$ and $M = 5$. The selection of suitable M and N can be guided by monitoring stabilization of statistics of the successive iterates $X^{(\tau)}$.

MAP Algorithm. Computing the minimum of (5.3) is, in general, a hard problem. The method of simulated annealing can be implemented to yield a sequence $\{X^{(\tau)}\}$ converging in distribution to a MAP estimator X^*. The procedure is similar to SR. The fundamental ideas are described in Pincus (1970), Cerný (1982), and Kirkpatrick, Gellatt and Vecchi (1983). See also Geman and Geman (1984) for applications to image processing.

For the design of *feasible* algorithms, we are guided by pragmatism as well as by the theoretical underpinnings of SR and simulated annealing. First we compute the ML reconstruction by EM. Then—in the language of simulated annealing—we "run" the physical system with posterior energy (5.3) at *zero temperature*. When our state-space (the range of values for $X(s)$) is discrete, this amounts to using Besag's method of Iterated Conditional Modes (ICM), Besag (1986). When the state-space is a continuous interval and the temporal index is also continuous ($\tau \in [0,\infty)$), we implement this step by performing gradient descent on (5.3) starting at the EM reconstruction. The local minimum of (5.3) obtained by ICM or by gradient descent is our approximate MAP estimate of X.

Note that ICM and gradient descent do not guarantee convergence to a global minimum of (5.3). The rationale for making a judicious choice for the initialization is to capture a "good" local minimum for the approximate MAP reconstruction.

Figure 2, Panels D, E, and F, shows approximate MAP reconstructions of the known *phantom* depicted in Figure 2A. First the ART of the phantom X was computed, for $n = 60$ sampling angles and $L = 64$ lateral sampling steps. The nonuniform attenuation function μ depicted in 2B was used to compute the ART; it builds a very substantial attenuation effect into the model. The Poisson data Y was generated to satisfy (3.1). Figure 2C shows the approximate ML reconstruction after 54 iterations of EM. For the Bayesian reconstructions, we used the prior of (4.1)–(4.3), with $\delta = 0.7$, and with the range of $X(s)$ in $[0,15]$ Each of the Bayesian reconstructions is computed by gradient descent starting from the EM estimate (EM-GrD). The effect of different choices for β on the degree of smoothing is apparent. We shall discuss estimation of β in the next section.

6. Parameter Estimation

The choice of β is critical. With $\beta = 0$ the estimator is undersmoothed, and in fact MAP estimation is just ML, since the prior is uniform. If β is too large, the estimator is too faithful to the prior and is oversmoothed. The parameter δ is also important, though we have found that (i) its value can usually be set based on information about the range of values $\{X(s)\}$, and (ii) reconstructions are not sensitive to moderate changes in δ. The discussion here will focus on β.

Because of the setting in which reconstruction algorithms are actually used, it is desirable to design estimation methods that work with a sample $\tilde{Y}$ of size one from the observable process. The isotope density $\tilde{X}$ is assumed to be drawn from a Gibbs prior with unknown β, but known δ (4.3). We shall estimate β from $\tilde{Y}$ and use the estimate $\hat{\beta}$ in the MMSE or MAP reconstruction

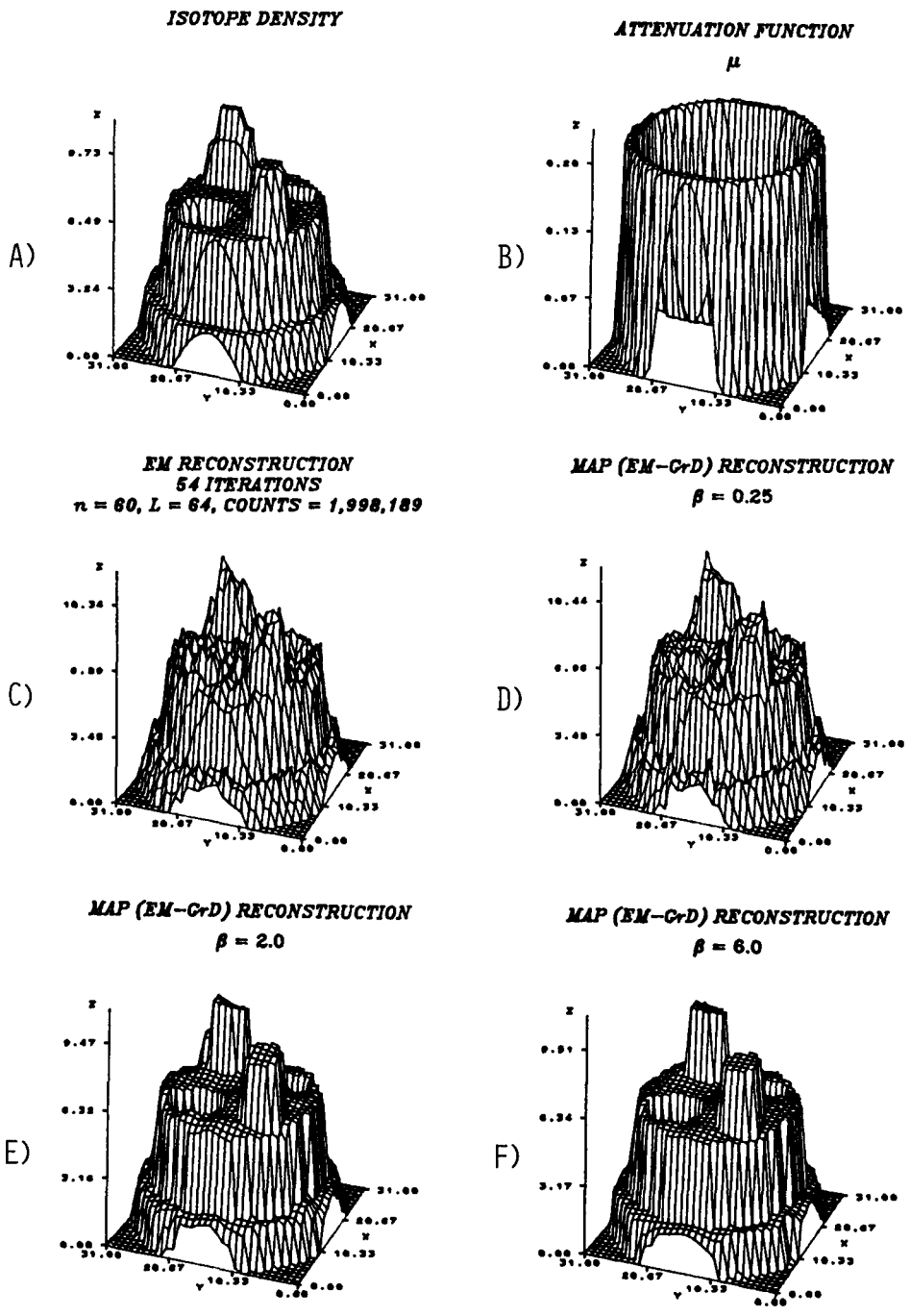
ISOTOPE DENSITY
ATTENUATION FUNCTION
μ
A)
B)
EM RECONSTRUCTION
54 ITERATIONS
n = 60, L = 64, COUNTS = 1,998,189
MAP (EM–GrD) RECONSTRUCTION
β = 0.25
C)
D)
MAP (EM–GrD) RECONSTRUCTION
β = 2.0
MAP (EM–GrD) RECONSTRUCTION
β = 6.0
E)
F)

Figure 2.

program. It is reasonable to do this with a single observation $\tilde{Y}$, since $\tilde{Y}$ contains a large amount of data about $\tilde{X}$, which, in turn, contains a large amount of data about the local energy function $U(X)$.

To be more explicit about the dependency on β of the prior and posterior distributions, we introduce the function

$$V(X) = \sum_{[s,t]} \phi(X(s) - X(t)) + \frac{1}{\sqrt{2}} \sum_{<s,t>} \phi(X(s) - X(t)).$$

V is just U/β. The prior is now written

$$\Pi(X) = \frac{1}{Z_\beta} \exp\{-\beta V(X)\}$$

and the posterior, given $\tilde{Y}$, is

$$\Pi(X \mid \tilde{Y}) = \frac{1}{Z_\beta(\tilde{Y})} \exp\{-\beta V(X) + \sum_{t \in D} [\tilde{Y}(t) \ln[(A_T X)(t)] - (A_T X)(t)]\}$$

Now $V(X)$ is a *complete-data sufficient statistic* for β. If we were able to observe $\tilde{X}$ directly, then we could, in principle, solve the likelihood equation

$$E_\beta[V(X)] = V(\tilde{X}) \tag{6.1}$$

for the ML estimate of β. The left-hand side of (6.1) is strictly decreasing in β and thus (6.1) yields a unique root $\hat{\beta}$.

Our situation is more complicated than this since we do not observe $\tilde{X}$, but instead we see only the *incomplete data* $\tilde{Y}$. We have a classic setup for application of EM. The EM algorithm, when it converges, will yield a root of the incomplete-data likelihood equation

$$E_\beta[V(X)] = E_\beta(V(X) \mid \tilde{Y}); \tag{6.2}$$

see Dempster, Laird and Rubin (1977). We note that there is no proof of uniqueness of roots of (6.2). Conceptually, (6.2) is solved at the intersection of two monotone decreasing functions of β. Whether (6.2) does admit multiple solutions is an open and elusive theoretical question.

To solve (6.2), the EM algorithm consists of two alternating steps—estimation of the right-hand side of (6.2) for prescribed β (E-step) and computation of the root $\tilde{\beta}$ of (6.2), substituting the current estimate of $E_\beta(V(X) \mid \tilde{Y})$ on the right-hand side. Specifically, we fix an initial $\beta = \beta^0$ and an initial $X = X^0$ (and hence V^0). Then solve

E-step. Estimate the complete-data sufficient statistic:

$$V^{(\tau+1)} = E_{\beta^{(\tau)}}(V(X) \mid \tilde{Y}) \tag{6.3a}$$

M-step. Determine $\beta^{(\tau+1)}$ as the solution of

$$E_\beta[V(X)] = V^{(\tau+1)}. \tag{6.3b}$$

The first step is done using SR, using say ten steps of SR and averaging the last five values of $V(X^{(\nu)})$. The second step is a simple root-finding calculation once the curve $E_\beta[V(X)]$ is known. Conveniently, the SR procedure simultaneously yields updates $X^{(\tau)}$ of the MMSE reconstruction. Thus (6.3a) and (6.3b) together give a completely data-driven method of reconstruction.

The construction of $E_\beta[V(X)]$ as a function of β can be done "off line", once and for all. We have done this using SR to simulate 230 configurations X from the prior (4.1) for β-values ranging from 0 to 6. Five replications were done at each of forty-six values of β. The resulting curve, fit by a cubic-spline regression function, is depicted in Figure 3. the calculation of this curve required forty-one hours of CPU time, using a highly optimized program on the 100 Megaflop Star Technologies ST100 Array Processor.

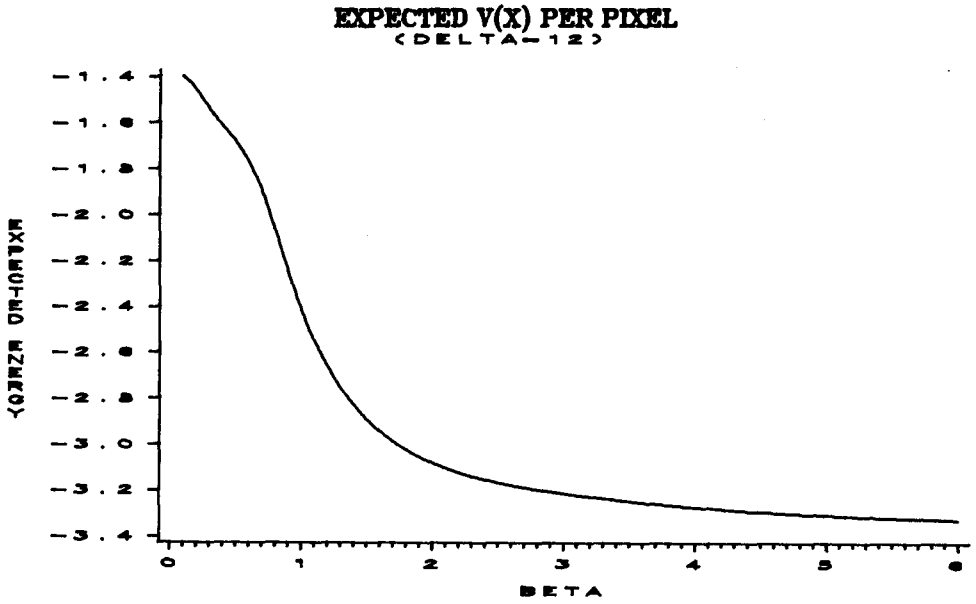

Figure 3.

J. Mertus (1987) has developed an efficient vectorized FORTRAN program for the EM estimation/reconstruction procedure described above. Each E-step, with ten sweeps of SR, takes on the order of seven minutes of CPU time on an IBM3090 or about three minutes on a CYBER 205, working on a 64×64 pixel lattice S, for isotope densities X having their support on a disk of diameter 44 pixels (about 22cm) and with a range of 64 grey levels. (These values correspond to our real data sets.) The computational requirements are enormous, but not prohibitive.

To circumvent the computational demands of EM, we have devised and experimented with a *moment method* for estimating β. The goal is to have a direct estimation method for β that can be applied to the observable $\tilde{Y}$ without requiring intermediate reconstruction of X. We construct a statistic $M(Y)$ based on the notion that the smoothness of Y will reflect the magnitude of β in the same way that the smoothness of X does. The exact form of $M(Y)$ is also guided by our knowledge of the Poisson distribution of Y and ability to compute theoretical moments of the Poisson random variables.

For the detector bin at angle θ_k and at sampling step σ_j, denote $t = (\sigma_j, \theta_k)$ and $t^+ = (\sigma_{j+1}, \theta_k)$. Also, introduce the notation $a(t) = (A_T \mathbf{1})(t)$, where $\mathbf{1}$ is the vector with components identically equal to one; $a(t)$ is simply the row-sum of A_T associated with the detector at location t. Then define the moment statistic

$$M(Y) = \sum_{k=1}^{n} \sum_{j=22}^{42} \left[\frac{Y(t)}{a(t)} - \frac{Y(t^+)}{a(t^+)} \right]^2 - \frac{Y(t)}{a^2(t)} - \frac{Y(t^+)}{a^2(t^+)} \tag{6.4}$$

The inner sum restricts the moment to the central part of the support of the isotope density to avoid edge effects. The expectation of $M(Y)$, for given X, is a measure of roughness of normalized

ART projections of X:

$$E(M(Y) \mid X) = \sum\sum T^2 \left[\frac{A_T X(t)}{a(t)} - \frac{A_T X(t^+)}{a(t^+)} \right]^2 . \tag{6.5}$$

We anticipate that the expectation $E_\beta[M(Y)]$ with respect to the prior will have the same general behavior as $E_\beta[V(X)]$ in (6.1). Accordingly, we define the moment estimate β^* of β as the root of the equation

$$E_\beta[M(Y)] = M(\tilde{Y}). \tag{6.6}$$

The effort to compute β^* is trivial, once the left-hand side of (6.6) is known as a function of β.

We have constructed the curve describing $E_\beta[M(Y)]$ using the same simulated X-data that generated $E_\beta[V(X)]$ in Figure 3. Figure 4 shows the resulting curve; it does, indeed, exhibit the same qualitative behavior as the curve in Figure 3.

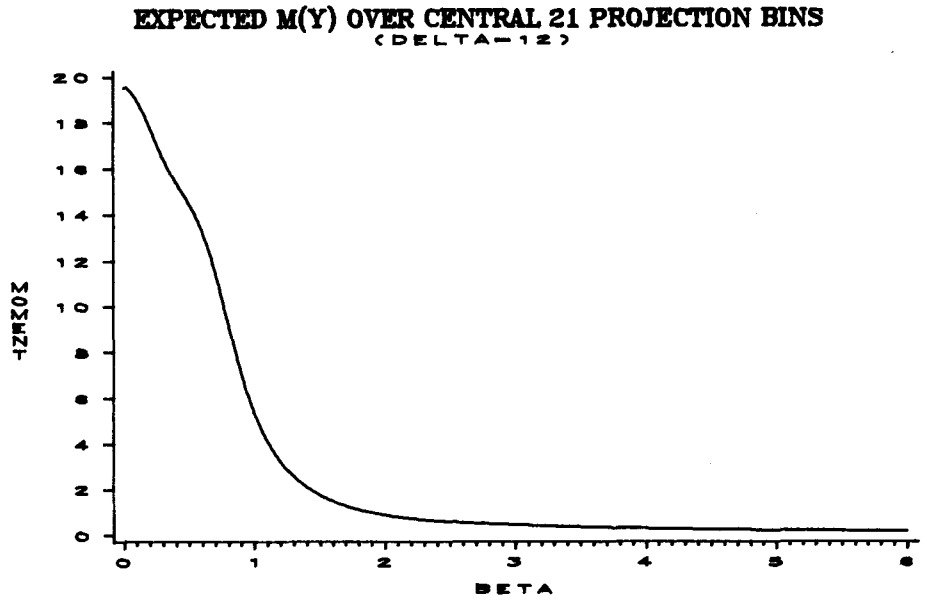

Figure 4.

A variety of experiments have been done with both the EM and moment method of estimating β. The most ideal circumstance, of course, is when the model truly fits the data.

In one such experiment, an X-array was generated from the prior (4.1) with $\beta = 1$. (As above, we used a 64×64 pixel lattice, 64 grey levels, a disk of diameter 44 pixels for the support of X, and a uniform attenuation function for the construction of A_T.) In implementing the E-step, ten passes of SR were performed and the last five values of $V(X^{(r)})$ were averaged to estimate the right-hand side of (6.3a). When $\beta^0 = 0.0$, the successive iterates of $\beta^{(r)}$ from the M-step were 0.63, 0.85, 0.95, 1.01, and 1.03. When $\beta^0 = 6.0$, the successive were 2.26, 1.15, 1.08, 1.05, 1.05, and 1.04. For the same X-array, five independent replications of the observable Y process were generated and the moment method yielded estimates β^* of 0.97, 1.00, 1.02, 1.06, and 0.98; the five estimates have mean 1.005 and standard deviation 0.034.

For more thorough testing of the moment method, a test set of X-arrays—independent of the set used to construct the curves in Figures 3 and 4—was generated with β-values ranging from 0.5 to 2.5. For each β, five X-arrays were generated, and for each X-array, five independent replications of the Y process were simulated. Figure 5 depicts the estimate errors $\beta^* - \beta$ for each of the twenty-five experiments at each β-value. The dispersion of the errors as a function of β is what one would anticipate from the slope of $E_\beta[M(Y)]$.

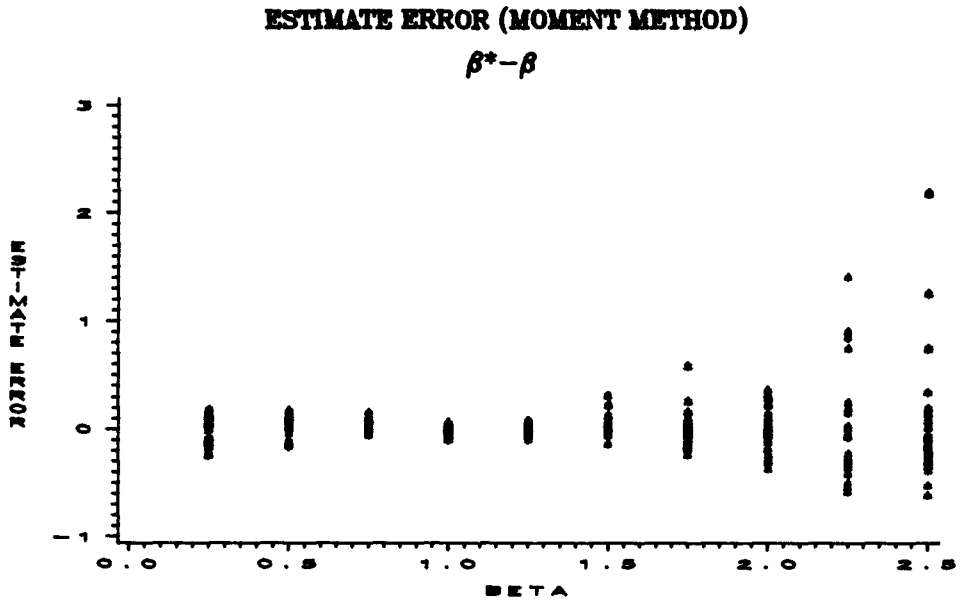

Figure 5.

7. Reconstruction Experiments

We report on two experiments which have been run on real and simulated data to learn about the performance of the Bayesian reconstruction methods in cases for which the underlying model does not fit exactly. One simulation experiment was designed to test the versatility and robustness of the methods to known departures from the model. The other experiment illustrates the performance of the algorithms on real data from a lung section.

The pseudo-grey-level images in Figures 6 and 7 associate high values in $[0, 63]$ with black and low values with white. *Our ability to present pictorial examples is limited by the printing process for this volume. Interested readers can obtain higher resolution copies of photographs on request to D.E. McClure.*

Experiment 1. A phantom isotope density (Figure 6A) was designed to have a combination of (i) large-scale structure, including subregions of Ω with considerable differences in intensity, and (ii) local irregularity of the same qualitative nature as that of sample functions from the Gibbs model (4.1)–(4.3), yet not precisely fitting the Gibbs model. Two functions were averaged to form the phantom. First, an array with a sharp spike in intensity (near the center, below the middle) was constructed. Second, an array was sampled from (4.1)–(4.3) with parameter values $\beta = 1$, and $\delta = 12$. Intuitively, the local structure of the average will be governed by the array sampled

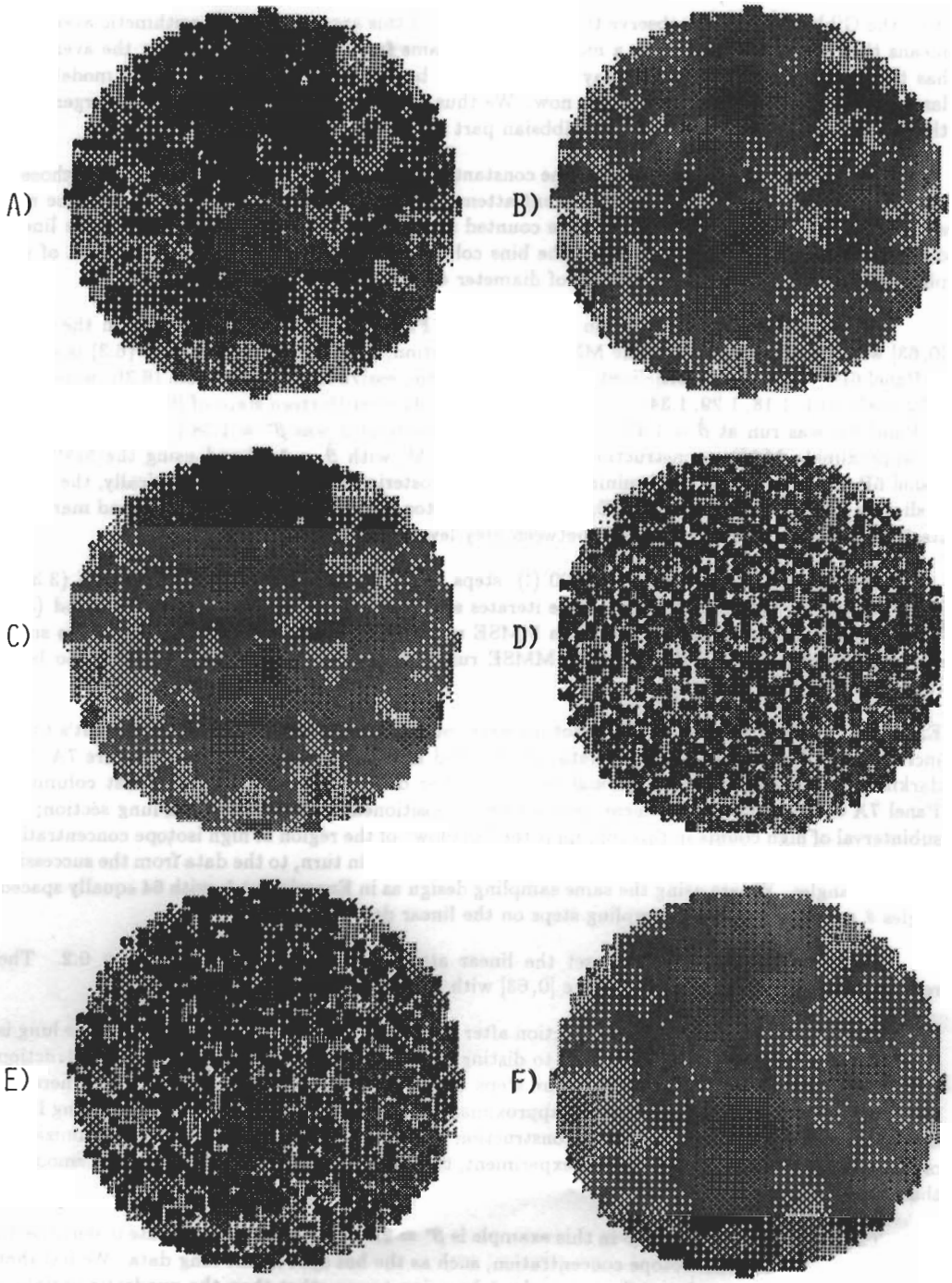

Figure 6. Simulated Data, 663,144 Total Counts.

from the Gibbs model. But observe that the rescaling of this array due to the arithmetic averaging means that it will not exactly fit a model from the same family. Roughly speaking, the averaging has the effect of smoothing the array so that it will be better described by a Gibbs model with larger β-value, assuming δ is fixed for now. We thus anticipate estimated values of β larger than the value $\beta = 1$ used to generate the Gibbsian part of the averaged phantom.

To simulate the emitted photons, the constant linear attenuation function $\mu \equiv 0.2$ was chosen, corresponding to approximately ten percent attenuation per centimeter for our scaling of the real system. A total of 663,144 photons were counted at 64 angles θ, with $L = 64$ bins on the linear detector array; in actuality, only 44 of the bins collect positive counts because the support of the phantom is contained in a smaller disk of diameter 44 pixels.

Reconstructions are depicted in Panels B–F of Figure 6. All were constructed on the range $[0, 63]$ with parameter $\delta = 12$. The MMSE reconstruction, with β estimated by EM (6.3) is shown in Panel 6B. When β was initialized at $\beta^0 = 0.0$, the successive EM iterates from (6.3b) were 0.52, 0.72, 0.85, 1.01, 1.18, 1.29, 1.34, 1.38, 1.40, 1.41, ..., 1.47 after thirteen steps of (6.3b). The MMSE in Panel 6B was run at $\hat{\beta} = 1.47$. (The moment estimate of β was $\beta^* = 1.38$.) Panel 6C depicts an approximate MAP reconstruction obtained by ICM, with $\hat{\beta} = 1.47$ and using the MMSE in Panel 6B to initialize the local minimization of the posterior energy. Characteristically, the MAP is slightly smoother than the MMSE; on a video monitor the difference is perceptible and manifests itself in apparent coarser transitions between grey levels in the MAP image.

The EM reconstruction after 5000 (!) steps of (3.3) is shown in Panel 6D. When (3.3) is run with double precision, the successive iterates still continue to increase the log-likelihood (3.2) after 5000 iterations. Panel 6E shows an MMSE run with a value of $\beta = 0.52$, which is too small (undersmoothing). Panel 6F shows an MMSE run with a value of $\beta = 4.40$, which is too large (oversmoothing).

Experiment 2. A total of 124,136 photons were counted from a cross-section of a patient's torso, including the lungs. The observed data are depicted in the so-called sinogram in Figure 7A. The darkness in the figure is proportional to the number of detected photons. The first column of Panel 7A corresponds to the linear detector being positioned to the right of the lung section; the subinterval of high counts in this column is the "shadow" of the region of high isotope concentration in the lung. The successive columns in Panel 7A correspond, in turn, to the data from the successive sampling angles. We are using the same sampling design as in Experiment 1, with 64 equally spaced angles θ and $L = 64$ lateral sampling steps on the linear detector array.

For the reconstructions,, we set the linear attenuation function again at $\mu \equiv 0.2$. The reconstructions were done on the range $[0, 63]$ with fixed $\delta = 12$.

Panel 7B shows the EM reconstruction after 5000 steps of (3.3). The "hot spot" in the lung is apparent, but local structure is difficult to distinguish. Panel 7C shows the MMSE reconstruction with β estimated at $\hat{\beta} = 4.56$ after four steps of the EM estimation procedure (6.3); here we initialized $\beta^0 = 6.0$. Panel 7D shows an approximate MAP reconstruction formed by applying ICM, setting $\hat{\beta} = 4.56$, and using the EM reconstruction in Panel 7B to initialize the local minimization of the posterior energy. Again in this experiment, the MAP reconstruction is somewhat smoother than the MMSE.

The moment estimate for β in this example is $\beta^* = 2.71$. The moment estimate is sensitive to sharp singularities in the isotope concentration, such as the hot spot in the lung data. We feel that the moment method can be made more robust by using terms other than the quadratic variation used in (6.4) for the summands that define the moment statistic. There are analytical obstacles, however, to calculating a bias correction for alternative summands, so that the expectation of the moment statistic, given X, is a function of differences alone, as (6.5) is.

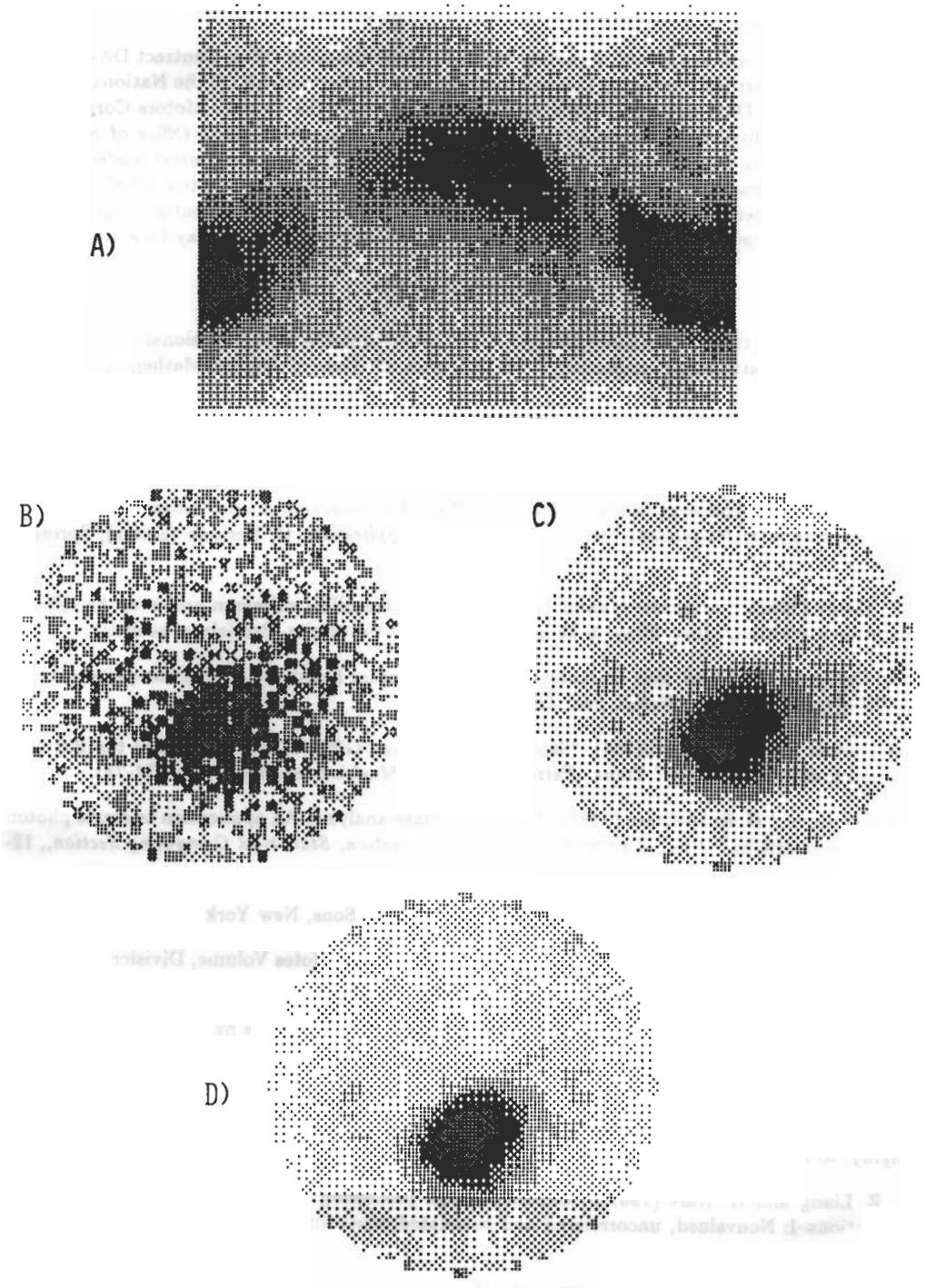

Figure 7. Real Lung Data, 124,136 Total Counts.

Acknowledgements

This research was supported in part by the U.S. Army Research Office Contract DAAL03-86-K-0171 to the Center for Intelligent Control Systems (Brown-Harvard-MIT), the National Science Foundation, Grant DMS-8352087, North American Philips, and the General Motors Corporation. We have also benefited from the support of the National Science Foundation Office of Advanced Scientific Computing Supercomputer Centers Program through computing resources made available at the John von Neumann Center and at the University of Illinois. John Mertus did a significant part of the experiments reported in §6 and §7. George Loriot has provided essential support in the development of algorithms and programs for the Star Technologies ST100 Array Processor.

BIBLIOGRAPHY

1. N. Accomando (1984), "Maximum likelihood reconstruction of a two-dimensional Poisson intensity function from attenuated projections," Ph.D. Thesis, Division of Applied Mathematics, Brown University.

2. J. Besag (1986), "On the statistical analysis of dirty pictures," *J. R. Statist. Soc.*, **B**, Vol. 48, 259-279.

3. T. Budinger, G. Gullberg, and R. Huesman (1979), "Emission computed tomography," in *Image Reconstruction from Projections: Implementation and Application,* G. Herman (editor), Springer-Verlag.

4. V. Cerný (1982), "A thermodynamic approach to the travelling salesman problem: An efficient simulation algorithm," Inst. Phys. & Biophys., Comenius University, Bratislava, preprint.

5. A. Dempster, N. Laird, and D. Rubin (1977), "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, **B**, Vol. 39, 1-38.

6. S. Geman and D. Geman (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. and Machine Intelligence*, **6**, 721-741.

7. S. Geman and D. E. McClure (1985), "Bayesian image analysis: An application to single photon emission tomography," *Proc. American Statistical Association, Statistical Computing Section,*, 12-18.

8. U. Grenander (1981), *Abstract Inference*, John Wiley & Sons, New York.

9. U. Grenander (1984), "Tutorial in Pattern Theory," Lecture Notes Volume, Division of Applied Mathematics, Brown University.

10. S. Kirkpatrick, C.D. Gellatt, and M.P. Vecchi (1983), "Optimization by simulated annealing," *Science*, **220**, 671-680.

11. E. Levitan and G.T. Herman (1987), "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. on Medical Imaging*, to appear.

12. Z. Liang and H. Hart (1987), "Bayesian image processing of data from constrained source distributions–I: Nonvalued, uncorrelated and correlated constraints," *Bull. Math. Biol.*, Vol. 49, 51-74.

13. Z. Liang and H. Hart (1987), "Bayesian image processing of data from constrained source distributions–II: Valued, uncorrelated and correlated constraints," *Bull. Math. Biol.*, Vol. 49, 75-91.

14. J. Mertus (1987), "Self-calibrating Bayesian methods for image reconstruction in emission tomography," Ph.D. Thesis, Division of Applied Mathematics, Brown University, in preparation.

15. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller (1953), "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, Vol 21, 1087-1091.

16. M.I. Miller, D.L. Snyder, and T.R. Miller (1985), "Maximum-likelihood reconstruction for single-photon emission computed tomography," *IEEE Trans. on Nuclear Science*, **NS-23**, 769-778.

17. M. Pincus (1970), "A Monte Carlo method for the approximate solution of certain types of constrained optimization problems," *Operations Research*, **18**, 1225-1228.

18. L.A. Shepp and Y. Vardi (1982), "Maximum likelihood reconstruction in positron emission tomography," *IEEE Trans. on Medical Imaging*, **1**, 113-122.

19. D.L. Snyder and M.I. Miller (1985), "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. on Nuclear Science*, **NS-32**, 3864-3872.

20. Y. Vardi, L.A. Shepp, and L. Kaufman (1985), "A statistical model for positron emission tomography," *JASA*, Vol. 80, 8-20 and 34-37.

SUMMARY

The reconstruction problem for SPECT (single photon emission computed tomography) is formulated as a statistical estimation problem: *estimate the nonhomogeneous intensity function of a two- (or three-) dimensional Poisson process from indirect observations.* Previously, this has been addressed using the principle of maximum likelihood, but the likelihood method does not incorporate spatial constraints. Alternatively, spatial information about the unknown intensity function can be described by a Gibbs prior distribution and this then leads to Bayesian methods for the reconstruction (estimation) problem. Bayesian reconstructions are described and illustrated by examples using both real and simulated data. A parameter estimation problem for the Gibbs prior distributions is posed. Two methods are suggested and illustrated for the subsidiary parameter estimation problem. Computational algorithms are given.

RÉSUMÉ

Nous considérons le problème de reconstruction de SPECT (single photon emission computed tomography) comme étant un problème d'estimation; c'est à dire que nous estimons la fonction d'intensité (nonhomogène) d'un processus Poissonien à 2 (ou 3) dimensions. Jusqu'à maintenant, ce problème a été traité en utilisant le principe du maximum de vraisemblance; mais cette méthode ne tient pas compte des contraintes spatiales. D'autre part, l'information spatiale sur la fonction d'intensité inconnue peut être traduite par l'emploi d'une distribution de Gibbs a priori, et nous sommes conduit à une méthode Bayesienne pour le problème de reconstruction. Nous décrivons des reconstructions Bayesiennes et donnons des exemples utilisant à la fois des données réelles et simulées. Nous posons des questions sur l'estimation des paramètres de la distribution a priori de Gibbs, et nous suggérons et donnons des exemples d'application de deux méthodes pour ce problème subsidiaire de l'estimation de paramètres. Nous donnons aussi les algorithmes utilisés.