PET Image Reconstruction: Methodology and Quantitative Accuracy

11

Bing Bai and Evren Asma

Contents

B. Bai, PhD (✉)
Toshiba America Medical Systems, Inc.,
2441 Michelle Drive, Tustin, CA 92780, USA
e-mail: bing.bai@toshiba.com

E. Asma, PhD
Toshiba Medical Research Institute USA, Inc.,
706 N Deerpath Drive, Vernon Hills, IL 60061, USA
e-mail: easma@tmriusa.com

Abstract

This chapter reviews the techniques developed for positron emission tomography (PET) image reconstruction and image property analysis. Both mathematical theory and practical considerations are introduced. We focus on the commonly used methods on commercial PET scanners, in particular model-based statistical reconstruction methods. We also briefly describe data corrections necessary for PET image reconstruction, which are important for reducing artifacts and improving quantitative accuracy. Finally some recent developments are described, including the reconstruction of time-of-flight (TOF) PET data and direct parametric image reconstruction.

11.1 Introduction

The theory of image reconstruction from projections has been developed since the early twentieth century [1] and flourished after the introduction of x-ray computed tomography (CT) [2]. Many methods have been proposed and implemented successfully for various imaging techniques including CT, positron emission tomography (PET), single-photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI).

M.M. Khalil (ed.), *Basic Science of PET Imaging*, DOI 10.1007/978-3-319-40070-9_11

Image reconstruction techniques can be broadly grouped into two categories: analytic and statistical methods. Analytic image reconstruction methods model the PET data as line integrals through the image. An exact or approximate formula is sought to invert the x-ray transform [3]. Despite the fact that the line integral model is only a very rough approximation to the underlying physics in PET and the fact that the statistical nature of measured data is ignored, analytic image reconstruction methods have been used extensively in clinical practice due to their speed. In the last 10 years, the rapid increase of computational power has also made model-based statistical reconstruction methods feasible for routine clinical studies. These methods employ system models to describe the scanner geometry and the data acquisition physics and statistical noise models to account for noise in the data. They use iterative numerical optimization algorithms to maximize a chosen cost function to obtain the final reconstructed image. As a result of these physical and statistical models, images reconstructed using model-based statistical methods have less noise and higher resolution compared to analytic reconstruction methods. All recent commercial PET scanners have model-based statistical reconstruction packages, and many clinics have switched to them for their routine clinical applications.

PET data acquisition involves counting of photon pairs as the radioactive isotope decays, which is a complex random process, and the data is usually very noisy. The spatial resolution of the data is limited due to the finite size of the detectors and other physical factors including positron range and noncollinearity of the photon pair. In general it is not possible to achieve the best image resolution and minimize noise at the same time. Thus one important goal of image reconstruction is to optimize the tradeoff between image noise and resolution.

This chapter reviews the techniques developed for PET image reconstruction and image property analysis. We introduce both mathematical theory and practical considerations, focusing on the methods commonly used on commercial PET scanners. We also briefly describe data corrections necessary for PET image reconstruction, which are important for reducing the artifacts and improving quantitative accuracy. The details of these corrections can be found in other chapters. Finally some recent developments are given, including the reconstruction of time-of-flight (TOF) PET data and direct parametric image reconstruction.

11.2 Analytic Image Reconstruction

11.2.1 PET Data Formation and Storage

In order to understand the theory of PET image reconstruction, let us first describe briefly how PET data is acquired and stored.

PET scanners are typically composed of multiple detector rings. Historically PET data was acquired in 2D mode with lead or tungsten septa inserted between detector rings to only allow for coincidences within the same or neighboring rings to be recorded and to reduce the number of scattered events. Figure 11.1 shows one ring of a typical PET gantry. When two photons are detected in the scanner, they are processed through electronics to check for various criteria such as whether each photon has the right amount of energy (typically between 350 and 650 keV) or whether the photons have arrived almost simultaneously (within 5–6 ns or less of each other). If all criteria are satisfied, then the two photons are recorded as a coincidence event. Each coincidence event can be stored separately as is the case with "list-mode" data, or alternatively, the total number of events detected at each detector pair can be stored. Usually such data is stored in a two-dimensional array for 2D PET indexed by radial element and view angle coordinates and is called a sinogram because measurements from a point source are concentrated on a sinusoid curve.

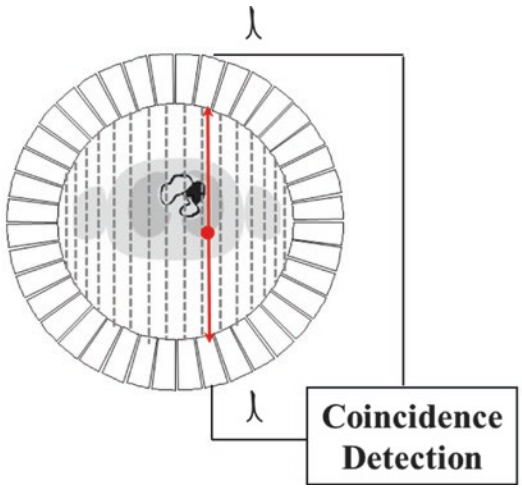

Fig. 11.1 One ring of a typical PET scanner and data processing

In modern PET scanners, the septa were removed for 3D data acquisition to improve the system sensitivity. For 3D PET, the data can be stored either as projections indexed by (x_r, y_r, θ, ϕ) or equivalently as stacks of sinograms indexed by (s, ϕ , z, θ), as shown in Fig. 11.2. Data on 3D PET scanners are typically stored in the stacked sinogram format and sometimes compressed by adding the neighboring angles or sinograms together.

11.2.2 Filtered Backprojection

11.2.2.1 Line Integral Model and Central Slice Theorem

A simple model for PET data assumes that the number of events detected at each detector pair is proportional to the integral of the radioactivity along the line connecting the centers of the two detectors. This is the basis upon which analytic image reconstruction algorithms are developed. Here we illustrate the line integral model for the 2D case.

Mathematically, the spatial distribution of the tracer is represented by a 2D continuous function $f(x,y)$. Measured projection data can then be approximated by the discrete samples of the x-ray transform of $f(x,y)$, which is defined by

$$p(s,\phi) = \int_{-\infty}^{\infty} f(s\cos\phi - t\sin\phi, s\sin\phi + t\cos\phi)\,dt \tag{11.1}$$

where s is the distance of the projection line to the center and ϕ is the projection angle, as shown in Fig. 11.2b.

An important result that underlies analytic image reconstruction is the central slice theorem. In 2D, the central slice theorem relates the 2D Fourier transform of the image to the 1D Fourier transform of its x-ray transform $p(s,\phi)$, as illustrated in Fig. 11.3:

$$F(\omega_s\cos\phi, \omega_s\sin\phi) = P(\omega_s;\phi) \tag{11.2}$$

where the 1D and 2D Fourier transforms are defined by

$$F(\omega_x,\omega_y) = \iint f(x,y)e^{-2\pi i(\omega_x x+\omega_y y)}dxdy$$
$$P(\omega) = \int p(s)e^{-2\pi i\omega s}ds \tag{11.3}$$

This relationship shows in 2D that the Radon transform uniquely describes any Fourier transformable image and that the image can be reconstructed by forming its 2D Fourier transform according to (Eq. 11.2) and taking an inverse 2D Fourier transform. Although this is a straightforward way of reconstructing 2D images, it requires the interpolation of the image's 2D Fourier transform onto a rectangular grid, and reconstructed image quality depends on the accuracy of the interpolation. As a result, direct Fourier reconstruction methods are not commonly used.

11.2.2.2 Backprojection Filtering

An image can be formed by integrating all the projections passing through a point and assigning the value of the integral to the point. This linear operation is called backprojection and can be expressed mathematically as:

$$b(x,y) = \int_0^{\pi} p(x\cos\phi + y\sin\phi, \phi)\,d\phi \tag{11.4}$$

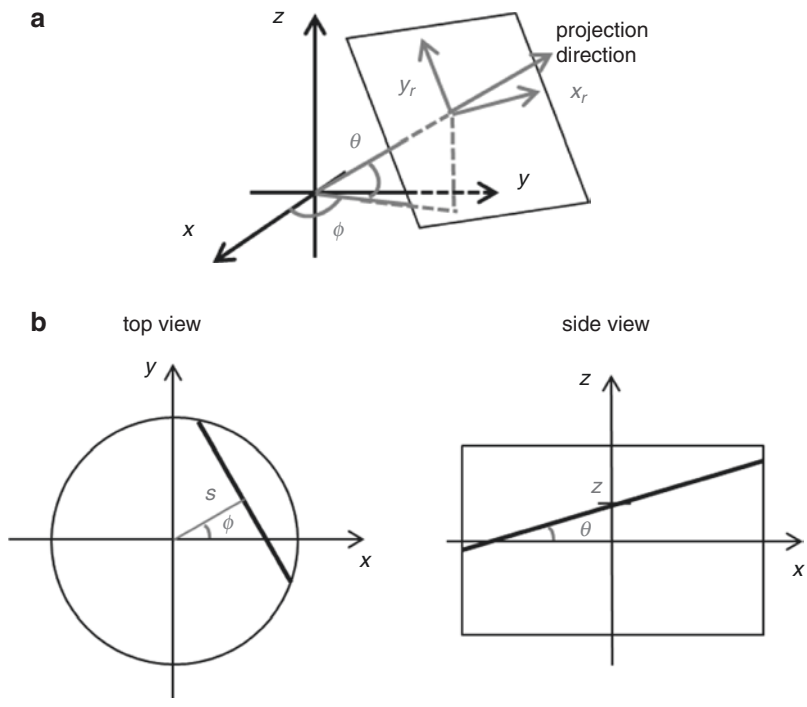

Fig. 11.2 3D PET data stored in (**a**) projection and (**b**) sinogram format

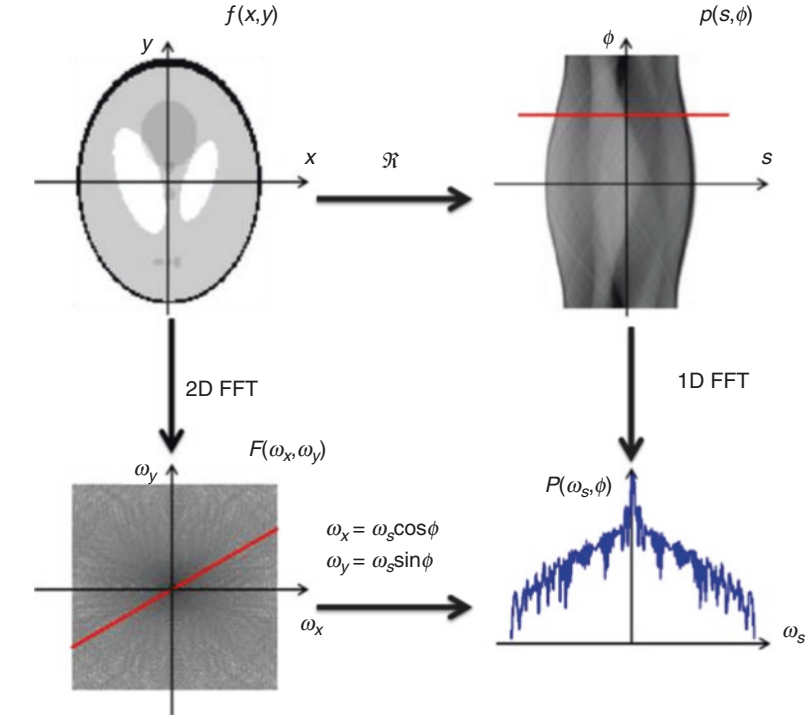

Fig. 11.3 2D central slice theorem

It can be shown that this backprojected image corresponds to the original image smoothed by the "1/r" blurring in Fourier transform domain:

$$B\left(\omega_x,\omega_y\right)=F\left(\omega_x,\omega_y\right)/\sqrt{\omega_x^2+\omega_y^2} \quad (11.5)$$

In order to recover the original image $f(x)$, we can filter this backprojected image, and the overall approach is known as backprojection filtering (BPF). A common approach is to switch the order of the linear backprojection and filtering operations and to apply filtering to the 1D projection data (usually by convolution with the filter function in projection space) before backprojection, thus called filtered backprojection or FBP [4]. As shown in (Eq. 11.5), the reconstruction filter has a frequency response of $|\omega|$. This so-called ramp filter amplifies high-frequency noise; therefore, in practice, a windowed ramp filter is used in FBP reconstruction and frequency components above a threshold (cutoff frequency) are set to zero. Frequently used window functions in 2D FBP include Shepp–Logan, Butterworth, and Hann windows [5, 6]. Figure 11.4 shows FBP images of a simulated NCAT phantom using different window functions.

11.2.3 Reconstruction of 3D PET Data

The 2D central slice theorem can be extended to 3D PET, resulting in the derivation of 3D FBP [7]. However, due to the limited axial coverage of PET scanners, not all oblique projections are recorded, and therefore some oblique projections through the imaging volume are missing. As a result, a typical 3D PET scanner corresponds to a shift–variant system where FBP cannot be directly applied to reconstruct the data. Figure 11.5 shows a direct line of response (LOR) which is acquired in 2D PET, an oblique LOR acquired in 3D PET and a missing oblique LOR in 3D PET that is not measured because the LOR does not intersect the detector surface.

One solution to the missing data problem is to estimate the truncated projections before applying 3D FBP. The 3D reprojection (3DRP) algorithm was proposed to estimate the missing data and has been a standard analytic reconstruction algorithm for 3D PET [8]. In 3DRP, unmeasured data is estimated by calculating the line integrals along the LORs through an initial image estimate obtained by applying 2D FBP to the non-oblique sinograms.

We note that 3DRP is based on the fact that 2D data is sufficient for reconstruction (e.g., 2D images reconstructed from line integrals on parallel planes can be stacked to form the final 3D image), and the goal of using additional 3D data is to improve the statistical properties of reconstructed images. While 2D and 3D reconstructions would be identical for reconstructions from noiseless data, at typical clinical data noise levels, by including the events detected by two detectors on different rings, 3D reconstructions significantly improve the signal-to-noise ratio (SNR) of the image.

Another way to reconstruct 3D PET data is to explore the redundancy in the data and to reduce it to a 2D dataset. Such a process is called rebinning. Since the computational cost of rebinning is negligible compared to the computational cost of image reconstruction, the resulting reconstruction speed is almost the same as that of 2D reconstructions.

A simple way to rebin 3D PET data is through the process called single slice rebinning (SSRB) [9] where a stack of 2D sinograms are created by placing detected events on the plane perpendicular to the scanner axis (z) and lying in the middle of the line connecting the two detectors that detected the event. Mathematically we have

$$p\left(s,\phi,z\right)=\sum_{\frac{z_1+z_2}{2}=z} p\left(s,\phi,z_1,z_2\right) \quad (11.6)$$

where s and ϕ are the coordinates of the line of response projected onto the x–y plane and z_1 and z_2 are the location of the two detector rings (z_1 may equal to z_2).

SSRB is only accurate when the activity is concentrated near the axis of the scanner. In cases of extended activity distributions, such as whole-body FDG scans, SSRB will become less accurate and introduce distortions. A more accurate rebinning method called Fourier rebinning

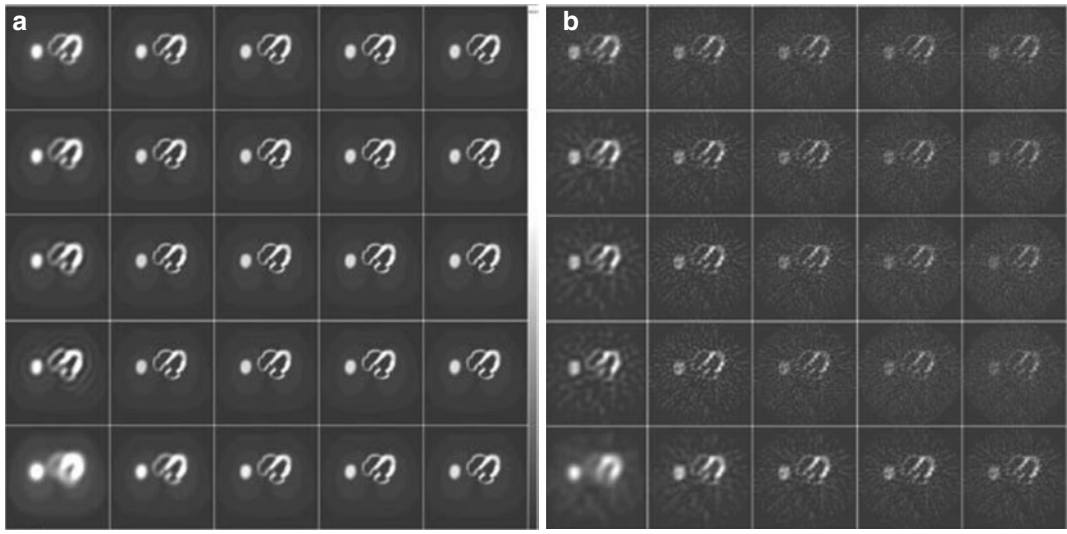

Fig. 11.4 Images reconstructed from simulated NCAT phantom (**a**) noiseless and (**b**) noisy data. *Top 4 rows* were reconstructed using Butterworth window with orders 2, 4, 8, and 32, respectively. *Bottom row* used Hann window. *Left to right*: cutoff frequencies of 0.1, 0.2, 0.3, 0.4, and 0.5 cycle/pixel, respectively (Reproduced from Tsui and Frey [6])

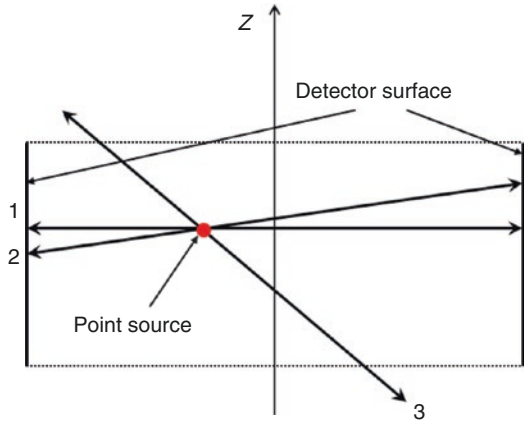

Fig. 11.5 Cross section of a PET scanner: events along LOR1 is acquired in 2D PET, LOR2 is acquired in 3D PET, while LOR3 is missing from the 3D PET data

(FORE) has been developed based on the relationship between the Fourier transforms of the 2D data and oblique data [10].

Following SSRB or FORE, a 2D reconstruction algorithm is applied to reconstruct each slice separately. The advantage of FORE is that it allows fast 2D reconstructions while retaining the SNR benefits from 3D data. Figure 11.6 shows a brain FDG scan data reconstructed using 3DRP and FORE+FBP. The image quality is comparable while the speed of FORE+FBP is more than ten times faster than 3DRP.

Although FORE has been widely used for clinical 3D PET image reconstruction, it is still based on the line integral geometric model and cannot model all the physical effects in PET data acquisition. In addition, FORE requires all the data corrections to be applied directly to the data (instead of being part of the system model), and therefore changes data noise properties which become difficult to model when a 2D iterative algorithm is used to reconstruct the rebinned dataset. As a result, fully 3D statistical reconstruction methods are preferable over FORE rebinning followed by 2D reconstruction methods.

11.3 Model-Based Statistical Reconstruction

Analytic image reconstruction methods inherently assume that noiseless line integrals of the image are available. In reality, data acquired in typical PET scanners are not exact line integrals, and there is significant statistical noise in almost all clinical datasets. Model-based statistical image

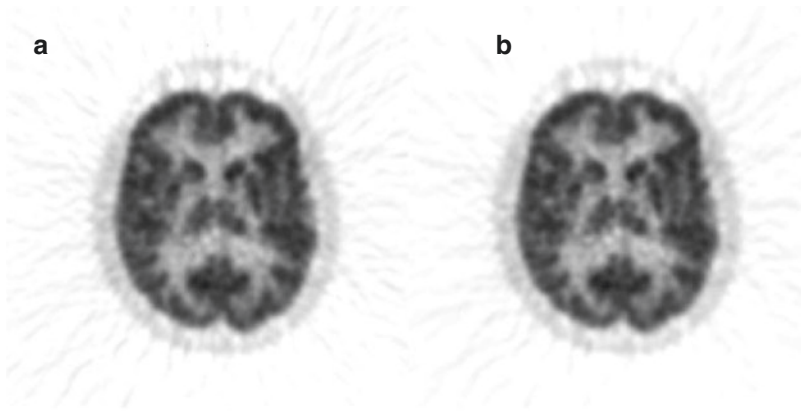

Fig. 11.6 A central transaxial slice of an FDG brain scan reconstructed using (**a**) 3DRP and (**b**) FORE+FBP [10]

reconstruction approaches are preferable due to their ability to model the physics and statistics of the imaging process and have become widely available in all commercial scanners. In this section, we first introduce statistical noise models for PET data and physical system models which describe the complex physics of data acquisition. We then review the methodology of maximum likelihood (ML) and maximum-a-posteriori (MAP) estimation and numerical optimization algorithms used to generate PET images. Due to the nonlinear, shift-varying and high-dimensional nature of statistical reconstruction algorithms, image resolution and noise analyses are complicated and are still an active research area. We will give a brief review on this topic in the end.

11.3.1 Noise Model

PET data is inherently noisy, and this fact has significant effects on reconstructed image quality. Statistical reconstruction algorithms model the noise in PET data explicitly and use iterative numerical optimization algorithms to solve the associated optimization problem. Over the last four decades, many model-based statistical image reconstruction algorithms have been proposed. Although the formulae of these methods are quite different, they are all designed to solve the following problem:

$$y = \bar{y}(x) + \text{noise} \tag{11.7}$$

where y is the measured data, $\bar{y}$ denotes the mean of the data, and x is the image of the unknown activity distribution. The number of events detected at a detector within a given time due to the radioactive decay inside the object can be accurately modeled by the Poisson distribution:

$$P(n \mid \lambda) = \frac{e^{-\lambda}\lambda^{n}}{n!} \tag{11.8}$$

where n is the number of decays and λ is the mean, which is equal to the variance.

If system dead time and detector pileup effects are ignored, then measured data can be modeled as independent Poisson random variables with joint distribution:

$$P(y \mid \bar{y}) = \prod_{i=1}^{M} \frac{e^{-\bar{y}_i}\bar{y}_i^{y_i}}{y_i!} \tag{11.9}$$

where y_i is the number of events detected in the ith LOR, $\bar{y}_i$ is the mean number of events in the ith LOR that can be calculated using the system model described in the next section, and M is the number of LORs.

While a Gaussian noise model may be used for low-noise data with acceptable accuracy, resulting in the weighted least square (WLS) approach for image reconstruction [11], most statistical approaches use the Poisson model.

In PET, a common practice is to subtract an estimate of the random events (e.g., using delayed window method) online to reduce the bandwidth needed for data transfer and storage space [12]. In that case, measured data at the *i*th LOR is given by

$$y_i = p_i - r_i \tag{11.10}$$

where p_i is the number of prompts and r_i is the number of delayed events for the *i*th LOR. While both p_i and r_i are independent Poisson random variables, the difference between the two is no longer Poisson. The distribution of y_i has a numerically intractable form. One can use Gaussian distribution as an approximation to the true distribution of the precorrected data. Yavuz and Fessler noticed that a simple but good approximation is to add $2\overline{r_i}$ to the data [13], where $\overline{r_i}$ is the mean of the delayed events estimated from the data and to model the resulting random variable as Poisson with mean and variance equal to $\overline{y}_i + 2\overline{r_i}$.

As we discussed previously, Fourier rebinning is commonly used to reduce the size of the dataset and reconstruction time. It has been shown that FORE rebinned data is no longer Poisson [14]. Similar to the approach for random precorrected data, Liu et al. used a simple scaling to match the mean and variance of the rebinned data [15]. Then a reconstruction algorithm based on the Poisson noise model can be used to reconstruct the resulting data.

11.3.2 System Model

The spatial resolution of PET is limited by several factors such as positron range, photon noncollinearity, and penetration and scattering of the photon in the detector. One critical limitation of analytic reconstruction methods is that these factors are neglected in the simple line integral model. With model-based statistical reconstruction, we use a system model to account for these resolution-deteriorating effects. Other factors may also be included in the system model such as the attenuation of the photons in the body, nonuniform efficiencies of the detectors, and random and scattered events.

In the absence of noise, we can model the data as a linear function of the image:

$$\overline{y}_i = \sum_{j=1}^{N} p_{ij} x_j + \overline{r}_i \tag{11.11}$$

or in matrix format $\overline{y} = Px + \overline{r}$

where $\overline{r}$ is the sum of mean of random and scattered events and P is the system matrix, which relates the image to the noiseless data, and can be expressed in factored form as [16]:

$$P = P_{\text{norm}} P_{\text{blur}} P_{\text{attn}} P_{\text{geom}} P_{\text{range}} \tag{11.12}$$

where P_{range} models the blurring due to positron range in image space [17, 18]; P_{geom} is the geometric projection matrix containing the geometric detection probabilities for each voxel and detector pair combination; P_{attn} is a diagonal matrix with attenuation factors; P_{blur} models the blurring in sinogram space due to photon pair noncollinearity, intercrystal penetration, and scattering; and P_{norm} is a diagonal matrix with normalization factors.

Matrix–vector multiplications with P and P^T (called forward and backprojection operations) are typically the most computationally intensive components of statistical reconstructions. Over the last decade, considerable amount of research has been done to speed up statistical reconstruction by reducing the time of forward and backprojections. Many algorithms have been developed that explore the symmetry of the scanner geometry [19] or use fast processors such as graphic processing units (GPUs) to achieve high performance at low cost [20].

It has been shown that the modeling of sinogram blurring is important for resolution recovery [16]. The central component of resolution recovery is the estimation of detector point spread functions (PSF). Such PSFs can be calculated analytically [21], estimated from Monte Carlo simulations [16], or measured from point source data [22, 23]. PSF kernels can either be estimated and stored nonparametrically [23], or the PSF measurements can be fit to a given model, such as asymmetric Gaussian functions [22], and the

resulting model parameters can be stored. It has also been shown that for Fourier rebinned data, PSF kernels can be estimated from point source data [23, 24].

Another recent trend is to use an image-space PSF model to account for all resolution degrading effects in the imaging process:

$$P = P_{\text{norm}} P_{\text{attn}} P_{\text{geom}} P_{\text{psf}} \tag{11.13}$$

The image-space PSF P_{psf} can be estimated from an initial reconstruction of point sources without any resolution recovery and is easy to implement as an image-space blurring operation. Recently shift–variant PSFs have been designed to model the degradation of image resolution toward the edge of the field of view (FOV) [25].

Figure 11.7 shows a Hoffman brain phantom data reconstructed with and without PSF modeling. The PSF image shows better resolution and contrast.

One caveat of PSF modeling is that it can result in edge artifacts that have been shown to overestimate activity by up to 40 % in phantom studies [26]. Methods to mitigate edge artifacts include filtering [27] and under-modeling the PSF kernels [28]; however these approaches come at the expense of reduced resolution recovery.

11.3.3 Maximum Likelihood Estimation Methods

Once we have the noise model and system model, statistical methods can be used to estimate the image. Maximum likelihood (ML) is a widely used statistical estimation method and has been applied to PET image reconstruction.

The likelihood function of the data is the probability of observing the data given the image. Usually the logarithm of the likelihood function is used for easier calculation because the logarithm is a one-to-one monotonic function:

$$L(x) = \log\left(\text{Prob}\left(y \mid x\right)\right) \tag{11.14}$$

ML estimation seeks the image that maximizes the log-likelihood function:

$$x_{ML} = \arg\max_{x \geq 0} L(x) \tag{11.15}$$

The nonnegativity constraint is due to the fact that the concentration of radioactivity is not negative.

The log-likelihood function under the Poisson noise model is given by (omitting the $y_i!$ term which does not depend on x)

$$L(x) = \sum_{i=1}^{M} y_i \log \bar{y}_i(x) - \bar{y}_i(x) \tag{11.16}$$

ML estimation is a classic optimization problem, where the cost function (or objective function) is the log-likelihood function (Eq. 11.16). There are many numerical algorithms that can be used to find the ML estimate of the image, such as coordinate ascent or gradient-based methods. One of the earliest approaches used for ML PET image reconstruction is the expectation–maximization (EM) algorithm [29]. EM is a general framework to compute the ML solution through the use of "complete" but unobservable data and is composed of two steps. The first step, called the E-step, involves the calculation of the conditional expectation of the complete data, and the second step, called the M-step, maximizes this conditional expectation with respect to the image. In PET, a very common choice for complete data is the number of events detected by the *i*th LOR that are emitted from the *j*th voxel. Shepp and Vardi first applied EM to emission image reconstruction [30], and Lange and Carson extended the work to transmission image reconstruction [31].

The update equation of EM algorithm for emission image reconstruction is given by

$$x_j^{k+1} = \frac{x_j^k}{\sum_{i'} P_{i'j}} \sum_i P_{ij} \frac{y_i}{\sum_{j'} P_{ij'} x_{j'}^k} \tag{11.17}$$

where y_i is the number of events acquired in the *i*th LOR, P_{ij} is the element in the system matrix P described in the previous section, and x_j^k is the estimate of the *j*th image voxel at the *k*th iteration.

The EM algorithm is usually initialized using a uniform image. A new image is then calculated using Eq. 11.17. This process is repeated until a

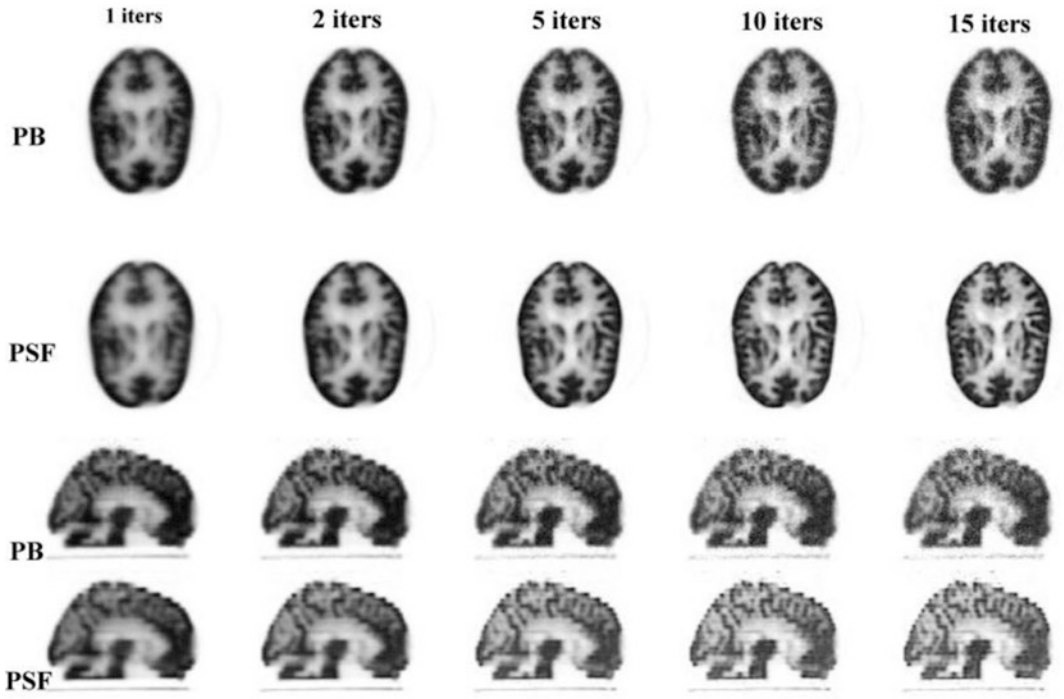

Fig. 11.7 Hoffman brain phantom reconstructions for various numbers of iterations. Upper images are transaxial views, and the lower images are sagittal views. *PB* parallel-beam, non-PSF [22]

certain number of iterations are reached. Typically a smoothing filter is applied afterward to reduce image noise.

It is interesting to note that the EM update equation can be derived in several ways. One approach is based on using the concavity of the Poisson log-likelihood function and Jensen's inequality [32]. In another work, Qi and Leahy showed that that EM algorithm is a functional substitution (FS) method [33]. FS is based on designing a surrogate function at each iteration which is easier to maximize than the original function. Under suitable conditions (equal function value and gradient with current estimated image, surrogate function always higher than the original function), it can be proven that the FS method converges to the maxima of the original function [34–36].

It has been shown that EM converges monotonically to the global maximum of the log-likelihood function and the image is guaranteed to be nonnegative if initialized with a nonnegative image [37]. These nice properties made EM a popular algorithm but unfortunately EM converges very slowly. Hundreds or thousands of iterations are usually necessary to ensure the convergence of EM. Hudson and Larkin observed that the convergence of EM algorithm could be significantly speeded up by dividing the projection data into nonoverlapping blocks, or subsets, and applying EM to each subset of the data [38] and named this method ordered subsets EM (OSEM).

OSEM can speed up the reconstruction almost linearly as a function of number of subsets in early iterations. As a result, it has been widely used in clinical PET and SPECT. The speedup of OSEM can be explained in part by viewing EM as a preconditioned gradient-ascent method:

$$x^{k+1} = x^k + D\left(x^k\right)\nabla L\left(x^k\right) \qquad (11.18)$$

where $D(x^k)$ is a diagonal matrix with $D_{jj} = x_j^k / \sum_i p_{ij}$ and $\nabla L\left(x^k\right)$

denotes the gradient of the log-likelihood calculated at the kth iteration image. When the image is far away from the solution, the gradient computed from a subset of the data provides a satisfactory direction for increasing the log-likelihood function value [39].

Although OSEM is fast in the beginning, it is not guaranteed to converge to the ML solution, and the convergence depends on the selection of the subsets. In the original OSEM paper, subsets are chosen such that the detection probability of each voxel is equal for each subset, which is called subset balance. It has been proven that with consistent data and under the condition of subset balance, OSEM converges to the ML solution [38]. In practice subset balance is difficult to achieve due to differences in sensitivity and attenuation. A common practice is to choose the projections in each subset with maximum angular separation to avoid directional artifacts. Another OSEM algorithm convergent with consistent data is rescaled block-iterative EM (RBI-EM), in which the OSEM equation is written in gradient-ascent form and a voxel-independent scaling factor is used to avoid the requirement of subset balance [40].

In general PET data is not consistent due to noise, in which case OSEM usually enters a limit cycle [39]. One way to make OSEM converge with noisy data is to reduce the number of subsets: however convergence is then slowed down. Alternatively, we can use OSEM in the earlier iterations and switch to a convergent algorithm in later iterations [41].

Among other ML estimation approaches, Browne and De Pierro proposed a row-action maximum likelihood algorithm (RAMLA) [42], where the stepsize η_k at the kth iteration satisfies the following conditions:

$$\lim_{k\to 0}\eta_k = 0, \sum_{k=0}^{\infty}\eta_k = +\infty \qquad (11.19)$$

RAMLA converges to ML solution if the log-likelihood function is strictly concave [42]. However the convergence rate can be very slow.

Another way to achieve convergence is through the use of an augmented cost function [39], and an example of this method is the convergent OSEM algorithm (COSEM) [43].

Despite its convergence issues, OSEM is still by far the most popularly used iterative reconstruction algorithm in clinical nuclear medicine. Figure 11.8 shows OSEM images from a monkey brain phantom reconstructed using different numbers of iterations. It shows that image noise increases quickly as number of iterations is increased. As a result, clinical OSEM images are usually reconstructed using less than five iterations, and a smoothing filter is applied post-OSEM to reduce image noise.

In Fig. 11.8 we also show an image reconstructed using the maximum-a-posteriori (MAP) method. In MAP reconstruction regularization is used to ensure a stable solution is reached, as demonstrated by the MAP image after 25 iterations, which shows fine structures and low noise. We will introduce MAP reconstruction methods in the next section.

11.3.4 Maximum-a-Posteriori Estimation Methods

11.3.4.1 Prior Functions

It has been observed that with ML estimators such as EM or OSEM, iterating beyond a certain point will dramatically increase noise and decrease image quality [44]. The reason is the inherent ill-posedness of the PET inverse problem. The system matrix is ill-conditioned so that small differences in the data (which naturally occur due to photon counting noise) produce large changes in the ML solution. It can be shown that the variance of the estimator increases as voxel sizes are reduced [33]. The noise in the OSEM image can be controlled by stopping the algorithm before convergence [45] or post-smoothing the image after many iterations [46]. An alternative approach to dealing with the instability of the ML problem is to use regularization in image reconstruction. Here we take the Bayesian formulation and view the regularization function as the log of the prior function. Thus we have the following *maximum-a-posteriori* (MAP) estimator:

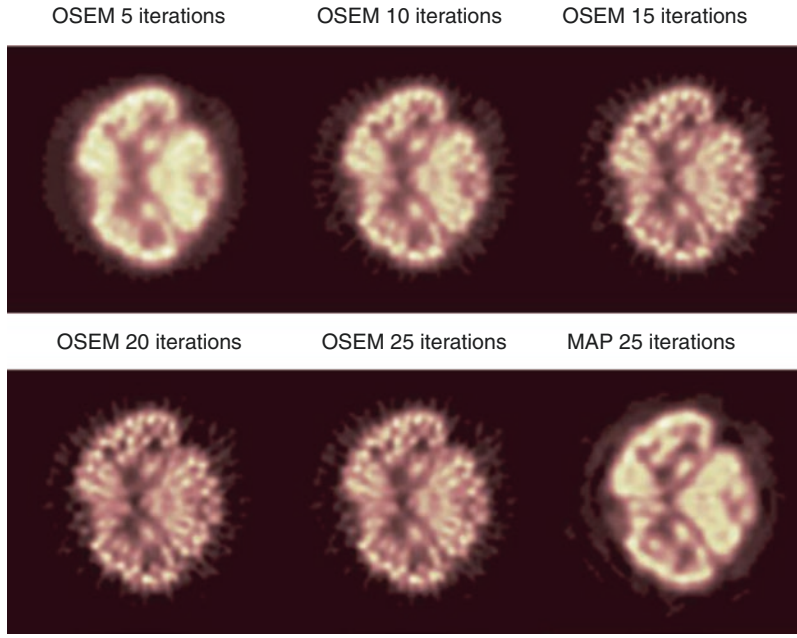

Fig. 11.8 Monkey brain phantom images reconstructed using OSEM with different number of iterations and MAP with 25 iterations

$$x_{MAP} = \arg\max_{x \geq 0} \Phi(x) = L(y \mid x) + p(x) \tag{11.20}$$

where $L(y|x)$ is the log-likelihood function and $p(x)$ is the log-prior (or penalty) function.

MAP estimation allows us to use other information about the image such as smoothness or anatomical information in the form of the prior function to improve image quality. The effect of the prior distribution is to choose among those images that have similar likelihood values the one that is most probable with respect to the prior.

There are many ways of designing the prior function for the image. One simple method is to ignore the statistical dependence of the pixels and treat each one separately. Independent Gaussian [21] and Gamma models [47] have been proposed. For these models, we need to estimate the mean image, which can be difficult and introduces significant bias. Several methods have been developed to estimate the mean image [48, 49]. The attraction of this model is that closed form update equations can be found in the generalized EM framework.

Independent voxel models are of limited value since the information we typically seek to capture in the prior is some degree of piecewise smoothness in the image. In order to model the local smoothness of the image, we can use Markov random field (MRF) models or Gibbs distributions [50]. The general form of a Gibbs distribution is given by

$$p(x \mid \beta) = \frac{1}{Z} e^{-\beta U(x)} \tag{11.21}$$

where Z is a normalization factor and $U(x)$ is the Gibbs energy function which is the sum of potentials. Each potential is defined on a subset of voxels called a "clique." The cliques are composed of neighboring voxels that are mutual neighbors [51]. β is a hyperparameter which controls the image smoothness.

The Gibbs distribution allows us to specify a prior in terms of local interactions since all MRFs have the property that the conditional probability of any voxel in the image depends only on the values of the voxels in a local neighborhood of that voxel. This not only allows us to model the desired local properties but also leads

to computationally tractable MAP reconstruction algorithms.

To define a MRF prior, we need to specify the form of statistical interaction (or conditioning) between neighboring voxels. In image reconstruction, the potential functions are usually defined on pairwise voxels:

$$U(x) = \sum_{j=1}^{N} \sum_{k \in N_j, k>j} \phi\left(x_j - x_k\right) \quad (11.22)$$

with the constraint that $\phi(x_j - x_k) \geq 0$ and that the function is monotonically nondecreasing in $\left|x_j - x_k\right|$.

The function $\phi(x_i - x_j)$ determines the properties of the prior. The simplest one is the quadratic penalty: $\phi(x_i - x_j) = (x_i - x_j)^2$. When the image has the same value at each voxel, the potential functions and hence the energy function $U(x)$ are zero, and the probability density function $P(x) = 1/Z \exp(-\beta U(x))$ is at its maximum value. As the difference between voxels increases, so does $U(x)$, with $P(x)$ decreasing correspondingly, indicating less likely images. The advantage of this weighting is that it tends to produce the most natural-looking images. However, quadratic prior functions are limited in their abilities to identify sharp changes in intensity. A large number of alternative functions have been explored that reduce the penalty for larger differences such as the generalized *p*-Gaussian model [52] and the Huber prior [16].

Recently total variation (TV) regularization has become increasingly popular in medical imaging, especially CT image reconstruction [53], but it can also be applied to PET and SPECT image reconstruction [54–56]. The TV prior is equivalent to the TV norm that is now commonly used in the context of sparse imaging [57].

11.3.4.2 Non-convex and Anatomical Priors

The above prior functions are convex. In order to further encourage sharp edges, non-convex functions have also been proposed [58]. Recently a relative distance prior has been proposed and used in clinical PET image reconstruction [59, 60]. The general goal of these functions is to assign lower probability to images that are not locally smooth but without over-penalizing occasional large changes that might correspond, for example, to organ boundaries.

When high-resolution anatomical information is available from co-registered MR or CT images, it is also possible to incorporate it into the prior function [61]. While anatomical and functional images clearly give very different views on the human body, it is also true that functional images, whether they represent metabolism, blood volume, or receptor binding, will exhibit a spatial morphology that reflects the underlying anatomy. It is reasonable to assume that most tracers exhibit distinct changes in activity across tissue boundaries, while inside each region the distribution is smooth, unless there is evidence to the contrary in the functional data itself. The important thing is that we do not force the abrupt change of activity across the boundaries but rather model changes in these locations as more likely. There are two main approaches to using anatomical information: In the first approach, edge or region information is used to penalize activity changes near the boundaries to a lesser degree. The other approach maximizes an information-based similarity measure between the anatomical and reconstructed functional image, such as the Kullback–Leibler (KL) distance, joint entropy (JE), or mutual information (MI) [61].

Figure 11.9 shows brain PET images reconstructed using FBP, MLEM, and MAP with JE prior. The MAP image has better resolution which clearly shows the fine structures in the brain.

Although there is improvement in quantitation and detection performance when the anatomical prior provides accurate information about the location of boundaries of lesions, these advantages are largely lost when the quality of the anatomical image deteriorates or when there is a mismatch between anatomical and functional boundaries. In addition, the resolution of images reconstructed using anatomical priors is anisotropic and spatially variant, making them more challenging to interpret. Anatomical priors have not been widely applied to clinical PET studies.

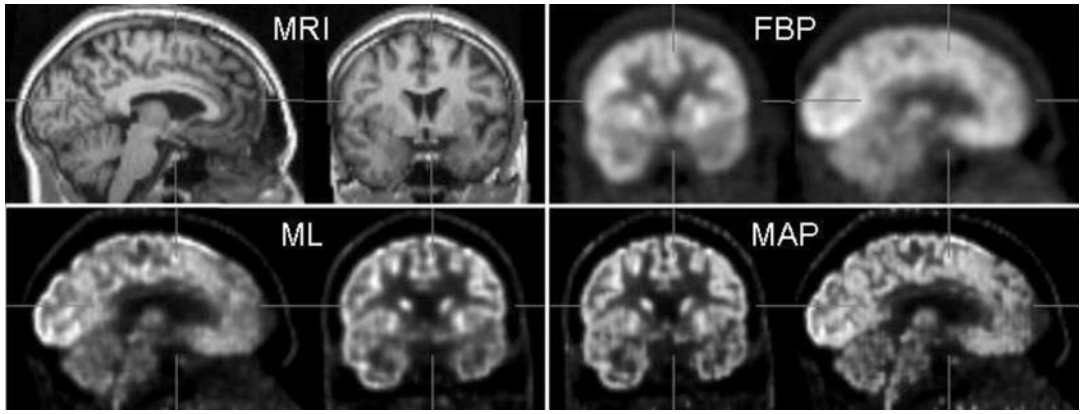

Fig. 11.9 Sagittal and coronal slices of MRI images and PET reconstruction with FBP, ML, and MAP (using JE prior) [62]

11.3.4.3 MAP Reconstruction Algorithms

Most of the optimization algorithms used for ML can be extended to MAP estimation. When EM is applied to MAP, the M-step has a closed form solution only when the prior function is spatially independent. For spatially coupled priors, an iterative method such as gradient or coordinate ascent can be applied in the M-step [63]. Green proposed the "one-step-late" (OSL) approach where the partial derivatives of the prior function are evaluated using the current estimate [64]. However the OSL method does not in general converge to the MAP solution, and the estimated image is not guaranteed to be nonnegative.

Another approach is to apply the FS methodology and design separable surrogate functions for the prior function [65]. In that case, a closed form solution can be found for the M-step.

Standard gradient-based methods can also be applied to the MAP estimation problem. For example, the preconditioned conjugate gradient algorithm has been used for PET image reconstruction [66]. The preconditioner is critical for the convergence of the algorithm. A simple diagonal preconditioner derived from the EM algorithm has been shown to be very effective [67]. Several methods have been proposed to approximate the ideal preconditioner (the inverse of the Hessian matrix), including using the inverse of the diagonal of the Hessian matrix [68], Fourier-based preconditioners [69], and matrix factorization [70].

Unlike the EM algorithm, the nonnegativity constraint in gradient-based methods needs to be handled explicitly. This can be achieved by several strategies, including restricting the step size [71], bent-line search followed by a second-line search [72], or using the active set [66]. Penalty functions [67] and interior point methods [68] have also been used.

MAP estimation problem is better conditioned than ML; thus reconstructed images are less noisy compared to ML. The convergence rate is also much faster, and good quality images can be reconstructed with tens of iterations instead of hundreds or thousands of iterations as in ML-EM. Another advantage of MAP estimators is that image quality can be controlled through carefully designed prior functions [33, 73]. We note that when the prior function contains data-dependent terms, the estimator is no longer truly a MAP estimator (these are typically referred to as penalized likelihood estimators since the second term is no longer a prior but a penalty that penalizes image roughness).

Figure 11.10 compares FBP with model-based statistical reconstruction of a monkey brain phantom data. The superior image quality resulting from the statistical algorithm is clear.

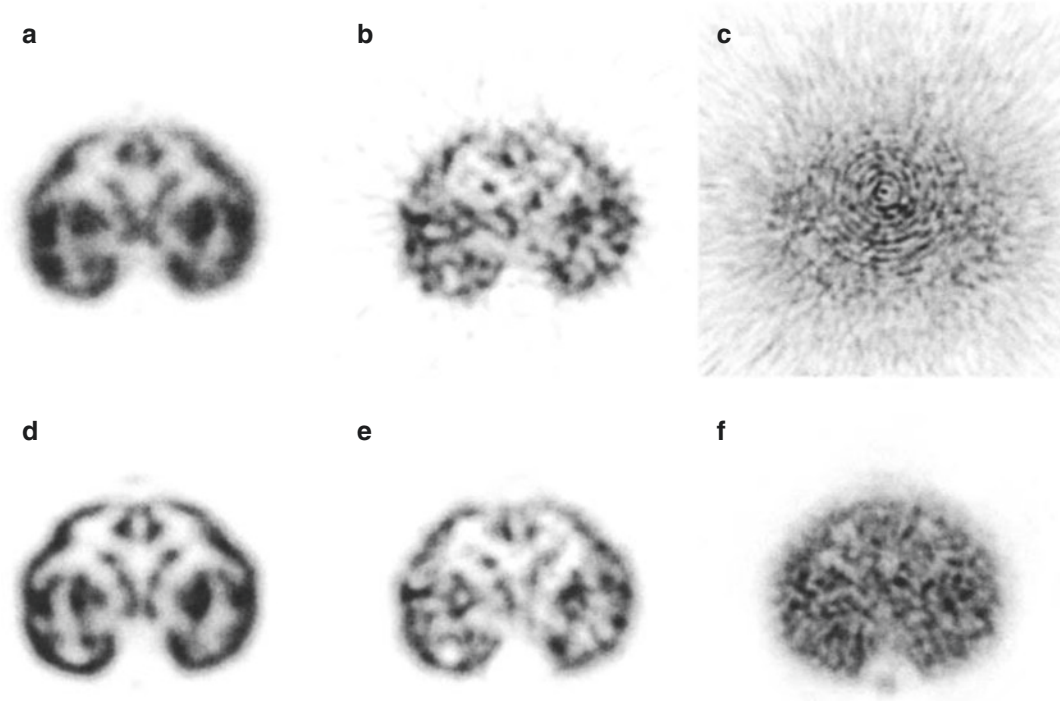

Fig. 11.10 Monkey brain phantom image from 60 scans. (**a**–**c**) FBP and (**d**–**f**) MAP. (**a**) Reconstructed from summed data. (**b**) From 1 scan. (**c**) Variance image [74]

11.3.5 Analysis of Image Properties

11.3.5.1 Image Resolution

Image properties of ML and MAP estimators are more difficult to analyze compared to analytic methods, because these methods are nonlinear and have large image and sinogram dimensions. For statistical image reconstruction, the system response is shift variant, especially for 3D systems. One way to quantify image resolution is by using the local impulse response (LIR) [75, 76]:

$$l^j(x) = \lim_{\delta \to 0} \frac{\mu\left(x + \delta e^j\right) - \mu(x)}{\delta} \quad (11.23)$$

where $\mu(x)$ is the mean of the estimated image when the true image is x. e^j denotes the unit vector where the *jth* voxel is 1 and all other voxels are 0. A good approximation to the mean of the estimated image can be obtained from the estimation of the noiseless data [76]. Analytic approximations of the LIR function for MAP estimators can also be derived [76]. The computation of the LIR involves the inversion of a large matrix composed of the Fisher information matrix (FIM) and the Hessian matrix of the prior function, an operation that is impractical for real data. Several Fourier transform-based approximations to the FIM have been proposed to make the inversion practical [77–79].

Interestingly, it has been shown that image resolution gets worse in regions with higher activity [76]. One can take the resolution analysis one step further and design spatially-variant weightings for the prior function to achieve desired image resolution, for example, to make the contrast recovery coefficient (CRC, defined as the peak of the LIR) spatially uniform [76, 77].

Not only is the resolution spatially variant for MAP, it is also non-isotropic. This results in distortion of the activity shape. Stayman and Fessler proposed to estimate the coefficients in

the prior function to match a predefined target response function in shift-invariant imaging systems [73]. Later the method was generalized to more realistic, shift–variant PET and SPECT systems [80].

The above methods require iterative computation of the coefficients in the prior function. A simple, non-iterative method has been used in PET image reconstruction to achieve count-independent resolution [81]. The resulting image resolution is roughly data-independent, thus can be pre-calculated for different values of the hyperparameter β using phantom measurements or simulated data. In real patient or animal scans, the hyperparameter β can be selected for desired resolution [81]. Figure 11.11 shows the calibration result and measured resolution from a validation phantom scan. It shows that the predicted (ED) resolution is very close to the actual measured resolution. The slight mismatch is caused by the line-source location differences in calibration and validation scans.

11.3.5.2 Image Noise

Analytic expressions of image noise for MLEM have been derived and shown to be accurate for low-noise data [82]. Similar analyses have also been performed for MAP reconstructions [77, 83]. These methods can be used to characterize the reconstructed image and optimize imaging systems and reconstruction algorithms. In general, images with higher resolution also have higher noise; thus the image reconstruction approach needs to balance resolution and noise in order to achieve optimal image quality.

11.3.5.3 Image Property Measurements

In practice, phantom scans are frequently used to measure image resolution, contrast, and noise. For resolution measurements a point source or line source is typically scanned, and the full width at half maximum (FWHM) is measured from the reconstructed image profile through the source. However, when evaluating statistical reconstruction algorithms, the point source or line source needs to be surrounded by a warm background to avoid potential biases caused by the nonnegativity constraint. Figure 11.12 shows the resolution vs. noise plots measured from an 8.3 cm diameter cylindrical phantom on a microPET Focus220 scanner. Five needles were inserted into the phantom along a line with

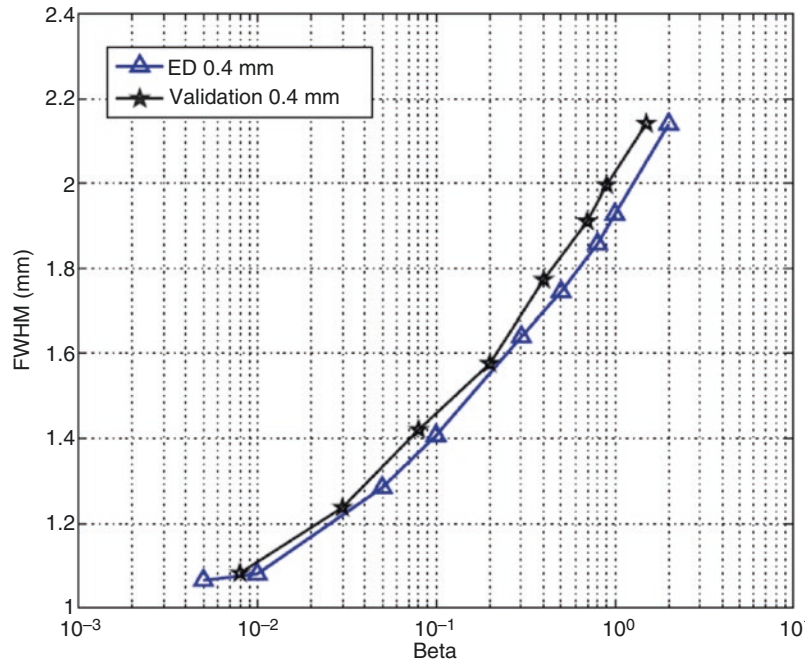

Fig. 11.11 Predicted and experimental resolution from phantom study. *ED* Experimentally generated table [81]

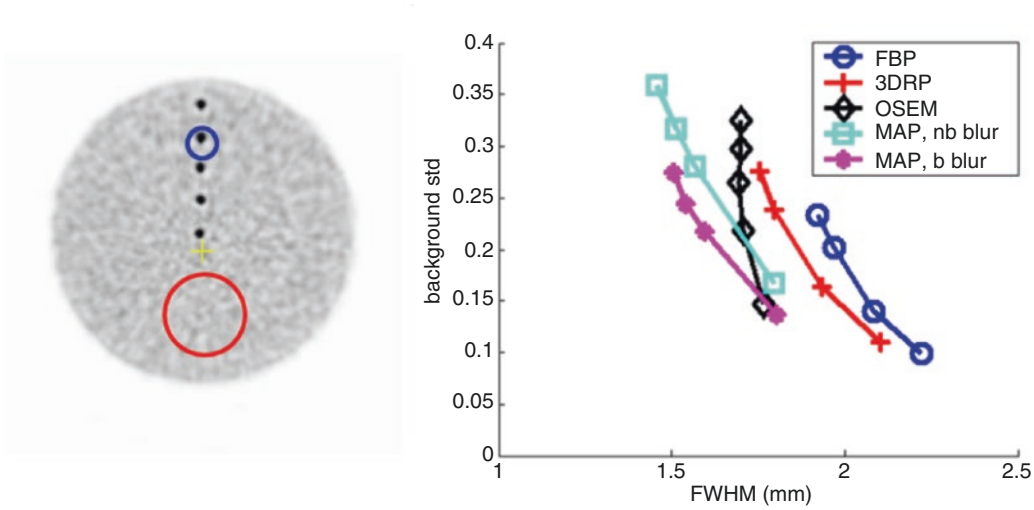

Fig. 11.12 Image resolution and noise measured using line sources in a cylindrical phantom. *Left*: central slice of the phantom. *Red circle* shows the background ROI; *blue circle* shows the ROI used to calculate resolution. *Right*: noise vs. resolution measurement

0.8 cm spacing in radial direction. One needle was placed at the center of the cylinder. The outer and inner diameters of the needles were 0.76 mm and 0.28 mm, respectively. The phantom was filled with a 69.8 MBq ^{18}F solution. The ratio between the concentrations of the activity in the needles and the background was 300:1. The phantom was placed at the center of the field of view and scanned for 15 min. Data were reconstructed using several algorithms including FBP, 3DRP, OSEM, and MAP with all the corrections. A large 3D region of interest (ROI) was selected in the background. The noise of the image is defined as the standard deviation of the 3D ROI normalized by the mean of the same ROI. Radial resolutions were calculated by selecting a small region around the line source 3 cm from the center in each image plane and fitting the profile through the peak to a Gaussian model. The FWHM of the 10 image planes at the axial center was averaged as the resolution of the image. For MAP reconstruction, four different smoothing parameters were used to achieve different resolution. System model using block and non-block blurring kernels was used for MAP reconstructions. The block blurring kernels model the PET scanner geometry more accurately. For OSEM a variety of number of iterations, ranging from 4 to 20, were used. FBP and 3DRP images were reconstructed with a ramp filter, followed by smoothing using Gaussian filters with different FWHM. The result shows that MAP with the block blurring kernels has the highest spatial resolution (smallest FWHM) at a given noise level, compared to other methods.

11.4 PET Image Quantitation

PET is a highly sensitive, quantitative molecular imaging technology, which is frequently used to measure the spatial and occasionally the temporal distribution of the radiotracers injected into the human or animal body. Accuracy of the PET measurement depends on many factors including patient physiology, scanner technology, and reconstruction method. Here we briefly describe the PET quantitation issues related to image reconstruction.

11.4.1 PET Data Corrections

The raw PET data is not only noisy but also distorted and corrupted during the acquisition process. Several corrections need to be applied in

order to obtain quantitatively accurate and artifact-free PET images.

11.4.1.1 Normalization

In PET, the detector response varies due to the intrinsic properties of the detectors, geometry and changing count-rate [84]. Normalization is used to compensate for these variations. Accurate normalization is essential for reducing artifacts and for accurate quantitation in PET. Modern PET and PET/CT scanners usually use component-based approaches, where the normalization factors are decomposed into several factors, for example, detector efficiencies, geometric factors, and dead time factors [84–86]. To measure the components, a known source of activity is scanned, and then the normalization factors are estimated by comparing the ideal numbers of coincidences and those actually measured. Rotating rod sources, uniform plane sources, and uniform cylindrical sources have been used for normalization.

The normalization factors can be estimated sequentially. For example, the detector efficiencies can be estimated first, and then the data can be corrected using these estimated detector efficiencies before estimating the geometric factors [86]. Alternatively we can jointly estimate the count-rate-independent factors using statistical estimation methods, where the statistical noise can also be taken into account to significantly reduce the number of events needed for normalization scans [85]. Further discussion on detector normalization can be found in Chap. 10.

11.4.1.2 Attenuation Correction

Due to the interactions between photons and body tissues through Compton scattering and photoelectric absorption, a substantial fraction of the 511 keV photons resulting from positron–electron annihilation is attenuated in the body, and the number of counts along the directions in which the photons originally travel is reduced. It has been estimated that as much as 80 % of the photon pairs emitted from the center of the human brain are lost due to attenuation [87] and the number is even higher for body imaging.

On dedicated PET scanners, the attenuation correction factors are usually measured using an external source. Two scans are needed for this purpose: a blank scan in which the FOV is empty and a transmission scan, which is taken with the subject in the scanner. Correction factors can be computed as the ratio of blank to transmission data. The blank scan can be acquired for a long time to reduce the noise; however the transmission scan is usually short and the resulting data is therefore noisy. As a result, the calculated attenuation correction factors are quite noisy. To address this problem, the reconstruction–reprojection method has been proposed. In this method, an image of attenuation coefficients is computed, and attenuation correction factors are calculated from the projection of this image [88].

With the development of combined PET/CT and PET/MR scanners, it is possible to use the coregistered CT or MR image to calculate the attenuation correction factors. For CT it is relatively straightforward since the CT image is a measurement of the attenuation coefficient of the body, although with a lower, poly-energetic spectrum. It has been shown that a piecewise linear transform can be used to convert the CT Hounsfield units (HU) to the attenuation coefficients of 511 keV photons [89]. For MR the conversion is more difficult because of the lack of signal from the bone [90]. PET attenuation correction using MR image is still a very active research area. See Chap. 9.

11.4.1.3 Scattered and Random Events

The dominant interaction between 511 keV photons and the surrounding atoms is Compton scattering [91]. In 3D PET the portion of scattered coincidences in all detected events can be as high as 30–50 % compared to 10–15 % in the 2D case [92]. Scattered events are mispositioned in the recorded data, which introduce a smooth background in the reconstructed PET image.

The scatter distribution depends on the emission source distribution, the scattering medium, and the detector geometry. Deterministic algorithms for calculating scatter distributions have been proposed [93, 94]. With the computational power available today, it is also possible to

estimate the scatter distribution using Monte Carlo simulations [95]. Since the distribution of scattered events is smooth, deterministic methods give reasonable results in most cases and are widely used, especially in commercial PET scanners [94].

A random event is recorded when two photons unrelated to each other are detected within the coincidence timing window. The fraction of random events depends on the amount of activity in the body and can be significant for studies using large amounts of radioactivity. Usually a "delayed window" method is used to estimate the amount of random events [12]. The measured "delayed events" are usually very noisy, and an iterative method has been proposed to estimate its mean [67].

Figure 11.13 shows a typical sequence of corrections applied to the data before it is input to a reconstruction algorithm. For analytic reconstruction methods, these corrections are applied to the data before the reconstruction. With statistical reconstruction methods, precorrection changes noise properties of the data, and the exact distribution of the corrected data is no longer Poisson and difficult to calculate. A better approach for statistical reconstruction methods is to incorporate the corrections in the system model, as previously presented.

11.4.2 PET Image Calibration

Value in the PET image represents the concentration of the radioisotope, usually in the units of Bq/ml or nCi/cc. For clinical PET/CT, a cross calibration procedure is usually performed. During cross calibration, a source (typically cylindrical ^{18}F or ^{68}Ge/^{68}Ga source) with known activity concentration is scanned in the scanner. A large ROI is drawn in the reconstructed image, and the mean of this ROI is measured to calculate the scaling factor that converts PET image values to standard units.

The decay of the radioactive nucleus and branching fraction also need to be considered when quantifying the PET image. PET scanner calibration and quality control procedures are reviewed in Chap. 10.

11.5 Recent Developments

11.5.1 Time of Flight PET

In time-of-flight (TOF) PET, the arrival times of the two 511 keV photons are measured. By calculating the difference between the arrival times, we can narrow down the location of the source to a portion of the LOR.

As with non-TOF-PET, both analytic and model-based statistical methods can be used to reconstruct TOF-PET data, with similar advantages and limitations. The additional TOF information can be used to improve image quality, especially for large patients.

All the model-based statistical reconstruction algorithms described in Section 3 can be extended directly to TOF data. The only changes are the

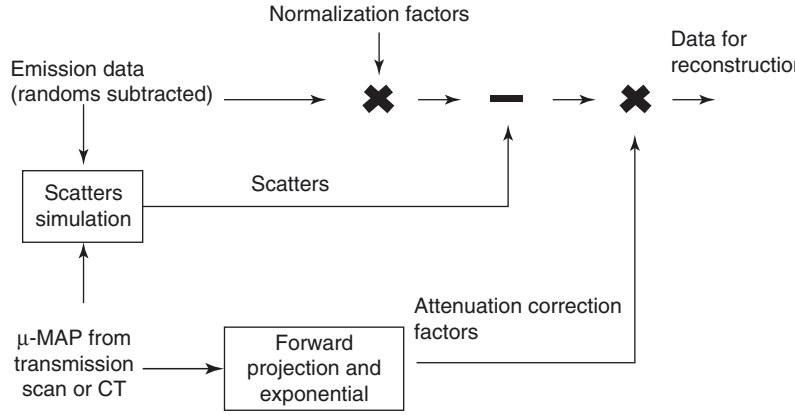

Fig. 11.13 Data corrections applied to PET data before reconstruction

following: (1) there is one more dimension of the data to represent TOF; (2) the system matrix needs to be modified to include TOF. For analytic reconstruction algorithms, TOF changes the reconstruction formula. The following discussion focuses on 2D case, and similar methods can also be derived for 3D TOF-PET.

2D TOF-PET data can be expressed as:

$$p(s,\phi,t) \approx \int_{-\infty}^{\infty} f(-s\sin\phi + l\cos\phi, s\cos\phi + l\sin\phi) h(t-l)\,dl \tag{11.24}$$

where t is the TOF variable. $h(\tau)$ is the TOF kernel, which is usually assumed to be a Gaussian function with FWHM determined by the system hardware and software.

It can be shown that the 2D Fourier transform of the 2D TOF sinogram in one projection is the FT of the image weighted by the 1D FT of the TOF kernel. FBP-type algorithms can be derived for TOF data. A general form of these methods in frequency domain is [96]

$$F(v) = \int_0^{\pi} \frac{W(\hat{u},v) P(\varphi,v)\,d\varphi}{\int_0^{\pi} W(\hat{u}',v) H(\hat{u}'\cdot v)\,d\varphi'} \tag{11.25}$$

where $F(v)$ is the 2D FT of the image, $P(\phi,v)$ is the 2D FT of the TOF projection at angle ϕ, $W(\hat{u},v)$ is a filter, which represents a weighting scheme of the projection data in frequency domain, $u = (-\sin\phi, \cos\phi)$ is the unit vector in the projection direction, and $H(\omega)$ is the 1D FT of the TOF kernel. Apparently the algorithm depends on the choice of the filter $W(\hat{u},v)$, which must satisfy the condition that

$$\int_0^{\pi} W(u,v) H(u\cdot v)\,d\phi \neq 0, \forall v \tag{11.26}$$

Due to the redundancy of the TOF data, FBP algorithm does not have a unique formula. In case of noiseless data, they give the same results. However, due to the noise of the data, the performance of the algorithms using different filters is not the same. A widely used method is confidence weighting, where the chosen filter is $W(u,v) = H(u\cdot v)$.

Given the current timing resolution of TOF-PET scanners (~500 pico-seconds, or 7.5 cm FWHM), the spatial resolution of TOF-PET images is essentially the same as non-TOF-PET. On the other hand, using the TOF information can improve the SNR of the image [96–98]. The improvement of SNR is approximately proportional to D/h, where D is the diameter of the object and h is the TOF spatial resolution. However TOF reconstruction is much slower due to the large dataset. Recently Cho et al. developed a generalized projection slice theorem and proposed a unified framework for mapping between different datasets, allowing us to explore and optimize the use of TOF and 3D PET information [99, 100]. Figure 11.14 displays the mappings between PET datasets.

Two Fourier rebinning algorithms of time-of-flight data (FORET) have been developed [100]. The 3D TOF data can be rebinned into 3D non-TOF data using FORET3D (or 2D non-TOF data using FORET2D), which is a weighted average of the TOF data in Fourier transform domain. The SNR of the data depends on the weighting scheme. Under reasonable assumptions, it can be proven that the best linear unbiased estimator of the 3D non-TOF data is achieved when using FORET3D to rebin 3D TOF data and the weight-

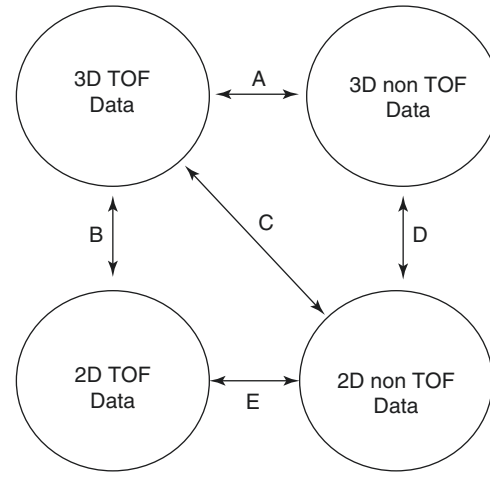

Fig. 11.14 Possible mappings between different PET data [100]

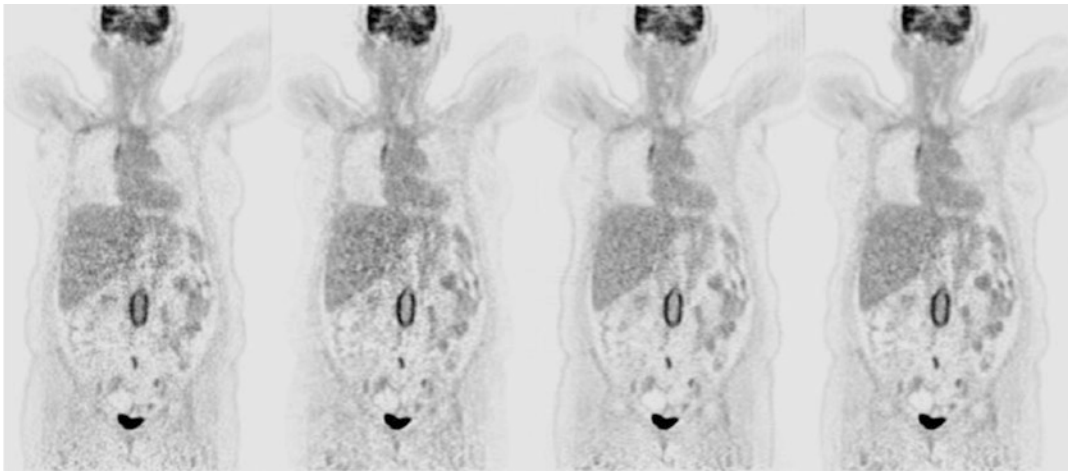

Fig. 11.15 FDG patient images. *Left to right*: non-TOF MAP, FORET2D + MAP, FORET3D + MAP, TOF MAP [24]

Table 11.1 Mean and noise measurements from the patient image in Fig. 11.15 [24]

Method	Mean	SD	%SD
TOF	0.577	0.109	18.9
FORET-3D	0.569	0.110	19.3
FORET-2D	0.617	0.151	24.5
Non-TOF	0.568	0.155	27.3

ing factors equal to the square of the Fourier transform of the TOF kernel [97]. The SNR gain is approximately a constant in this case. For FORET2D, which rebins 3D TOF data into 2D non-TOF data, it can be shown that the SNR gain is roughly constant for each 2D sinogram, which can be calculated very accurately [24]. Figure 11.15 shows a patient data reconstructed from non-TOF, FORET2D, FORET3D, and TOF data using a statistical algorithm. Table 11.1 shows the noise in the images measured from an ROI in the liver of the patient image in Fig. 11.15. The smoothing parameters were selected such that all the images have the same resolution (6 mm FWHM). These results show that the quality of FORET3D image is similar to TOF image, while FORET2D image is noisier than FORET3D, but still better than the non-TOF image. Since FORET uses FFT, the computation is very fast. The reconstruction speed of FORET3D is similar to non-TOF reconstruction, which is about five times faster than 3D TOF reconstruction. FORET2D can further increase the speed by a factor of 5, at the expense of more noise in the image.

11.5.2 Parametric Image Reconstruction

We focus on the reconstruction of static data in this chapter. PET is a dynamic modality where data can be acquired continuously for a period of time and useful kinetic parameters can be calculated from it [101]. A straightforward approach is to divide the data into many frames, reconstruct each frame, and then calculate the kinetic parameters (either voxel-wise or ROI-wise) from the time–activity curves extracted from these images. Another class of image reconstruction methods are designed to reconstruct kinetic parameters from the measurements directly [102–104]. Figure 11.16 shows a whole-body FDG patient study. The patient was stepped through the scanner twice and Patlak image calculated from the list-mode data. Static PET images were also calculated from the two passes. We note that although the visual difference between Patlak image and static images is subtle, the Patlak image is a voxel-wise quantitative image and the values are more informative. These direct and indirect parametric reconstruction methods require the estimation of the arterial input function which can either be measured during the scan by

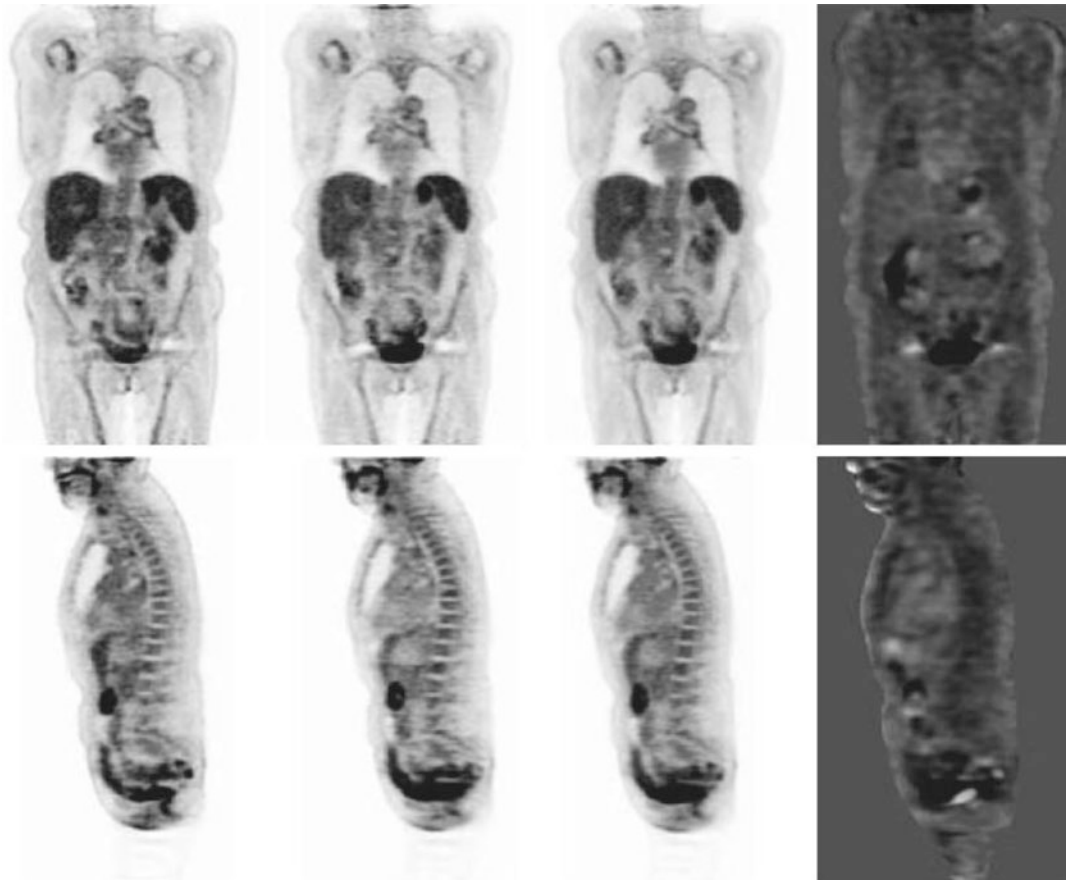

Fig. 11.16 Whole-body FDG patient images. From *left to right*: first frame, second frame, Patlak slope, %DSUV [102]

arterial blood sampling or from reconstructed multiframe images. In Fig. 11.16, the arterial input function was estimated using a hybrid approach which combines a population-based approach with a simplified exponential model [104].

Conclusions

The last two decades have seen tremendous progress in PET image reconstruction. Once considered slow and only used for research studies, model-based statistical reconstruction methods are now applied in nuclear medicine clinics every day for tumor detection and staging, assessment of therapy response, and other applications. By combining an appropriate noise model with an accurate system model, we are able to improve image resolution and noise properties, which increase the sensitivity and specificity of clinical studies. There has also been a lot of work in regularized reconstruction algorithms where we can effectively control image properties and optimize them for different tasks by designing appropriate regularization functions that might also include anatomical information. This is in contrast to the analytic methods, where the only control options are the ramp filter cutoff frequencies and any post-filters applied to the images. As a result, optimal results can be obtained for different imaging applications.

Coupled with the development of PET instrumentation, including the new detector material and data correction techniques, PET image quality and quantitative accuracy have improved significantly. Statistical PET image reconstruction is still an active research field, especially in TOF-PET, MR-PET, and parametric imaging.

References

1. Radon J. On determination of functions by their integral values along certain multiplicities. Ber Sachische Akad Wiss LeipzigGermany. 1917;69: 262–77.
2. Hounsfield GN. Computerized transverse axial scanning (tomography): part 1. Description of system. Br J Radiol. 1973;46(552):1016–22.
3. Natterer F. Computerized tomography. In: The mathematics of computerized tomography. New York: John Wiley and Sons Inc; 1986. p. 1–8.
4. Shepp LA, Logan B. The Fourier reconstruction of a head section. IEEE Trans Nucl Sci. 1974;21:21–33.
5. Farquhar TH, Chatziioannou A, Chinn G, Dahlbom M, Hoffman EJ. An investigation of filter choice for filtered back-projection reconstruction in PET. Presented at the Nuclear Science Symposium and Medical Imaging Conference, vol. 2; 1997. p. 1042–6.
6. Tsui B, Frey E. Analytic image reconstruction methods in emission computed tomography. In: Quantitative analysis in nuclear medicine imaging. New York: Springer; 2006. p. 82–106.
7. Colsher JG. Fully three-dimensional positron emission tomography. Phys Med Biol. 1980;25(1):103–15.
8. Kinahan PE, Rogers JG. Analytic 3D image reconstruction using all detected events. IEEE Trans Nucl Sci. 1989;36(1):964–8.
9. Daube-Witherspoon ME, Muehllehner G. Treatment of axial data in three-dimensional PET. J Nucl Med. 1987;28(11):1717–24.
10. Defrise M, Kinahan PE, Townsend DW, Michel C, Sibomana M, Newport DF. Exact and approximate rebinning algorithms for 3-D PET data. IEEE Trans Med Imaging. 1997;16(2):145–58.
11. Fessler JA. Penalized weighted least-squares image reconstruction for positron emission tomography. IEEE Trans Med Imaging. 1994;13(2):290–300.
12. Hoffman EJ, Huang S-C, Phelps ME, Kuhl DE. Quantitation in positron emission computed tomography: 4. Effect of accidental coincidences. J Comput Assist Tomogr. 1981;5(3):391–400.
13. Yavuz M, Fessler JA. Statistical image reconstruction methods for randoms-precorrected PET scans. Med Image Anal. 1998;2(4):369–78.
14. Comtat C, Kinahan PE, Defrise M, Michel C, Townsend DW. Fast reconstruction of 3D PET data with accurate statistical modeling. IEEE Trans Nucl Sci. 1998;45(3):1083–9.
15. Liu X, Comtat C, Michel C, Kinahan P, Defrise M, Townsend D. Comparison of 3-D reconstruction with 3D-OSEM and with FORE+OSEM for PET. IEEE Trans Med Imaging. 2001;20(8):804–14.
16. Qi J, Leahy RM, Cherry SR, Chatziioannou A, Farquhar TH. High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner. Phys Med Biol. 1998;43(4):1001–13.
17. Bai B, Ruangma A, Laforest R, Tai YC, Leahy RM. Positron range modeling for statistical PET image reconstruction. Presented at the Nuclear Science Symposium and Medical Imaging Conference Record, vol. 4; 2003. p. 2501–5.
18. Fu L, Qi J. A residual correction method for high-resolution PET reconstruction with application to on-the-fly Monte Carlo based model of positron range. Med Phys. 2010;37:704.
19. Hong I, Chung S, Kim H, Kim Y, Son Y, Cho Z. Ultra fast symmetry and SIMD-based projection-backprojection (SSP) algorithm for 3-D PET image reconstruction. IEEE Trans Med Imaging. 2007;26(6):789–803.
20. Pratx G, Chinn G, Olcott PD, Levin CS. Fast, accurate and shift-varying line projections for iterative reconstruction using the GPU. IEEE Trans Med Imaging. 2009;28(3):435–45.
21. Huesman RH, Klein GJ, Moses WW, Qi J, Reutter BW, Virador PR. List-mode maximum-likelihood reconstruction applied to positron emission mammography (PEM) with irregular sampling. IEEE Trans Med Imaging. 2000;19(5):532–7.
22. Panin VY, Kehren F, Michel C, Casey M. Fully 3-D PET reconstruction with system matrix derived from point source measurements. IEEE Trans Med Imaging. 2006;25(7):907–21.
23. Tohme MS, Qi J. Iterative image reconstruction for positron emission tomography based on a detector response function estimated from point source measurements. Phys Med Biol. 2009;54(12):3709.
24. Bai B, Lin Y, Zhu W, Ren R, Li Q, Dahlbom M, DiFilippo F, Leahy RM. MAP reconstruction for Fourier rebinned TOF-PET data. Phys Med Biol. 2014;59(4):925.
25. Rahmim A, Qi J, Sossi V. Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls. Med Phys. 2013;40:064301.
26. Bai B, Esser PD. The effect of edge artifacts on quantification of Positron Emission Tomography. Presented at the Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC); 2010. p. 2263–6.
27. Tong S, Alessio AM, Thielemans K, Stearns C, Ross S, Kinahan PE. Properties and mitigation of edge artifacts in PSF-based PET reconstruction. IEEE Trans Nucl Sci. 2011;58(5):2264–75.
28. Watson CC. Estimating effective model kernel widths for PSF reconstruction in PET. Presented at the Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC); 2011. p. 2368–74.
29. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Methodol. 1977;39:1–38.
30. Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. IEEE Trans Med Imaging. 1982;1(2):113–22.
31. Lange K, Carson R. EM reconstruction algorithms for emission and transmission tomography. J Comput Assist Tomogr. 1984;8(2):306–16.
32. De Pierro AR. On the relation between the ISRA and the EM algorithm for positron emission tomography. IEEE Trans Med Imaging. 1993;12(2):328–33.

33. Qi J, Leahy RM. Iterative reconstruction techniques in emission computed tomography. Phys Med Biol. 2006;51(15):R541–78.
34. Lange K, Fessler JA. Globally convergent algorithms for maximum a posteriori transmission tomography. IEEE Trans Image Process. 1995;4(10):1430–8.
35. Hunter DR, Lange K. A tutorial on MM algorithms. Am Stat. 2004;58(1):30–7.
36. Jacobson MW, Fessler JA. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. IEEE Trans Image Process. 2007;16(10):2411–22.
37. Vardi Y, Shepp LA, Kaufman L. A statistical model for positron emission tomography. J Am Stat Assoc. 1985;80(389):8–20.
38. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans Med Imaging. 1994;13(4):601–9.
39. Ahn S, Fessler JA. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. IEEE Trans Med Imaging. 2003;22(5):613–26.
40. Byrne CL. Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods. IEEE Trans Image Process. 1998;7(1):100–9.
41. Li Q, Ahn S, Leahy R. Fast hybrid algorithms for PET image reconstruction. Presented at the Nuclear Science Symposium and Medical Imaging Conference Record, vol. 4; 2005. p. 1851–5.
42. Browne J, De Pierro AB. A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography. IEEE Trans Med Imaging. 1996;15(5):687–99.
43. Hsiao IT, Huang HM. An accelerated ordered subsets reconstruction algorithm using an accelerating power factor for emission tomography. Phys Med Biol. 2010;55(3):599–614.
44. Snyder DL, Miller MI, Thomas LJ, Politte DG. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. IEEE Trans Med Imaging. 1987;6(3):228–38.
45. Veklerov E, Llacer J. Stopping rule for the MLE algorithm based on statistical hypothesis testing. IEEE Trans Med Imaging. 1987;6(4):313–9.
46. Llacer J, Veklerov E, Coakley KJ, Hoffman EJ, Nunez J. Statistical analysis of maximum likelihood estimator images of human brain FDG PET studies. IEEE Trans Med Imaging. 1993;12(2):215–31.
47. Lange K, Bahn M, Little R. A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. IEEE Trans Med Imaging. 1987;6(2):106–14.
48. Alenius S, Ruotsalainen U. Bayesian image reconstruction for emission tomography based on median root prior. Eur J Nucl Med. 1997;24(3):258–65.
49. Hsiao T, Rangarajan A, Gindi G. A new convex edge-preserving median prior with applications to tomography. IEEE Trans Med Imaging. 2003;22(5):580–5.
50. Besag J. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Ser B Methodol. 1974;36:192–236.
51. Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell. 1984;6(6):721–41.
52. Bouman CA, Sauer K. A unified approach to statistical tomography using coordinate descent optimization. IEEE Trans Image Process. 1996;5(3):480–92.
53. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys Med Biol. 2008;53(17):4777.
54. Panin VY, Zeng GL, Gullberg GT. Total variation regulated EM algorithm [SPECT reconstruction]. IEEE Trans Nucl Sci. 1999;46(6):2202–10.
55. Bai B. An interior-point method for total variation regularized positron emission tomography image reconstruction; 2012. p. 83136B1–83136B6.
56. Wang G, Qi J. Edge-preserving PET image reconstruction using trust optimization transfer, Presented at the 12th International Meeting on fully three-dimensional image reconstruction in radiology and nuclear medicine; 2013. p. 70–3.
57. Candes EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory. 2006;52(2):489–509.
58. Geman S, McClure D. Bayesian image analysis: an application to single photon emission tomography. Presented at the American Statistical Association; 1985. p. 12–8.
59. Ahn S, Ross SG, Asma E, Miao J, Jin X, Cheng L, Wollenweber SD, Manjeshwar RM. Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET. Phys Med Biol. 2015;60(15):5733.
60. Nuyts J, Bequé D, Dupont P, Mortelmans L. A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. IEEE Trans Nucl Sci. 2002;49(1):56–60.
61. Bai B, Li Q, Leahy RM. Magnetic resonance-guided positron emission tomography image reconstruction. Presented at the Seminars in Nuclear Medicine, vol. 43;2013. p. 30–44.
62. Nuyts J. The use of mutual information and joint entropy for anatomical priors in emission tomography. Presented at the Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE, vol. 6; 2007. p. 4149–54.
63. Hebert T, Leahy R. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. IEEE Trans Med Imaging. 1989;8(2):194–202.

64. Green PJ. Bayesian reconstructions from emission tomography data using a modified EM algorithm. IEEE Trans Med Imaging. 1990;9(1):84–93.
65. De Pierro AR. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. IEEE Trans Med Imaging. 1995;14(1):132–7.
66. √ú Mumcuoglu E, Leahy RM, Cherry SR. Bayesian reconstruction of PET images: methodology and performance analysis. Phys Med Biol. 1996;41(9):1777.
67. Mumcuoglu EU, Leahy R, Cherry SR, Zhou Z. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. IEEE Trans Med Imaging. 1994;13(4):687–701.
68. Johnson CA, Seidel J, Sofer A. Interior-point methodology for 3-D PET reconstruction. IEEE Trans Med Imaging. 2000;19(4):271–85.
69. Fessler JA, Booth SD. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. IEEE Trans Image Process. 1999;8(5):688–99.
70. Chinn G, Huang S-C. A general class of preconditioners for statistical iterative reconstruction of emission computed tomography. IEEE Trans Med Imaging. 1997;16(1):1–10.
71. Kaufman L. Implementing and accelerating the EM algorithm for positron emission tomography. IEEE Trans Med Imaging. 1987;6(1):37–51.
72. Kaufman L. Maximum likelihood, least squares, and penalized least squares for PET. IEEE Trans Med Imaging. 1993;12(2):200–14.
73. Stayman JW, Fessler JA. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. IEEE Trans Med Imaging. 2000;19(6):601–15.
74. Chatziioannou A, Qi J, Moore A, Annala A, Nguyen K, Leahy R, Cherry SR. Comparison of 3-D maximum a posteriori and filtered backprojection algorithms for high-resolution animal imaging with microPET. IEEE Trans Med Imaging. 2000;19(5):507–12.
75. Wilson DW, Tsui BMW. Spacial resolution properties of FB and ML-EM reconstruction methods. Presented at the Nuclear Science Symposium and Medical Imaging Conference; 1993. p. 1189–93.
76. Fessler JA, Rogers WL. Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. IEEE Trans Image Process. 1996;5(9):1346–58.
77. Qi J, Leahy RM. Resolution and noise properties of MAP reconstruction for fully 3-D PET. IEEE Trans Med Imaging. 2000;19(5):493–506.
78. Fessler JA. Analytical approach to regularization design for isotropic spatial resolution. Presented at the Nuclear Science Symposium Conference Record, 2003 IEEE, vol. 3; 2003. p. 2022–6.
79. Vunckx K, Zhou L, Matej S, Defrise M, Nuyts J. Fisher information-based evaluation of image quality for time-of-flight PET. IEEE Trans Med Imaging. 2010;29(2):311–21.
80. Stayman JW, Fessler JA. Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems. IEEE Trans Med Imaging. 2004;23(3):269–84.
81. Li Q, Bai B, Cho S, Smith A, Leahy R. Count independent resolution and its calibration. Presented at the 10th International Meeting on fully three-dimensional image reconstruction in radiology and nuclear medicine; 2009. p. 223–6.
82. Barrett HH, Wilson DW, Tsui BM. Noise properties of the EM algorithm: I. Theory. Phys Med Biol. 1994;39(5):833–46.
83. Fessler JA. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. IEEE Trans Image Process. 1996;5(3):493–506.
84. Badawi R, Marsden P. Developments in component-based normalization for 3D PET. Phys Med Biol. 1999;44(2):571.
85. Bai B, Li Q, Holdsworth CH, Asma E, Tai YC, Chatziioannou A, Leahy RM. Model-based normalization for iterative 3D PET image reconstruction. Phys Med Biol. 2002;47(15):2773–84.
86. Defrise M, Townsend DW, Bailey D, Geissbuhler A, Michel C, Jones T. A normalization technique for 3D PET data. Phys Med Biol. 1991;36(7):939–52.
87. Cherry SR, Phelps ME. Imaging brain function with positron emission tomography. In: Brain Mapping Methods. New York: Academic; 1996. p. 191–221.
88. Riederer SJ. Application of the noise power spectrum to positron emission CT self-absorption correction. Med Phys. 1981;8(2):220–4.
89. Carney J, Townsend DW, Rappoport V, Bendriem B. Method for transforming CT images for attenuation correction in PET/CT imaging. Med Phys. 2006;33(4):976–83.
90. Hofmann M, Pichler B, Schölkopf B, Beyer T. Towards quantitative PET/MRI: a review of MR-based attenuation correction techniques. Eur J Nucl Med Mol Imaging. 2009;36(1):93–104.
91. Cherry SR, Sorenson JA, Phelps ME. Physics in nuclear medicine. Philadelphia: Elsevier Health Sciences; 2012.
92. Watson CC, Newport D, Casey ME. A single scatter simulation technique for scatter correction in 3D PET. In: Three-dimensional image reconstruction in radiology and nuclear medicine. Boston: Kluwer Academic Publishers; 1996. p. 255–68.
93. Ollinger JM. Model-based scatter correction for fully 3D PET. Phys Med Biol. 1996;41(1):153.
94. Watson CC. New, faster, image-based scatter correction for 3D PET. IEEE Trans Nucl Sci. 2000;47(4):1587–94.
95. Holdsworth H, Levin C, Janecek M, Dahlbom M, Hoffman E. Performance analysis of an improved 3-D PET Monte Carlo simulation and scatter correction. IEEE Trans Nucl Sci. 2002;49(1):83–9.

96. Watson CC. An evaluation of image noise variance for time-of-flight PET. IEEE Trans Nucl Sci. 2007;54(5):1639–47.
97. Ahn S, Cho S, Li Q, Lin Y, Leahy RM. Optimal rebinning of time-of-flight PET data. IEEE Trans Med Imaging. 2011;30(10):1808–18.
98. Tomitani T. Image reconstruction and noise evaluation in photon time-of-flight assisted positron emission tomography. IEEE Trans Nucl Sci. 1981;28(6): 4581–9.
99. Cho S, Ahn S, Li Q, Leahy RM. Analytical properties of time-of-flight PET data. Phys Med Biol. 2008;53(11):2809–21.
100. Cho S, Ahn S, Li Q, Leahy RM. Exact and approximate Fourier rebinning of PET data from time-of-flight to non time-of-flight. Phys Med Biol. 2009;54(3):467–84.
101. Carson RE. Tracer kinetic modeling in PET. In: Positron emission tomography. London: Springer; 2005. p. 127–59.
102. Kamasak ME, Bouman CA, Morris ED, Sauer K. Direct reconstruction of kinetic parameter images from dynamic PET data. Med. Imaging IEEE Trans. 2005:636–50.
103. Karakatsanis NA, Lodge MA, Tahari AK, Zhou Y, Wahl RL, Rahmim A. Dynamic whole-body PET parametric imaging: I. Concept, acquisition protocol optimization and clinical application. Phys. Med. Biol. 2013;58:7391.
104. Zhu W, Bai B, Conti PS, Li Q, Leahy RM. Data correction methods for wholebody Patlak imaging from list-mode PET data. Presented at the 12th International Meeting on fully 3D image reconstruction in radiology and nuclear medicine, Lake Tahoe; 2013. p. 213–16.