

TOPICAL REVIEW • OPEN ACCESS

## 4D image reconstruction for emission tomography

To cite this article: Andrew J Reader and Jeroen Verhaeghe 2014 *Phys. Med. Biol.* **59** R371

View the [article online](#) for updates and enhancements.

You may also like

- [A generative adversarial network \(GAN\)-based technique for synthesizing realistic respiratory motion in the extended cardiac-torso \(XCAT\) phantoms](#)  
Yushi Chang, Zhuoran Jiang, William Paul Segars et al.
- [Application of adaptive kinetic modelling for bias propagation reduction in direct 4D image reconstruction](#)  
F A Kotasidis, J C Matthews, A J Reader et al.
- [Target volume and motion position evaluation of four-dimensional cone-beam CT: comparison with 4D-CT using dynamic thorax phantom](#)  
Yasuhiro Doi, Yoshinobu Shimohigashi, Yohei Yotsuji et al.

## Topical Review

# 4D image reconstruction for emission tomography

Andrew J Reader<sup>1,2</sup> and Jeroen Verhaeghe<sup>3</sup>

<sup>1</sup> Division of Imaging Sciences and Biomedical Engineering, Department of Biomedical Engineering, King's College London, St. Thomas' Hospital, London, UK

<sup>2</sup> Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

<sup>3</sup> Molecular Imaging Centre Antwerp, University of Antwerp, Antwerp, Belgium

Received 20 March 2014, revised 30 May 2014

Accepted for publication 24 June 2014

Published 31 October 2014

## Abstract

An overview of the theory of 4D image reconstruction for emission tomography is given along with a review of the current state of the art, covering both positron emission tomography and single photon emission computed tomography (SPECT). By viewing 4D image reconstruction as a matter of either linear or non-linear parameter estimation for a set of spatiotemporal functions chosen to approximately represent the radiotracer distribution, the areas of so-called 'fully 4D' image reconstruction and 'direct kinetic parameter estimation' are unified within a common framework. Many choices of linear and non-linear parameterization of these functions are considered (including the important case where the parameters have direct biological meaning), along with a review of the algorithms which are able to estimate these often non-linear parameters from emission tomography data. The other crucial components to image reconstruction (the objective function, the system model and the raw data format) are also covered, but in less detail due to the relatively straightforward extension from their corresponding components in conventional 3D image reconstruction. The key unifying concept is that maximum likelihood or maximum *a posteriori* (MAP) estimation of either linear or non-linear model parameters can be achieved in image space after carrying out a conventional expectation maximization (EM) update of the dynamic image series, using a Kullback-Leibler distance metric (comparing the modeled image values with the EM image values), to optimize the desired parameters. For MAP, an image-space penalty for regularization purposes is required. The benefits of 4D and direct reconstruction reported in the literature are reviewed, and furthermore demonstrated with simple simulation examples. It is clear that the future of reconstructing dynamic or functional emission



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

tomography images, which often exhibit high levels of spatially correlated noise, should ideally exploit these 4D approaches.

Keywords: image reconstruction, positron emission tomography, single photon emission computed tomography

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Positron emission tomography (PET) and single photon emission computed tomography (SPECT) provide a powerful array of capabilities for functional and molecular imaging in the human body, through detection of trace concentrations of radioactively-labeled compounds within any chosen volume of interest. Emission tomography reveals functional processes relevant to a range of fields which are of great importance to human health, such as neurology (Herholz *et al* 2004, Jones and Rabiner 2012), oncology (Muehllehner and Karp 2006) and cardiology (Mc Ardle *et al* 2012), and there is considerable versatility arising from the growing array of radioactively-labeled compounds (called radiotracers) which are in existence and also under development (Schirmacher *et al* 2013). Example radiotracers for PET include water ( $[^{15}\text{O}]\text{H}_2\text{O}$ ) for imaging blood flow, fluorodeoxyglucose ( $[^{18}\text{F}]\text{FDG}$ ) for imaging glucose metabolism, and neurotransmitter-specific ligands (e.g.  $[^{11}\text{C}]\text{raclopride}$ ) for imaging receptors in the brain. For SPECT, examples include  $[^{99\text{m}}\text{Tc}]\text{HMPAO}$  (exametazime) for imaging blood flow in the brain,  $[^{99\text{m}}\text{Tc}]\text{sestamibi}$  for myocardial perfusion imaging and  $[^{99\text{m}}\text{Tc}]\text{MDP}$  (methylene diphosphonate) for bone imaging. PET and SPECT are complementary to magnetic resonance imaging (MRI), and with the increasing establishment of simultaneous PET-MR imaging (Judenhofer *et al* 2007, Schlemmer *et al* 2008, Delso *et al* 2011), the powerful capabilities of PET are now becoming available in combination with the anatomical and functional imaging capacities of MR (such as functional MRI (fMRI)).

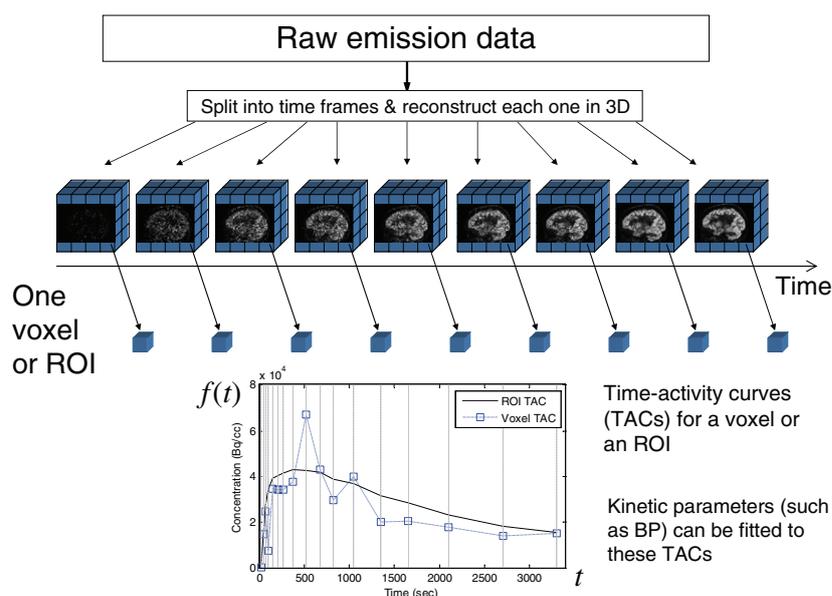
Both PET and SPECT usually involve intravenous administration of a radiotracer to the patient or research participant being studied, followed by external detection of the high-energy photons which are emitted from the distribution of the radiotracer within the body. The chemical sensitivity of PET and SPECT is very high (nanomolar or even picomolar concentrations of a radiotracer can be imaged and quantified), but there are however sensitivity limitations for detecting the high-energy photons, with often well below 5 percent of the total emitted photons being detected. Consequently, the number of photons detected by the scanners is typically in the range of  $10^4$  -  $10^9$  counts (depending on the study type and duration), which is low when compared to the count levels encountered in conventional non-dynamic transmission computed tomography (CT) for anatomical imaging.

It is the relatively limited number of photons detected in PET and SPECT, combined with the desire to reconstruct 3D images (with anything in the range of  $\sim 64^3$  to  $\sim 256^3$  image values), which means that *noise* in the 3D images is often, by far, the dominant concern for emission tomography. This has led not only to hardware developments to improve detection sensitivity, but also to developments in the algorithms used to reconstruct these images. It has now been amply demonstrated that the choice of reconstruction algorithm can have a significant impact on image quality (e.g. Comtat *et al* 1998), and extensive research over more than two decades into emission tomographic 3D image reconstruction methods has resulted in noise reduction and even resolution improvements (Rahmim *et al* 2013) via statistical iterative image reconstruction techniques (Leahy and Qi 2000). These iterative methods permit more accurate system and statistical modeling of the count-limited emission data to be incorporated into the reconstruction algorithm. Nonetheless, even with these advances in 3D reconstruction, the problem of

limited photon counts persists, and in particular is exacerbated when dynamic emission imaging is carried out, as the limited counts are divided among many different time frames (figure 1). In dynamic emission imaging the time course of the 3D radiotracer distribution after injection is sought, so that biologically useful parameters can be estimated from these time courses (such as blood flow, metabolic rates or binding potentials). Conventionally this involves independently reconstructing 3D time-frame images followed by post-reconstruction temporal modeling or kinetic fitting, in order to estimate kinetic and functional parameters within voxels (known as parametric imaging) or within regions of interest (ROIs). The noise in the 3D image frames can lead to unnecessarily high noise levels (and even bias) in the final fitted parameters, and sometimes the kinetic modeling can even fail. Ideally the noise distribution in the reconstructed 3D images, which is spatially-correlated and can also be object dependent, should be accurately accounted for in such post-reconstruction analyses. However, for the frequently-used iterative image reconstruction algorithms these noise distributions are rarely known precisely, being dependent on the number of iterations employed and other reconstruction parameters.

It is in this context that one of the recent surges of research interest in the field of reconstructing images from PET or SPECT data has occurred: the area of *fully 4D image reconstruction*. These 4D reconstruction approaches, which will be reviewed in this article, seek to reduce the aforementioned problems of limited counts in dynamic emission tomography by incorporating a model of the temporal behavior of the radiotracer directly into the image reconstruction algorithm. These models are usually chosen to provide physiologically-meaningful constraints within image reconstruction, and as a special case the kinetic model which is normally used on a post-reconstruction basis can itself be placed within the reconstruction algorithm. Often these temporal models use far fewer parameters to describe the time-course of the radioactive distribution compared to the number of time frames. For example, 3 coefficients or kinetic parameters can be used to describe a time-activity curve (TAC) instead of 26 distinct time frame samples. This reduction in the number of parameters to estimate, and the corresponding constraints the temporal models impose on the TACs during the image reconstruction process, deliver notable noise reduction in the time series of 3D images obtained. Nonetheless, it is important to note that the conventional independent-frame reconstruction approach is often followed by post-reconstruction fitting of the TACs which also provides noise reduction, but any benefits of TAC constraints during reconstruction are lost and furthermore accurate post-reconstruction modeling of the noise in the TACs can be difficult as previously mentioned. In contrast, if fully 4D reconstruction uses the kinetic model which is normally applied after reconstruction within the reconstruction, it has the advantage of estimating the end-point kinetic parameters directly from the raw data. This means the known noise-model of the raw emission data (which is often modeled extremely well as a space-time dependent Poisson process) is accurately accounted for when estimating the final parameters, sometimes offering variance and even bias reduction compared to post-reconstruction fitting of independently reconstructed images.

It is also worth mentioning at this point that in particular for SPECT imaging, fully 4D reconstruction offers the additional advantage of accounting for the changing activity distribution during a single rotation of the SPECT camera (Gullberg *et al* 2010). This is an important feature because whenever the activity distribution changes during the course of camera rotation, inconsistent projection data will be obtained (e.g. the projections taken at 0 degrees are not consistent with the projections taken a few moments later at 90 degrees, due to the dynamically changing activity distribution in the field of view (FOV)). Inconsistent projection data can lead to biases in the reconstructed images when the changing activity is not taken into account during the reconstruction. Emission tomographs which collect all projection data simultaneously (as is most typically the case with PET scanners) do not of course have this problem of inconsistent projections, but can nonetheless still benefit from the noise reduction available with 4D reconstruction.



**Figure 1.** Noise in dynamic 3D imaging (4D tomography). The raw emission data may be in list-mode format, or more conventionally stored as sinograms or projections for each time frame. A time series of 3D images is reconstructed, and time-activity curves (TACs) can be extracted for a voxel or region of interest (ROI), so that a kinetic model can be fitted to find functional parameters. Nine illustrative time frame images (out of 17, covering 60 min of a [ $^{18}\text{F}$ ]flumazenil PET scan) are shown along with the 17 point TACs. Central sagittal slices of every other frame (i.e. frames 1, 3, 5, ..., 17) of the 3D frame reconstructions are shown, demonstrating the elevated noise levels at the short time sampling intervals (the earlier, shorter frames, are noisier). The voxel-level TAC, due to noise, demonstrates an erratic time course of radiotracer concentration which is physiologically implausible. Depending on the study type, there may be up to 30 or more separate time frames, with frame durations ranging from a few seconds up to many minutes (in this figure the frame duration is represented by the width of the grey vertical lines).

## 2. Brief overview of statistical image reconstruction

To lay a foundation for the current review of 4D reconstruction in emission tomography, it is beneficial to briefly review the key concepts of statistical image reconstruction, which apply whether it is 2D, 3D, 4D or even 5D/6D reconstruction (5D and 6D applies when respiratory and/or cardiac gating is used in addition to dynamic 3D imaging, (e.g. Verhaeghe *et al* 2007)). We are concerned here with statistical image reconstruction, rather than analytic reconstruction methods such as filtered backprojection. Analytic methods, generally speaking, have less accurate models of the imaging system than statistical image reconstruction, but the noise in the reconstructed images can be relatively easily calculated (e.g. Watson 2009). There are five key components to be considered for statistical image reconstruction: i) the parameters to be estimated, ii) the form of the acquired data, iii) the forward model, iv) the objective and v) the algorithm. These five components are implied in the 3D reconstruction review by Lewitt and Matej (2003).

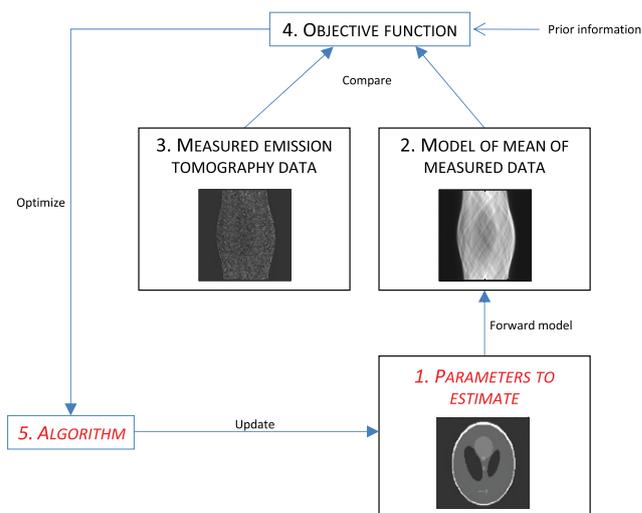
At its core, all image reconstruction is concerned with spatially-localized parameter estimation, so it is important to start with a clear definition of the *parameters* which are being sought. Once the desired parameters have been chosen, the format of the *acquired raw PET or SPECT data* needs to be carefully considered, as in general the required parameters do not

directly relate in a one-to-one fashion with the measured data. A typical choice for the parameters is the mean radioactive concentration in each voxel in the FOV of the scanner, whereas instead the measurements are a set of sinograms, projections or list-mode data. We therefore need to describe the mapping from the parameters of interest (e.g. the mean radioactive concentration in each voxel) to the space in which the acquired data were acquired. This mapping, which should accurately capture the physics of radioactive emissions as well as the scanner's response, is often broadly referred to as the *forward model*, or the system matrix, as will be seen further below. A simple example is the discrete 3D x-ray transform (Defrise *et al* 1995), which takes line integrals through an object represented by voxels to deliver sets of 2D parallel projections (so the discrete x-ray transform maps voxel values to projection data values). Having clearly defined the parameters which are sought, and how these map to the acquired data format, we then need to decide on the *objective function* (also known as the cost function). The objective function defines the way in which we do, or do not, require the estimated parameters to agree with the measured data. For example, one might require the forward projection of the voxel values to agree with the measured sinograms by requiring the sum of the square difference between each forward projected value and each measured value to be minimized (a least squares objective). Finally an *algorithm* is needed which is capable of generating estimates of the desired parameters which are in accordance with the chosen objective function.

All of these aspects, summarized in figure 2, will be explored in detail in turn, with a specific focus on the topic of 4D image reconstruction and the current state of the art in the field, for both PET and SPECT imaging. As will be seen however, the crucially different components needed in 4D reconstruction primarily relate to the *parameters to be estimated*, and the *algorithms used to achieve this*. Therefore the primary focus will be on those two core components in reconstruction, with less emphasis on the objective, data and system model, as these latter aspects remain relatively unchanged for 4D reconstruction and hence the current reviews of 3D statistical image reconstruction (such as those by Lewitt and Matej (2003), Qi and Leahy (2006) and Reader and Zaidi (2007)), already provide excellent coverage. Note further that there are many 4D data processing methods, which in a strict sense are not really 4D reconstruction methods at all (e.g. preprocessing of sinogram data through use of principal components analysis (Razifar *et al* 2006), factor analysis and related techniques (Wernick *et al* 1999), and filtering methods (e.g. Turkheimer *et al* 2003, McLennan and Brady 2009)) and so are not the focus of this review of 4D reconstruction. Such alternative 4D data processing methods are already reviewed by others in previous reviews of 4D reconstruction (see Tsoumpas *et al* 2008a, Rahmim *et al* 2009).

Finally, many motion correction strategies also fall under the category of 4D or higher dimensional image reconstruction, but the vastly growing volume of work on motion-corrected and/or gated reconstruction (e.g. Li *et al* 2006) is beyond the scope of this review. Nonetheless, genuinely 4D reconstruction methods with simultaneous motion estimation (i.e. not ad hoc inter-reconstruction or post-reconstruction motion estimation methods) are now emerging (e.g. Jacobson and Fessler 2006, Grotus *et al* 2009).

One final point to note for this section is the acquired data format. Emission tomography reconstruction algorithms have often required the data to be in sinogram or projection format, whereby each event (a pair of coincidence photons in PET, or a single photon in SPECT) is histogrammed into a projection bin, with each projection data bin approximately associated with a line of response (LOR) through the scanner FOV. However, with increasing information now being measured for each event (e.g. photon detection positions, arrival times and energies), the use of projection data format becomes increasingly sub-optimal. As an example, for dynamic imaging, an entire projection data set is needed for each time frame (a time sampling interval), so that not only is considerable storage required for multiple time frames,



**Figure 2.** The five core components in image reconstruction, shown in five boxes. The parameters are used to generate a representation of the emitting radioactive distribution in the FOV, which is then forward modeled to deliver a model of what the mean measured data would be for such a distribution. The objective function assesses the agreement of the noisy measured data with this model of the mean data (possibly also assessing agreement with prior knowledge), and an algorithm then updates the parameters in a way which will lead to an improvement in the agreement required by the objective function. Components 1 and 5 (in italics) require significant changes compared to 3D reconstruction in order to achieve fully 4D (and ‘direct’) image reconstruction.

but furthermore there is loss of information as events within a certain time sampling interval are grouped together. The same applies spatially: events can be grouped into the same projection bin through use of axial spans (Fahey 2002), leading again to information loss. For this reason list-mode data can be both more efficient and more accurate (see e.g. Parra and Barrett 1998, Reader *et al* 1998), whereby each event with all its attributes (position, time, energy, and for example the time of flight (TOF) estimate for PET, or for example the gantry position for rotating SPECT systems) is simply recorded sequentially in a file. Both dynamic sinogram data and list-mode data can be used for 4D image reconstruction.

### 3. Representations of radioactivity distributions in 3D and 4D

#### 3.1. 3D case

At the heart of 3D reconstruction is a parameterized model of the spatial distribution of the radioactive concentration in the scanner FOV. A typical example in PET imaging would be the distribution of the radiotracer [ $^{18}\text{F}$ ]FDG in the human brain. One can model, or represent, the 3D radioactive distribution as a function  $f$  of position  $\mathbf{r} = [x \ y \ z]^T$  in the FOV by summing a limited number of  $j=1 \dots J$  spatial functions  $b_j(\mathbf{r})$ :

$$f(\mathbf{r}; \boldsymbol{\theta}) = \sum_{j=1}^J b_j(\mathbf{r}; \boldsymbol{\theta}_j) \quad (1)$$

where each function is parameterized by its own unique parameter vector  $\boldsymbol{\theta}_j$ , consisting of  $p=1 \dots P$  parameters. In general, each spatial function  $b_j(\mathbf{r}; \boldsymbol{\theta}_j)$  may itself also consist of one or

more functions of linear or non-linear parameters. The overall parameter vector  $\boldsymbol{\theta}$  on the left hand side of equation (1) is therefore a  $JP$ -dimensional vector, composed of  $P$ -dimensional subvectors  $\boldsymbol{\theta}_j$ , one for each spatial function  $b_j$ . Image reconstruction then involves estimating  $P$  parameters for each one of the  $J$  spatial functions in (1).

The common practice in 3D emission tomography is to predefine the precise size, shape and position of the spatial functions used in equation (1) before image reconstruction, and we will refer to this fixed collection as a set of  $j=1\dots J$  spatial basis functions  $\phi_j(\mathbf{r})$ . This leaves image reconstruction with the task of only finding a scaling parameter (i.e. the coefficient)  $\theta_j$  for each of the predefined functions  $\phi_j(\mathbf{r})$ . In this case, with only one parameter per spatial function,  $P=1$ , equation (1) simplifies to

$$f(\mathbf{r}; \boldsymbol{\theta}) = \sum_{j=1}^J \theta_j \phi_j(\mathbf{r}) \quad (2)$$

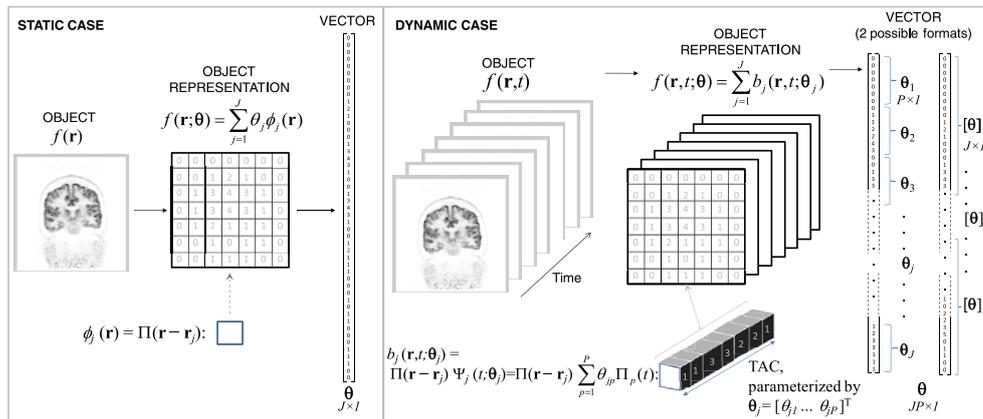
An example of equation (2) would be to predefine a set of various volumes or regions of interest, and to use these as the spatial basis functions. However, by far the most widely-used case is to simplify still further, by using predefined functions that are shifted copies of each other:

$$f(\mathbf{r}; \boldsymbol{\theta}) = \sum_{j=1}^J \theta_j \phi(\mathbf{r} - \mathbf{r}_j) \quad (3)$$

where the dependency of  $\phi$  on  $j$  is no longer present, as the uniqueness of each spatial basis function  $\phi$  is now only its particular fixed position in 3D space, specified by  $\mathbf{r}_j$ . In equations (1)–(3) the spatial functions can possibly overlap with one another in space, but the typical choice of  $\phi$ , as illustrated in figure 3, is the pixel (for 2D reconstruction) or the voxel (for 3D reconstruction). For these cases  $\phi$  is a 2D or 3D rectangle function:

$$\phi(\mathbf{r} - \mathbf{r}_j) = \Pi(\mathbf{r} - \mathbf{r}_j) \quad (4)$$

with positions  $\mathbf{r}_j$  chosen to avoid overlap of any of the  $J$  functions. With this choice of spatial basis function the overall representation of the radioactive distribution is in effect captured by the vector of parameters  $\boldsymbol{\theta}$ , merely a list of numbers which are arranged for display purposes into a 3D array of values to form a grey scale image (see again figure 3). Whilst the pixel or voxel are common, many other spatial basis functions have been proposed, as of course there are infinitely many 3D object representations available. Some examples include spherically symmetric basis functions ('blobs') (Lewitt 1992), or as mentioned, choosing regions of interest (ROIs). If one knows *a priori* the precise regions and locations in the radioactive distribution which are of interest, then these can be used as spatial basis functions, for direct estimation of the radioactive concentration contained in those regions (Carson 1986). Note that even for that case, just as for 3D voxel or 'blob' functions, just one coefficient parameter  $\theta_j$  is required for each spatial basis function, as the model of equation (2) still applies. Equations (2)–(4) cover the majority of models in the 3D reconstruction literature, and these are all examples of *linear parameter estimation*, as the parameters being estimated are merely coefficients which scale the amplitudes of predefined basis functions. However, more generally, one should not insist on such a simplification, and instead use equation (1) which allows parameters to be estimated which not only specify the coefficient (amplitude) of the spatial basis functions, but which also *specify the shape* of the spatial functions. This generalization will help a great deal with a unified understanding of 4D image reconstruction in emission tomography, as in 4D tomography choosing more general temporal functions has great utility. To illustrate the point for 3D spatial functions, equation (1) could be used to describe the 3D



**Figure 3.** Left: 2D object representation using a  $J \times 1$  vector of parameters  $\boldsymbol{\theta}$  in conjunction with a choice of basis function  $\phi(\mathbf{r})$ . In the example static case a 2D rectangle function (pixel) is used as the spatial basis, and each pixel has a coefficient  $\theta_j$  to represent the level of radiotracer concentration within that pixel-sized region of the FOV. Right: for the dynamic 2D case a 3D rectangle function is used as an example spatiotemporal basis (pixel with extension into time as a third dimension). In this case a  $JP \times 1$  vector  $\boldsymbol{\theta}$  is needed, which can be organized as either i) being composed of subvectors  $\boldsymbol{\theta}_j$ , with each subvector containing  $P$  parameters (in this example  $P$  matches the number of time frames  $T$ , and  $\boldsymbol{\theta}_j$  can be interpreted directly as a TAC for pixel  $j$ ) or ii) being composed of subvectors  $[\boldsymbol{\theta}]_p$ , with each subvector being a parametric image (in the example each such subvector is a time-frame image composed of  $J$  pixels). The illustration extends naturally to the case of voxels and dynamic 3D imaging. NB: a simplification which is widespread in the literature will be adopted after section 3 in this article: voxels will be used as the spatial basis functions, so that  $J=V$ , where  $V$  is the number of voxels.

radioactive spatial distribution as a collection of 3D Gaussian functions, each function having  $P=5$  parameters contained in the subvector  $\boldsymbol{\theta}_j = [\theta_{j1} \theta_{j2} \theta_{j3} \theta_{j4} \theta_{j5}]^T$ :

$$b_j(\mathbf{r}; \boldsymbol{\theta}_j) = \theta_{j1} \exp\left(-\frac{(x - \theta_{j2})^2 + (y - \theta_{j3})^2 + (z - \theta_{j4})^2}{\theta_{j5}}\right) \quad (5)$$

Use of equation (5) means we are no longer seeking to estimate just the coefficient  $\theta_{j1}$  of the spatial function, but furthermore its position  $[\theta_{j2} \theta_{j3} \theta_{j4}]^T$  in the 3D FOV along with its width (proportional to the variance)  $\theta_{j5}$ . In static image reconstruction in 2D and 3D, the generality of equation (1), with the example case of equation (5), has remained largely ignored. However, for 4D reconstruction where such a generalization has great utility in the temporal domain, this is far from the case.

A final point to note for this section is that any given representation of the distribution of the radiotracer will have approximations compared to the unknown true distribution. Consideration of the issues involved with representations of functions of continuous variables is dealt with rigorously and at length in Barrett and Myers (2004), and an excellent overview of 3D spatial basis functions for emission tomography reconstruction is contained in Lewitt and Matej (2003).

### 3.2. 4D case

The aforementioned generalized representations for 3D reconstruction help to clarify the current state of the art in 4D image reconstruction for emission tomography. First, we will extend the general 3D object representation equation (1) to now include the time dimension:

$$f(\mathbf{r}, t; \boldsymbol{\theta}) = \sum_{j=1}^J b_j(\mathbf{r}, t; \boldsymbol{\theta}_j) \quad (6)$$

In equation (6) the time course, or the dynamics, of the 3D radioactive concentration in the scanner FOV is now represented through the use of a 4D function  $f(\mathbf{r}, t)$ , which is a superposition of  $j=1 \dots J$  spatiotemporal functions, each of which has  $P$  parameters stored within a subvector  $\boldsymbol{\theta}_j$  of  $\boldsymbol{\theta}$ . These functions can in general overlap in space and time. Unlike for the 3D case in the previous section, it is far more common to use more than one parameter ( $P > 1$ ) for a given function  $b_j$ . It is convenient, according to the context, to order and access the elements of  $\boldsymbol{\theta}$  by considering either subvectors  $\boldsymbol{\theta}_j$  or subvectors  $[\boldsymbol{\theta}]_p = \{\boldsymbol{\theta}_{jp}, j = 1 \dots J\}$  (see for example figure 3, where these two cases correspond to TACs and image frames respectively). A simplification of equation (6), assumed by the *entire* current literature on 4D emission tomography reconstruction in PET and SPECT, is to first factorize each spatiotemporal function into a fixed choice of spatial basis function  $\phi_j(\mathbf{r})$  in product with a temporal function  $\Psi_j(t; \boldsymbol{\theta}_j)$ .

With the factorization we have:

$$f(\mathbf{r}, t; \boldsymbol{\theta}) = \sum_{j=1}^J \phi_j(\mathbf{r}) \Psi_j(t; \boldsymbol{\theta}_j) \quad (7)$$

The spatial basis function  $\phi_j(\mathbf{r})$  is now scaled by the temporal function  $\Psi$  (instead of the parameter  $\theta_j$  in equation (2)), and hence the parameterization of the spatiotemporal function is now incorporated exclusively into the definition of  $\Psi(t; \boldsymbol{\theta}_j)$ , with  $\phi_j(\mathbf{r})$  merely indicating the spatial volume occupied by  $\Psi$ . If the spatial basis functions do not overlap, then the temporal function  $\Psi$  is in fact the TAC for that given spatial volume. If the spatial basis functions do overlap, then the TAC at a given sub-volume in the FOV can be found by summing over all the overlapping contributions, as will be shown below. With the factorizing simplification of equation (7), all 4D reconstruction methods in the literature proceed to focus exclusively on a specific parameterization of the temporal function  $\Psi$ . The parameterization of  $\Psi(t; \boldsymbol{\theta}_j)$  with  $P$  parameters can be written more explicitly as:

$$\Psi_j(t; \boldsymbol{\theta}_j) = \Psi_j(t; \theta_{j1}, \theta_{j2}, \dots, \theta_{jP}) \quad (8)$$

The goal will be to estimate these  $P$  parameters for each and every temporal function indexed by  $j$ , to find the whole  $JP$ -dimensional parameter vector  $\boldsymbol{\theta}$ . With this vector, one can then form the 4D radioactive distribution via equation (7).

For representation in computer memory the 4D continuous model of equation (7) will be discretized and the following vector notation can be used

$$[\mathbf{f}(\boldsymbol{\theta})]_{vt} = f_{vt}(\boldsymbol{\theta}) = \sum_{j=1}^J [\boldsymbol{\phi}]_{jv} [\boldsymbol{\Psi}(\boldsymbol{\theta}_j)]_{jt} = \sum_{j=1}^J \phi_{jv} \Psi_{jt}(\boldsymbol{\theta}_j) \quad (9)$$

where the vector  $\mathbf{f}$  is a time-series ( $t=1 \dots T$  frames) of 3D images ( $v=1 \dots V$  voxels) contained in a  $VT$ -dimensional vector  $\mathbf{f}$ . Each element of  $\mathbf{f}$  is a sample of the 4D continuous radioactivity concentration modeled by (7), averaged over the time interval represented by time frame index  $t$ , and averaged over the sub-volume represented by voxel index  $v$ . The spatial functions  $\phi$  may of course not be voxels, and so the  $JV$ -dimensional vector  $\boldsymbol{\phi}$  contains all the  $v=1 \dots V$  voxel-volume samples of each of the  $j=1 \dots J$  spatial basis functions  $\phi_j(\mathbf{r})$ . The  $JT$ -dimensional vector  $\boldsymbol{\Psi}$  contains the  $t=1 \dots T$  time-frame samples of all the  $j=1 \dots J$  temporal functions  $\Psi_j(t; \boldsymbol{\theta}_j)$ .

With this discretization scheme both  $\mathbf{f}$  and  $\boldsymbol{\Psi}$  can be linear or non-linear functions of  $\boldsymbol{\theta}$ , but when  $\mathbf{f}$  is evaluated and sampled for all  $v=1 \dots V$  and all  $t=1 \dots T$  for a particular parameter vector  $\boldsymbol{\theta}$ , the time

series of images contained in  $\mathbf{f}$  can then be operated on by a conventional linear forward model (system matrix), irrespective of whether a linear or non-linear parameterization of  $\mathbf{f}$  was used. For this reason we have deliberately defined equation (9) in terms of vectors rather than matrices, in order to preserve the use of matrices for system modeling in section 5 below. Sampling  $\mathbf{f}$  prior to applying the system model may be sub-optimal for some choices of  $\phi$  (such as ‘blobs’), but this pre-discretization approach is highly practical for implementation of 4D image reconstruction algorithms.

One final point to note is the special simplified case used in the majority of the 4D literature, where non-overlapping pixels or voxels are used for  $\phi$ , such that each voxel-volume sample  $v$  corresponds directly to a single basis function  $\phi_j(\mathbf{r})$ . This means that the indices  $v$  and  $j$  are equivalent,  $J$  and  $V$  are equivalent, and  $\phi=1$  within the voxel volume. Therefore the temporal function  $\Psi$  can be directly interpreted as the TAC in a given voxel (where a voxel can be labeled by  $v$  or  $j$ ):

$$f_v(\boldsymbol{\theta}) = f_{jt}(\boldsymbol{\theta}_j) = \Psi_{jt}(\boldsymbol{\theta}_j) \quad (10)$$

With this direct association of a unique TAC modeled by  $\boldsymbol{\theta}_j$  with each unique voxel  $j$ , we can obtain a set of  $P$  parametric images, with each parametric image given by the subvector  $[\boldsymbol{\theta}]_p = \{\theta_{jp}, j=1 \dots J\}$ . Thus estimation of the  $JP$ -dimensional (or equivalently  $VP$ -dimensional) vector  $\boldsymbol{\theta}$  can also be referred to as *parametric image reconstruction*. For the remainder of this review we will focus our attention on this simplified and widely-used case of employing non-overlapping voxels.

## 4. Parameterization of 4D radioactivity distributions

### 4.1. Linear parameterization

A common choice of parameterization of the temporal function  $\Psi$  is the linear case, using the parameters  $\boldsymbol{\theta}$  to define scale factors. The parametric images are therefore coefficient images,  $[\boldsymbol{\theta}]_p$ , one for each of the  $p=1 \dots P$  temporal basis functions. The conventional case of separate time-frame 3D reconstructions corresponds to using temporally-shifted rectangle functions  $\Pi(t)$  as the basis functions:

$$\Psi_j(t; \boldsymbol{\theta}_j) = \sum_{p=1}^P \theta_{jp} \Pi_p(t) \quad (11)$$

In equation (11) there are the same number of parameters ( $P$ ) as there are total time frames ( $T$ ), and the  $p=1 \dots P$  rectangle basis functions each cover unique non-overlapping time sampling intervals (see table 1), with start and end times such that the entire time duration of the scan is covered. The start and end times of each rectangle function are chosen according to the expected temporal dynamics of the emission data, and it is common to have up to 20 or 30 such frames to cover 60 min of data.

Less trivial temporal basis functions can be chosen to describe the temporal function, but still using a linear parameterization:

$$\Psi_j(t; \boldsymbol{\theta}_j) = \sum_{p=1}^P \theta_{jp} B_p(t) \quad (12)$$

This linear case, equation (12) combined with equation (7), actually covers a significant portion of the current 4D image reconstruction literature. The key difference between a large number of the 4D reconstruction methods is in the precise choice of the temporal basis functions  $B(t)$  (e.g. from splines (Nichols *et al* 2002) through to basis functions related to radiotracer kinetics (Wang *et al* 2008)) along with the number  $P$  of these functions to use.

**Table 1.** Review of choices of temporal functions  $B(t)$  for 4D reconstruction which require only a *linear parameterization* (simple scale factor for each function), along with how many such functions are used (i.e. the number of parameters,  $P$ , per temporal function  $\Psi$ ).  $T$  is the number of time frames one seeks to reconstruct.

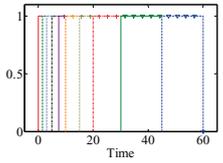
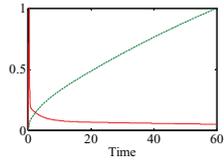
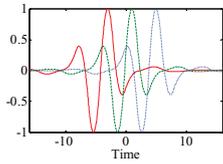
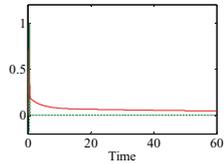
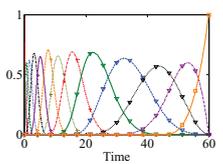
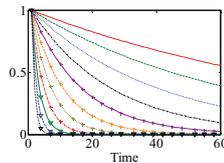
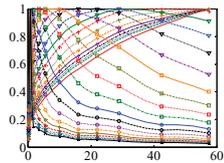
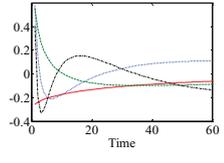
Temporal basis $B(t)$ and number, $P$ , used	Comments, references	Temporal basis $B(t)$ and number, $P$ , used	Comments, references
<p><i>Rectangle functions, <math>P = T</math></i></p> 	<p>No physiological meaning, no noise-reduction.</p> <p>Standard independent frame 3D reconstruction.</p>	<p><i>Patlak basis functions, <math>P = 2</math></i></p> 	<p>Finding the coefficients provides direct estimates of the functional parameters <math>\theta_1 = c</math> and <math>\theta_2 = K_i</math>.</p> <p><i>Input function <math>C_{PL}(t)</math> needed.</i></p> $B_1 = C_{PL}(t); B_2 = \int_0^t C_{PL}(t') dt'$ <p>(Wang <i>et al</i> 2008), (Tsoumpas <i>et al</i> 2008b), (Tang <i>et al</i> 2008, 2010), (Angelis <i>et al</i> 2011b)</p>
<p><i>Wavelets, <math>P &lt; T</math></i></p> 	<p>No physiological meaning, noise-reduction only.</p> <p>(Verhaeghe <i>et al</i> 2008)</p>	<p><i>Logan basis functions, <math>P = 2</math></i></p> 	<p>Provides direct estimates of the functional parameters <math>\theta_1 = DV</math> and <math>\theta_2 = c</math>, or <math>\theta_1 = DVR</math> if a reference region rather than input function used.</p> $B_1 = C_{PL}(t); B_2 = -\frac{d}{dt}C_{PL}(t)$ <p>(Rahmim <i>et al</i> 2012)</p>
<p><i>Splines, <math>P &lt; T</math></i></p> 	<p>No physiological meaning, noise-reduction only.</p> <p>(Nichols <i>et al</i> 1999), (Reutter <i>et al</i> 2000), (Nichols <i>et al</i> 2002), (Li <i>et al</i> 2007), (Verhaeghe <i>et al</i> 2007)</p>	<p><i>Decaying exponential basis functions, <math>P &gt; T</math></i></p> 	<p>Predetermined logarithmically-spaced decay constants.</p> <p>(Hebber <i>et al</i> 1997)</p>
<p><i>Data-derived basis functions, <math>P &lt; T</math></i></p>	<p>No physiological meaning, noise-reduction only. Based on non-negative matrix factorization.</p> <p>(Reader <i>et al</i> 2006a), (Sitek <i>et al</i> 1999, 2000, 2001).</p>	<p><i>Spectral-analysis basis functions, <math>P &gt; T</math></i></p> 	<p>Uses the input function or a data-derived surrogate generating function.</p> <p>Can be implemented as an alternating <i>joint estimation</i> of coefficients and the generating function.</p> <p>(Reader <i>et al</i> 2007), (Verhaeghe <i>et al</i> 2010), (McLennan and Brady 2010)</p>
<p><i>Singular vectors from a filtered backprojection reconstruction, <math>P &lt; T</math>, or, ROI TACs from population data, <math>P &lt; T</math></i></p>	<p>First case has no physiological meaning.</p> <p>(Matthews <i>et al</i> 1997)</p>	<p><i>Orthogonalized exponential basis (spectral analysis)</i></p> 	<p>First few left singular vectors of the exponential functions convolved with the input function (singular vectors shown only).</p> <p>(Maltz 2000) (Maltz 2001)</p>

Table 1 provides an overview of the 4D image reconstruction literature for this very popular linear parameterization case of using (12) in conjunction with (7).

Considering table 1 there is an important distinction of two categories of linear temporal basis functions: those that do, and those that do not have a direct link with biological/functional parameters of interest to medical imaging. The cases that do not have such an interpretation can only really be viewed as a means of limiting noise in the temporal data by imposing correlations between the various time points in the data, and after reconstruction there will still often be the need to perform a kinetic analysis to estimate actual biological parameters of interest, such as the distribution volume (DV), binding potential (BP) or the tracer influx rate ( $K_i$ ). Reviews of these biological and functional parameters of interest are given in Bentourkia (2007) and in Innis *et al* (2007), and are not the focus of this present review. Note that post-reconstruction estimation of kinetic parameters can result in bias issues if the reconstruction algorithm does not permit negative values (Reilhac *et al* 2008, Verhaeghe and Reader 2010).

If the basis functions are just a means of reducing noise, then one could also consider the alternative approach of limiting noise through appropriate definition of the objective function (e.g. using temporal roughness penalties) for the reconstruction task, as will be covered in section 6 below. It is an interesting observation that for 3D reconstruction the use of different objective functions (which include a spatial roughness penalty, (e.g. Hebert and Leahy 1989)) has proven to be a more popular approach to noise reduction than changing the spatial basis functions (e.g. use of blobs, sieves (Snyder and Miller 1985) and even ROIs). In contrast, the current 4D literature reveals the opposite trend: relatively little emphasis on temporal penalties, with instead a large emphasis on changing the temporal basis functions. Conceivably, the best approach would be to combine spatiotemporal penalties in the objective function with an appropriate choice of parameterized spatiotemporal functions, but this remains an interesting area of research which has not really been addressed in the literature thus far.

Returning to the case of basis functions which do have a direct functional / biological interpretation, there are three main choices of temporal basis functions with a linear parameterization, as shown in table 1. These correspond to the cases of Patlak, Logan and spectral analysis, as will be described in detail below. When such functions are used one is implicitly carrying out *direct kinetic parameter estimation*. This means the parameters found by the reconstruction algorithm directly from the raw emission data will be the end-point biological measures of interest, without the need for any further processing (unlike the case for merely noise-reducing basis functions). Here the word 'direct' means the avoidance of an additional fitting stage after reconstruction (post-reconstruction processing often introduces approximations in modeling the noise in the images), it does not mean the estimation is non-iterative.

An important component used by these functionally meaningful basis functions is an input function, or alternatively a reference function. The input function, which will be denoted by  $C_{PL}(t)$  in this review, refers to the concentration of free and intact radiotracer in the arterial plasma, as a function of time. It is of fundamental importance as one normally needs to know both the input  $C_{PL}(t)$  to a system (the system being biological tissue in our case) as well as the output from the system (which is the TAC  $\Psi(t)$  in our case), if any information about the system itself is to be inferred. The presence of the radiotracer in the arterial plasma provides the input supply for tissue to be able to take up the tracer. The tissue will then either allow the tracer to be readily returned to the blood, bind the tracer (reversible binding) or somehow (e.g. metabolically) trap the tracer in the tissue (irreversible binding). However, measuring the input function is an involved and invasive process, and for this reason reference tissue methods (Lammertsma and Hume 1996, Gunn *et al* 1997) have been devised. In this latter case, the uptake of radiotracer in a tissue of interest is assessed relative to a reference tissue, so as to infer functional / biological parameters of interest in the tissue, provided that the behaviour of

the reference tissue is indeed a good reference (e.g. provided that it is known *a priori* that the reference tissue contains little or no capacity for any form of binding of the tracer, whilst at the same time having the same degree of non-specific binding as the target tissue of interest).

**4.1.1. Patlak basis functions.** For the case of modeling irreversible uptake of the radiotracer into a tissue, Patlak devised a linear model (Patlak *et al* 1983, Gjedde 1995) relating any time activity curve  $C(t)$  to the tracer influx constant  $K_i$ :

$$\frac{C(t)}{C_{PL}(t)} = K_i \frac{\int_0^t C_{PL}(t') dt'}{C_{PL}(t)} + c \text{ for } t > t_{ss} \quad (13)$$

where the format of equation (13) is designed to allow a conventional linear regression with slope  $K_i$  and offset  $c$ . In effect the TAC can be modeled by our parameterized temporal function  $\Psi$  as the sum of just two basis functions (Tsoumpas *et al* 2008b):

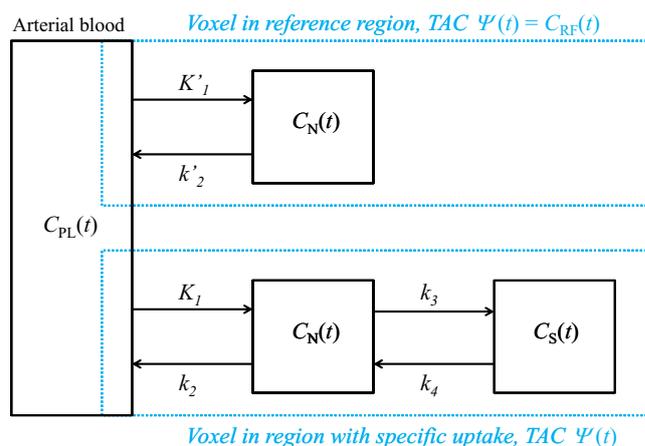
$$\Psi_j(t; \theta_j) = \sum_{p=1}^2 \theta_{jp} B_p(t) = \theta_{j1} C_{PL}(t) + \theta_{j2} \int_0^t C_{PL}(t') dt' \text{ for } t > t_{ss} \quad (14)$$

where  $t_{ss}$  is the time at which the steady state is reached,  $C_{PL}(t)$  is the input function, and the two coefficients,  $\theta_{j1}$  and  $\theta_{j2}$ , of the two basis functions, can be relabeled as  $c$  and  $K_i$  respectively. The radiotracer net influx rate constant  $K_i$  has particular significance: for example when the PET radiotracer [ $^{18}\text{F}$ ]FDG is used, it is proportional to the glucose metabolic rate. In table 1 it is shown that, to date, at least four distinct research groups have implemented direct estimation of  $K_i$  through use of these Patlak basis functions (Wang *et al* 2008, Tsoumpas *et al* 2008b, Tang *et al* 2010, Angelis *et al* 2011b).

**4.1.2. Logan basis functions.** The Patlak approach only applies to radiotracers with essentially irreversible uptake in tissue. A similar approach to Patlak was devised by Logan (Logan 2000) for reversible uptake of tracers (e.g. radioligands binding to neuroreceptors). However, in its original form the Logan method does not constitute a linear model for a TAC. Nonetheless, linear forms based on the Logan approach have been devised and used as basis functions to permit direct kinetic parameter estimation (Wang and Qi 2013)

$$\Psi_j(t; \theta_j) = \theta_{j1} C_{PL}(t) + \theta_{j2} \left[ -\frac{d}{dt} C_{PL}(t) \right] \text{ for } t > t_{ss} \quad (15)$$

where the derivative of the input function  $C_{PL}(t)$  after the steady state time  $t_{ss}$  is negative, and so the right most term in parentheses is a positive basis function, such that  $\theta_{j1}$  and  $\theta_{j2}$  can be directly interpreted as the distribution volume (DV) and the negative of the intercept  $c$  of a Logan plot respectively. The DV has a useful physiological interpretation as being the level to which tissue concentrates the radiotracer relative to the blood at equilibrium (so a DV of 2 means the tracer is twice as concentrated in the tissue as in the blood). This is pertinent for imaging of receptors in the brain, and the model finds use for example in PET, in modeling the binding of [ $^{11}\text{C}$ ]raclopride to available  $\text{D}_2/\text{D}_3$  receptors, so that images of DV allow one to infer the levels of endogenous dopamine (which competes with raclopride for binding to  $\text{D}_2/\text{D}_3$  receptors). The one example in the literature of direct use of Logan-derived basis functions for direct reconstruction is Rahmim *et al* (2012), but in that work an integrated form of equation (15) is used.



**Figure 4.** Kinetic modeling in emission tomography, using compartmental models.  $C_N(t)$  and  $C_S(t)$  are the concentrations of the tracer in the non-specifically bound / non-displaceable (N) compartment and the specifically bound (S) compartment of the tissue respectively. See table 3 for definitions of all other terms.

4.1.3. *Spectral analysis basis functions.* A widespread model for the uptake of a radiotracer into tissue for PET and SPECT is that of compartmental modeling (see figure 4). By considering the transfer of the radiotracer between the blood and the tissue, and transfer between bound and unbound states within the tissue, one can devise a set of differential equations describing the time course of radiotracer concentration in the tissue. The solution to these equations turns out to be a convolution of the input function with a sum of exponentials (with the number of exponential functions related to the number of compartments in the model (Gunn *et al* 2001)).

The so-called spectral analysis model (Cunningham and Jones 1993) exploits this solution by using a set of exponential functions which cover all plausible tissue behaviours. The method makes no assumptions of reversible or irreversible uptake, and models the TAC as a large collection of basis functions, each of which is an exponential with a different decay constant convolved with the input function:

$$\Psi_j(t; \theta_j) = \sum_{p=1}^P \theta_{jp} [ C_{PL}(t) \otimes \exp(-\beta_p t) ] \tag{16}$$

where  $\otimes$  denotes convolution. From this it is easy to compute  $K_1$  as simply the sum over all the coefficients of the basis functions, with DV found by summing the coefficients with each one divided by its respective decay constant  $\beta$ . If appropriate  $K_i$  can also be found (see for example Meikle *et al* (1998)).

4.2. *Non-linear parameterization*

Returning to equation (8), we can of course also consider a non-linear parameterization of the temporal function  $\Psi$ , which is very useful for kinetic modeling, as many functional / biological parameters of interest are in fact non-linear. The most commonly encountered kinetic models for the dynamically changing radiotracer distribution at a given location in the scanner FOV are shown in figure 4. All these models are captured by one general expression:

$$\Psi_j(t; \theta_j) = \theta_{j1} C_1(t) + \theta_{j2} C_2(t) \otimes h(t; \theta_{j3}, \theta_{j4}, \dots, \theta_{jp}) \tag{17}$$

Just as equation (12) expressed a general form of spatiotemporal basis which covers a significant portion of the 4D reconstruction literature (linear parameterization), so also equation (17) expresses a general form which covers the vast majority of the remaining cases (non-linear parameterization) in the 4D reconstruction literature.

Equation (17) indicates that the overall TAC at a position in the scanner FOV is composed of a spatially-invariant TAC  $C_1(t)$  (which in all known models is either the time-course of radiotracer concentration in the whole blood  $C_{WB}(t)$  or else a reference function  $C_{RF}(t)$ ) added on to a convolution of a second TAC  $C_2(t)$  convolved with a tissue response function  $h(t)$ . This second TAC, in all known models, is either the input function  $C_{PL}(t)$  or the reference function  $C_{RF}(t)$ . The tissue response function in essentially all currently devised models is either one exponential or a sum of exponential functions. Table 2 summarizes the literature for the important case of equation (17) (with table 3 providing definitions of terms), showing the models used for each spatially localized TAC in the FOV. Figure 5 shows an example TAC from this model.

#### 4.3. Special cases of TAC models

**4.3.1. Joint estimation methods (linear models).** An approach which strictly comes under the category of using basis functions to denoise the reconstruction, but which is inspired by the physiologically known processes of radiotracer uptake, is to jointly estimate both the coefficients for a set of temporal basis functions as well as a discrete function which resembles the input function.

The model is inspired by spectral analysis (see previous section 4.1.3.), by predefining a fixed number of  $p=1 \dots P$  exponential basis functions, each with a fixed decay constant  $\beta_p$ , where for each exponential basis function we need to find a coefficient image  $[\theta]_p$ . However, in addition, we need to find a discretely sampled ‘generating function’ (which can be viewed as a surrogate for the unknown input function), which is defined by its values at a fixed number  $N$  of time points. Hence we need to estimate  $P$  parameters for each voxel, giving a total of  $JP$  parameters to find, as well as an additional  $N$  parameters (the generating function, just as the input function, is assumed to be the same for all voxels). Therefore, the TAC for a voxel is modeled by:

$$\Psi_j(t; \theta_j) = \sum_{p=1}^P \theta_{jp} \left[ C_{GF}(t; \theta^*) \otimes \exp(-\beta_p t) \right] \quad (18)$$

where  $C_{GF}(t)$  is the generating function (parameterized by  $N$  values contained in the parameter vector  $\theta^*$ ), as described in Verhaeghe and Reader (2013), based on the idea of Reader *et al* (2007). The approach is to alternate between estimation of the coefficient images and estimation of the generating function, by holding the generating function constant and then the coefficients constant during each estimation respectively.

This work has led on to improved methods for direct kinetic parameter estimation reliant on joint estimation of a reference TAC as well as kinetic parameters (Wang and Qi 2009b). Related methods, even more ambitious than the above, seek to estimate the true arterial input function which has physiological meaning (Yetik and Qi 2006).

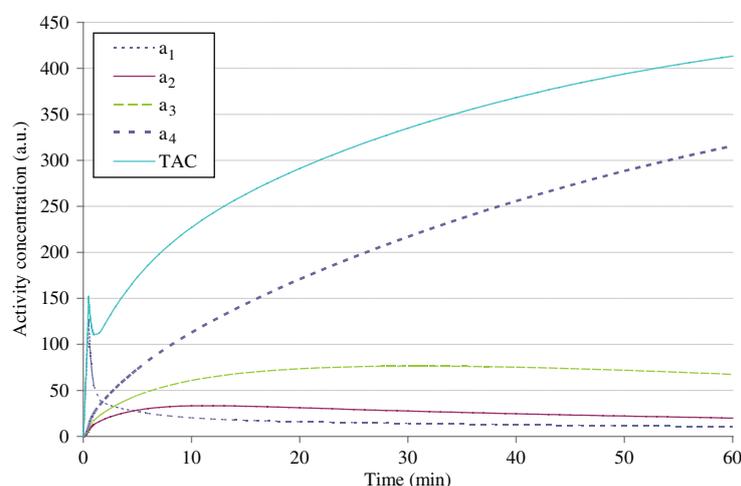
**4.3.2. Factorization approaches (linear models).** Other joint-estimation methods, that estimate both coefficient images as well as a set of temporal basis functions directly from the data (i.e. fully data derived, not imposing the constraint of a convolution of a generating function with decaying exponentials), can be considered as matrix factorization methods.

**Table 2.** Review of non-linear models for the temporal functions  $\Psi$  used in 4D image reconstruction for emission tomography.

Model name	$\theta_1, \theta_2$ and $C_1(t), C_2(t)$	Tissue response function, $h(t)$	How to obtain kinetic parameters from the $\theta$ parameters	Example references of application to 4D reconstruction
$\Psi(t; \theta_1, \theta_2, \dots, \theta_p) = \theta_1 C_1(t) + \theta_2 C_2(t) \otimes h(t; \theta_3, \theta_4, \dots, \theta_p)$				
1 TCM, no blood volume	$\theta_1 = 0$ $\theta_2 = 1$ $C_1(t) = 0$ $C_2(t) = C_{PL}(t)$	$\theta_3 \exp[-\theta_4 t]$	$K_1 = \theta_3$ $k_2 = \theta_4$	(Reader <i>et al</i> 2006b, Rakvongthai <i>et al</i> 2013)
1 TCM, with blood volume	$\theta_1 = F_v$ $\theta_2 = 1 - F_v$ $C_1(t) = C_{WB}(t)$ $C_2(t) = C_{PL}(t)$	$\theta_3 \exp[-\theta_4 t]$	$K_1 = \theta_3$ $k_2 = \theta_4$	(Zeng <i>et al</i> 1995, Huesman <i>et al</i> 1998, Reutter <i>et al</i> 1998, Yan <i>et al</i> 2009a, 2009b, Su <i>et al</i> 2013, Rahmim, Mohy-ud-Din 2014)
2 TCM, with blood volume	$\theta_1 = F_v$ $\theta_2 = 1 - F_v$ $C_1(t) = C_{WB}(t)$ $C_2(t) = C_{PL}(t)$	$\theta_3 \exp[-\theta_4 t] + \theta_5 \exp[-\theta_6 t]$	$K_1 = \theta_3 + \theta_5$ $k_2 = \frac{\theta_3 \theta_4 + \theta_5 \theta_6}{\theta_3 + \theta_5}$ $k_3 = \frac{\theta_3 \theta_5 (\theta_4 - \theta_6)^2}{(\theta_3 + \theta_5)(\theta_3 \theta_4 + \theta_5 \theta_6)}$	(Kamasak <i>et al</i> 2005)
Simplified reference tissue model (SRTM)	$\theta_1 = R_1$ $\theta_2 = k_2 - \frac{k_2 R_1}{1 + BP}$ $C_1(t) = C_{RF}(t)$ $C_2(t) = C_{RF}(t)$	$\exp\left[-\frac{k_2 t}{1 + BP_{ND}}\right]$	$k_4 = \frac{\theta_4 \theta_5 (\theta_3 + \theta_5)}{\theta_3 \theta_4 + \theta_5 \theta_6}$ $BP_{ND} = \frac{\theta_2 + \theta \theta_3}{\theta_3} - 1$	(Wang and Qi 2009b, Gravel and Reader 2013)

**Table 3.** Definition of terms for the kinetic models described in table 2.

Term	Definition
$K_1$	Rate constant for transfer of tracer from plasma to tissue ( $\text{mL}\cdot\text{min}^{-1}\cdot\text{cm}^{-3}$ )
$k_2$	Rate constant for return of tracer from tissue to blood ( $\text{min}^{-1}$ )
$k_3$	Rate constant for transfer of tracer from free to bound compartment ( $\text{min}^{-1}$ )
$k_4$	Rate constant for transfer of tracer bound to free compartment ( $\text{min}^{-1}$ )
$F_v$	Fractional volume of blood in the tissue
$K^*_1$	Rate constant for transfer of tracer from plasma to reference tissue ( $\text{mL}\cdot\text{min}^{-1}\cdot\text{cm}^{-3}$ )
$R_1$	Ratio of the delivery in the tissue region of interest compared to that in the reference region (ratio of influx, $K_1/K^*_1$ sometimes referred to as $R_f$ )
$BP_{ND}$	Binding potential - the ratio of the concentration of radioligand in the tissue to that of the reference region at equilibrium
$C_{WB}(t)$	Time-variant concentration of tracer in the arterial blood (including plasma and blood cells) ( $\text{Bq}\cdot\text{mL}^{-1}$ )
$C_{PL}(t)$	Time-variant concentration of tracer in plasma, the “input function” ( $\text{Bq}\cdot\text{mL}^{-1}$ )
$C_{RF}(t)$	Time-variant concentration of tracer in reference tissue region ( $\text{Bq}\cdot\text{mL}^{-1}$ )



**Figure 5.** Example TAC generated from the non-linear model for the case of 3 tissue compartments, showing the components that are added to generate the overall TAC. The components are  $a_1 = F_v C_{WB}(t)$ ,  $a_2 = (1 - F_v) C_{PL}(t) \otimes \theta_3 \exp[-\theta_4 t]$ ,  $a_3 = (1 - F_v) C_{PL}(t) \otimes \theta_5 \exp[-\theta_6 t]$  and  $a_4 = (1 - F_v) C_{PL}(t) \otimes \theta_7 \exp[-\theta_8 t]$ .

Such factorization has been considered as part of the 4D image reconstruction in Reader *et al* (2006a) and Sitek *et al* (2001). Note as mentioned before that there are many factorization approaches in the literature, but the focus here is on methods that have been incorporated into true 4D image reconstruction algorithms.

To describe these methods it is convenient to use the voxel-spatial basis function model given previously by equation (10) in combination with a discretization of equation (12) to obtain:

$$f_{jt}(\boldsymbol{\theta}) = \Psi_{jt}(\boldsymbol{\theta}) = \sum_{p=1}^P \theta_{jp} B_{pt} \quad (19)$$

where  $\mathbf{B}=\{B_{pt}\}$  is a  $PT$ -dimensional vector holding the temporal samples  $t = 1 \dots T$  for each of the  $p = 1 \dots P$  temporal basis functions (usually  $P < T$ ). Hence the discrete  $\mathbf{f}(\boldsymbol{\theta})$  is decomposed into coefficient images  $[\boldsymbol{\theta}]_p$  for each of the  $p = 1 \dots P$  temporal basis functions,  $[\mathbf{B}]_p$ . These two components are treated as unknowns so that both  $\boldsymbol{\theta}$  and  $\mathbf{B}$  are estimated, usually in an alternating fashion by holding one constant whilst estimating the other, and vice versa.

The differences between the various implementations lie in i) the constraints on the vectors  $\boldsymbol{\theta}$  and  $\mathbf{B}$ , ii) the cost function to quantify the approximation of the true object by equation (19) and iii) the algorithm used to find the two components. In Reader *et al* (2006a) the Kullback-Leibler distance objective is optimized using an EM type algorithm, so that all components of  $\boldsymbol{\theta}$  and  $\mathbf{B}$  are inherently non-negative, which in itself also provides noise reduction. In Sitek *et al* (2001) a quadratic penalty for negative elements in  $\boldsymbol{\theta}$  and  $\mathbf{B}$  is added to a least-squares objective that was optimized using a conjugate gradient method. It is important to note that the estimated basis functions are not unique and are not physiologically meaningful. Therefore the decomposition should only be seen as a means to reduce noise in the final spatiotemporal image.

**4.3.3. Shape constraints.** A somewhat different approach to using a pre-defined or estimated temporal parameterization is to consider conventional rectangle basis functions but to additionally introduce a shape constraint so that the final TAC is either just decaying, just increasing or else increasing up until a given time point (or interval) where the peak value is reached, after which the TAC is only allowed to decay. This approach was primarily developed for SPECT and has been termed the dSPECT (Bauschke *et al* 1999, Celler *et al* 2000, Farncombe *et al* 2001, Blinder *et al* 2004) but the approach can be applied to PET imaging as well. Different implementations also differ in cost function and optimization strategy.

**4.3.4. Joint estimation and mixed models.** An interesting mixed example of combining linear with non-linear parameterization in SPECT has been presented by Reutter (Reutter *et al* 2005). This approach combines a B-spline model for some regions (i.e. using the model of equation (12) with splines) with a one-tissue compartment model parameterization (a model like equation (17)) for the others, and furthermore seeks to estimate the input function and the region boundaries from the projections.

## 5. System model

Having taken care to define the parameters of interest which represent the 4D spacetime radiotracer distribution in the FOV, we now move on to consider how this parameterized object representation is imaged by a PET or SPECT scanner. The process of modeling the physics of radioactive emissions from the radiotracer distribution, along with how these emissions are detected and processed by the tomograph, constitutes the process of *system modeling*. This is also known as the forward model in the field of inverse problems, where, in general, one ultimately seeks to obtain some approximate inverse of this model, in order to estimate parameters representative of the object which gave rise to the measured data. Whilst system modeling can be based on the object representation function  $f$  of the continuous variables  $\mathbf{r}$  and  $t$ , we will focus on system modeling for the  $VT$ -dimensional vector  $\mathbf{f}(\boldsymbol{\theta})$ , which is the space-time discretization of the representation  $f(\mathbf{r}, t; \boldsymbol{\theta})$  of the 4D radiotracer distribution.

In statistical image reconstruction the system model is required to model the *mean* or the *expectation* of the measured emission data for a given object vector  $\mathbf{f}$ , where, for 4D,  $\mathbf{f}$  is a time-series of 3D images. For the majority of emission tomographs the space of the measurements, whether noisy or not, can be regarded as an  $IT$ -dimensional vector space. The  $I$

dimensions typically correspond to the number of distinct sinogram bins (including the time-of-flight bins if appropriate), and  $T$  corresponds to the number of time samples or frames. So for dynamic imaging one needs to map the  $VT$ -dimensional  $\mathbf{f}$  (a time series of images) to a space of  $IT$  dimensions (a time series of sinograms, for example). As will be seen below, this will be achieved using a very large  $IT \times VT$  matrix, called the system matrix.

Whilst we have an infinite number of choices for the modeling and parameterization of  $\mathbf{f}$  (via  $\boldsymbol{\theta}$  and the spatiotemporal functions chosen) we are in contrast necessarily constrained by the scanner hardware in terms of modeling the mean of the measured data. In effect, there are for most scanners only a finite number of possible event types that can be detected. For example, there are only a finite number of discrete sinogram bins for a given scanner, with each bin being sensitive to a particular weighted volume of the FOV, defined by a sensitivity function. Both list-mode data and sinogram / projection data can be regarded in this same way, as there will nearly always be a finite precision (effectively a sampling interval) for any measured event. We will denote the dynamic measured data for an emission tomograph by a single vector  $\mathbf{m}$ , of  $IT$  dimensions, which can be regarded as a time-series of  $I$ -dimensional subvectors, each subvector being a dataset from a static acquisition.

Any given component  $m_{it}$  of the total measured data vector  $\mathbf{m}$  has a time index  $t$ , corresponding to a time sampling interval of start time  $t_s$ , and end time  $t_e$ , with duration  $\Delta t$ . This could be anything from 1 ms (e.g. the high resolution research tomograph (HRRT) PET scanner (de Jong *et al* 2007) records list-mode data to within 1 ms time sampling) to 60 min (e.g. a non-dynamic acquisition, collecting a static set of sinograms). However, in the following we will focus on the commonly used case of matched time sampling intervals for both  $\mathbf{f}$  and  $\mathbf{m}$ .

Assuming that the emission tomography scanner can be represented by a system matrix  $\mathbf{A} = \{a_{it,vt}\}_{IT \times VT}$  operating on this vector  $\mathbf{f}$ , we will model the mean of  $\mathbf{m}$  by a vector  $\mathbf{q}$ :

$$\mathbf{q}(\boldsymbol{\theta}) = \mathbf{A}\mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\rho} \quad (20)$$

where  $\boldsymbol{\rho} = \{\rho_{it}\}_{IT}$  is a model or estimate of any background component not accounted for by  $\mathbf{A}$  (such as randoms, or scatter). Each component of  $\mathbf{q}$  is given by

$$q_{it}(\boldsymbol{\theta}) = \sum_v a_{it,vt} f_v(\boldsymbol{\theta}) + \rho_{it} \quad (21)$$

With the practical pre-discretization of  $f$  used here (i.e. using  $\mathbf{f}(\boldsymbol{\theta})$  rather than  $f(\mathbf{r}, t; \boldsymbol{\theta})$ ) the system matrix  $\mathbf{A}$  contains elements  $a_{it,vt}$  specifying the intersection of each voxel volume  $v$  at time  $t$  with each potentially time-dependent sensitivity function  $\xi_{it}(\mathbf{r})$  associated with each data element  $i$ . This models the probability that a radioactive emission from within a given voxel  $v$ , during time sampling interval  $t$ , gives rise to a measured count in an element  $it$  of the  $IT$ -dimensional data space for the same time sampling interval  $t$ . A common example of  $\xi_{it}(\mathbf{r})$  would be a tube or line of response through the FOV, so that  $a_{it,vt}$  gives the intersection of a tube or line of response  $i$ , at a time  $t$  with a voxel  $v$  at time  $t$ . This permits a practical and fast on-the-fly calculation of the elements of matrix  $\mathbf{A}$ , avoiding the phenomenal storage requirements, using methods such as the Siddon algorithm to calculate intersection lengths of lines with voxels (Siddon 1985, Zhao and Reader 2003). But more advanced models of the scanner geometry (Markiewicz *et al* 2005), resolution (Panin *et al* 2006, Kotasidis *et al* 2011) and even photon scatter (Markiewicz *et al* 2007) can in principle be included in  $\mathbf{A}$ . Furthermore, a commonly adopted simplification when implementing (21) for 4D imaging is to use exactly the same system matrix for every time frame, so that  $a_{it,vt}$  is obtained simply by  $a_{i,v}$ , but note the important need to account for radioactive decay (see equations (22) and (23) below).

However system models should in general account for scanner responses varying with time, which can occur for many reasons, including i) scanner motion (e.g. camera rotation in

some SPECT systems), and ii) the total activity in the FOV changing with time, as scanners do not respond in a strictly linear fashion to such changes, due to limitations in scanner instrumentation/electronics. Even if only rarely done to date, incorporating a time-varying  $\mathbf{A}$  into 4D reconstruction is not conceptually difficult to do.

Notice that the system modeling equation (20) presented here for 4D reconstruction is more general than conventional modeling in 3D image reconstruction, as  $\mathbf{f}$  can be either a linear function of  $\boldsymbol{\theta}$  (such as a discretization of (12)) or a non-linear function of  $\boldsymbol{\theta}$  (such as a discretization of (17)) whilst nonetheless preserving a linear forward projection model.

More realistic system models often use the simple line intersection model as just part of the system matrix. One proceeds, stage by stage, to model what happens with the radioactive emissions from the approximate representation  $\mathbf{f}(\boldsymbol{\theta})$ . For PET, the system matrix  $\mathbf{A}$  is normally factorized into a resolution modeling component (e.g. a convolution matrix  $\mathbf{H}$  to model positron range (Reader *et al* 2003)), a geometric detection probability component (e.g. the discrete x-ray transform which performs line integrals,  $\mathbf{X}$ ), and normalization, attenuation and radioactive decay components (diagonal matrices  $\mathbf{N}$ ,  $\mathbf{L}$  and  $\mathbf{D}$  respectively),  $\mathbf{A}=\mathbf{LNXHD}$ , giving:

$$\mathbf{q}(\boldsymbol{\theta}) = \mathbf{LNXH}\mathbf{D}\mathbf{f}(\boldsymbol{\theta}) \quad (22)$$

The diagonal matrix  $\mathbf{D}$ , modeling radioactive decay, can also be modified to account for frame length variations in the temporal sampling of  $\mathbf{f}$ . Finally any background components such as the model of the mean of the time-dependent scatter and random events (for PET) are added on:

$$\mathbf{q}(\boldsymbol{\theta}) = \mathbf{LNXH}\mathbf{D}\mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\rho} \quad (23)$$

Of course, the above system model for PET can easily be extended to account for time of flight, primarily by modifying the matrix  $\mathbf{X}$  to hold time-of-flight kernels rather than lines, to project into different time-of-flight sinogram bins. For SPECT the attenuation cannot be handled by a diagonal matrix, and the discrete x-ray transform is also too approximate: it is necessary to account for depth-dependent sensitivity functions of the collimator holes, and depth-dependent attenuation factors in the system model (Zeng *et al* 1991).

As mentioned at the beginning of this section, the system model concerns *the mean*, or the expectation, of the data obtained from the overall emission tomography measurement process. Conceptually, the mean data can be regarded as what would be obtained by repeating precisely the same noisy emission scan a very large number of times and then taking the average of these independent data set realizations. The count limited effects of a single scan, resulting in noisy measured data, are not modeled in the system model, but are dealt with statistically in the objective function, covered in the next section for the 4D reconstruction context.

## 6. Objective functions for 4D reconstruction

The objective function for image reconstruction describes the goal for the parameter estimation process: how should the estimated parameters agree, or not, with the actually measured emission data  $\mathbf{m}$ ? The objective functions for 4D image reconstruction remain relatively unchanged from the 3D cases, and are simply extended in the obvious way to include the time dimension. For both 4D PET and SPECT imaging, the probability of obtaining  $m_{it}$  counts in data bin  $i$  at time index  $t$  if on average there are  $\mu_{it}$  counts for that data bin, is well modeled as a Poisson distribution:

$$\Pr(m_{it}|\mu_{it}) = \frac{\exp[-\mu_{it}]\mu_{it}^{m_{it}}}{m_{it}!} \quad (24)$$

In image reconstruction, we are modeling the mean counts  $\mu_{it}$  in bin  $i$  at time index  $t$  by  $q_{it}(\boldsymbol{\theta})$  as described in the previous section, so we can now find the probability of obtaining measured  $m_{it}$  counts in data bin  $i$  at time  $t$  for a given parameter vector  $\boldsymbol{\theta}$ :

$$\Pr(m_{it}|\boldsymbol{\theta}) = \frac{\exp[-q_{it}(\boldsymbol{\theta})][q_{it}(\boldsymbol{\theta})]^{m_{it}}}{m_{it}!} \quad (25)$$

Now, to find the parameters of interest  $\boldsymbol{\theta}$ , we can define the likelihood of  $\boldsymbol{\theta}$  given the fact that  $m_{it}$  occurred by:

$$l(\boldsymbol{\theta}|m_{it}) = \Pr(m_{it}|\boldsymbol{\theta}) \quad (26)$$

and so we need to find a  $\boldsymbol{\theta}$  which maximizes this likelihood. However, we have of course many more measurements than just one  $m_{it}$  for one data bin, so we can define the likelihood of a given parameter vector  $\boldsymbol{\theta}$  given the entire measured data vector  $\mathbf{m}$  by:

$$\lambda(\boldsymbol{\theta}|\mathbf{m}) = \prod_{t=1}^T \prod_{i=1}^I \frac{\exp[-q_{it}(\boldsymbol{\theta})][q_{it}(\boldsymbol{\theta})]^{m_{it}}}{m_{it}!} \quad (27)$$

where it is assumed that the integer number of counts in each element of the measured data vector  $\mathbf{m}$  are all independent of one another, hence the product of the individual probabilities. The objective of the reconstruction is therefore to maximize  $\lambda$  with respect to the parameter vector  $\boldsymbol{\theta}$ , where any given  $\boldsymbol{\theta}$  defines a 4D radiotracer distribution  $\mathbf{f}$  and therefore also a mean data vector  $\mathbf{q}$  by the system model of the previous equation (20),  $\mathbf{q}(\boldsymbol{\theta}) = \mathbf{A}\mathbf{f}(\boldsymbol{\theta}) + \boldsymbol{\rho}$ . It is much simpler for optimization purposes to consider the natural logarithm of equation (27), the Poisson log-likelihood, and to ignore the constant term which is independent of the parameters:

$$\ln \lambda(\boldsymbol{\theta}|\mathbf{m}) = \sum_{t=1}^T \sum_{i=1}^I (m_{it} \ln q_{it}(\boldsymbol{\theta}) - q_{it}(\boldsymbol{\theta})) \quad (28)$$

which will attain its maximum at the same value of  $\boldsymbol{\theta}$  as for (27) since the logarithm is a monotonically increasing function. It is very useful to note at this point that the Poisson log-likelihood (28) is related to a distance measure known as the Kullback-Leibler (KL) distance (Barrett and Myers 2004):

$$D^{(KL)}(\mathbf{m}, \mathbf{q}(\boldsymbol{\theta})) = \sum_{t=1}^T \sum_{i=1}^I \left( m_{it} \ln \frac{m_{it}}{q_{it}(\boldsymbol{\theta})} + q_{it}(\boldsymbol{\theta}) - m_{it} \right) = -\ln \lambda(\boldsymbol{\theta}|\mathbf{m}) + \text{const}. \quad (29)$$

Therefore maximizing the Poisson log-likelihood (28) is equivalent to minimizing the KL distance (29):

$$\underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ln \lambda(\boldsymbol{\theta}|\mathbf{m}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} D^{(KL)}(\mathbf{m}, \mathbf{q}(\boldsymbol{\theta})) \quad (30)$$

As will be seen later, the KL distance is an extremely important measure of agreement between two vectors, as in fact both Poisson maximum likelihood and penalized likelihood estimation of linear or non-linear choices of parameters  $\boldsymbol{\theta}$  from tomographic emission data can legitimately be achieved in *image space* provided this distance measure is used. The key, as will be shown, is the use of an EM tomographic image update  $\mathbf{f}^{(\text{EM})}$  (in place of the noisy measured data  $\mathbf{m}$  in (28)) for fitting the modeled  $\mathbf{f}(\boldsymbol{\theta})$ , using the KL distance as the objective. We will come back to this in the next section.

A commonly reported problem with seeking an estimate of  $\boldsymbol{\theta}$  which maximizes (28) (i.e. the maximum likelihood estimate of  $\boldsymbol{\theta}$ ), is that with a noisy measured vector  $\mathbf{m}$ , the estimate

of  $\theta$  is also noisy. It should first be noted that this frequently reported problem presupposes a choice of space and time functions and a parameterization  $\theta$  of the radiotracer distribution such that noisy solutions are even permitted. This does not have to be the case, and in fact one can select spatiotemporal functions which implicitly limit the representations of the radiotracer distribution  $f(\mathbf{r}, t)$  to a subspace which permits only smooth reconstructions (analogous to the method of sieves for the purely spatial case, (Snyder and Miller 1985)). If for example one chooses regions or volumes of interest as the spatial basis functions, with a positivity constraint on  $\theta$ , then the reconstructed image can only consist of these uniform regions, no matter how noisy the data are, such that the famous ‘night sky’ images cannot even be obtained. Likewise, if temporally extensive temporal basis functions are chosen with a positivity constraint on the coefficients of these functions, it is impossible to obtain visually noisy TACs.

Nonetheless, a popular approach to noise reduction has been to keep a choice of spatiotemporal functions which does permit very noisy reconstructions (representations of  $f(\mathbf{r}, t)$ ), and instead to place restrictions on the parameter vector  $\theta$  by regularizing the objective function. Therefore instead of seeking just to maximize the likelihood (the ML estimate of  $\theta$ ) one can use a maximum *a posteriori* (MAP) objective. In this case one includes the presumed probability of any given  $\theta$ :

$$\Pr(\theta) = \frac{\exp[-\beta U(\theta)]}{Z} \quad (31)$$

where  $Z$  is a normalizing constant which can be ignored since we will subsequently only be concerned with maximization with respect to  $\theta$ , and  $\beta$  is the regularization parameter, controlling the level of importance of the prior ( $\beta=0$  implies no prior probability, all  $\theta$  being equally probable).  $U$  is any function which gives a large value for parameter vectors  $\theta$  which are deemed improbable, and low values for vectors  $\theta$  which one presumes to be more probable, based on prior knowledge of what should be expected from a PET or SPECT scan. A common choice is to define  $U$  as an energy function, as will be described below. Using this probability in product with the likelihood gives what is known as the posterior probability of the parameters  $\theta$ , given  $\mathbf{m}$ :

$$\lambda^{(MAP)}(\theta | \mathbf{m}) = \prod_{t=1}^T \prod_{i=1}^I \frac{\exp[-q_{it}(\theta)] [q_{it}(\theta)]^{m_{it}}}{m_{it}!} \frac{\exp[-\beta U(\theta)]}{Z} \quad (32)$$

The log of the posterior probability is therefore given by

$$\ln \lambda^{(MAP)}(\theta | \mathbf{m}) = \sum_{t=1}^T \sum_{i=1}^I (m_{it} \ln q_{it}(\theta) - q_{it}(\theta)) - \beta U(\theta) + const \quad (33)$$

Equation (33) is known variously as the Poisson log-posterior or the penalized Poisson log-likelihood, and from the previous discussion can also be seen as the negative of a penalized KL distance (to within a constant).

There are countless possibilities for the choice of  $U$ , primarily viewed as an energy function, and the reader is referred to the 3D image reconstruction literature for consideration of example choices for 3D spatial regularization (see for example the review of Qi and Leahy (2006)). Temporal penalties have also been proposed, such as penalizing differences between the reconstructed TAC and its best compartmental model fit curve (Kadrmas and Gullberg 2001), as well as Gibbs priors (Lee *et al* 2005, Gravier *et al* 2006). In fact for direct parametric image reconstruction some interesting possibilities arise (Wang and Qi 2012, Wang and Qi 2013): one can regularize according to an energy function defined by the activity images  $\mathbf{f}(\theta)$  and/or regularize

according to an energy function defined by the parametric images  $[\boldsymbol{\theta}]_p = \{\theta_{jp}, j=1\dots J\}$  for  $p=1\dots P$ , as will now be explained.

First, the basic form of the energy term used for 4D parametric reconstruction, when using voxels and regularizing by the activity, is given by Kamasak *et al* (2005), Wang and Qi (2012), Rakvongthai *et al* (2013):

$$U(\boldsymbol{\theta}) = \sum_{t=1}^T \sum_{j=1}^J \sum_{l \in N_j} w_{jl} (\Psi_{jt}(\boldsymbol{\theta}_j) - \Psi_{lt}(\boldsymbol{\theta}_l))^2 \quad (34)$$

where the total energy is given by the weighted ( $w_{jl}$ ) sum over the neighbourhood  $N_j$  of every voxel  $j=1\dots J$ , summing the square difference between the activity at voxel  $j$ , and the activities at voxels  $l$  within the neighbourhood of voxel  $j$ . More generally one can use a function  $\mathcal{T}$  to transform the parameter vector  $\boldsymbol{\theta}_j$  for a given voxel  $j$  into one of the kinetic parameters (for example, mapping  $\boldsymbol{\theta}_j$  to obtain  $k_4$ , as shown in table 2 for the 2 TCM model). This allows regularization of the kinetic parametric images as well

$$U(\boldsymbol{\theta}) = \sum_{p=1}^P \sum_{j=1}^J \sum_{l \in N_j} w_{jl} (\mathcal{T}_{jp}(\boldsymbol{\theta}_j) - \mathcal{T}_{lp}(\boldsymbol{\theta}_l))^2 \quad (35)$$

With equations (34) and (35) it is important to realize that the relative weighting of the energy term will matter: one might need to assign more weight to parametric images which are known to suffer from more noise, such as the microparameters  $k_2$  and  $k_4$ . Also, as mentioned, both parametric and activity based regularization are simultaneously possible.

With regard to the choice of weighting factors  $w_{jl}$ , all the work on 3D regularization can be applied, such as in particular the use of MR images as anatomical priors which can guide the choice of these weighting factors, as proposed by Bowsher for 3D reconstruction (Bowsher *et al* 1996). Tang *et al* as well as Bousse *et al* applied a modified version of the Bowsher prior to use MR regularization in direct parametric image reconstruction (Tang *et al* 2008, 2010, Bousse *et al* 2012).

Despite the attraction of MAP estimation, the ML objective still remains popular primarily because many implementations never seek an actual ML estimate, but rather an incompletely converged reconstruction which often has a more desirable signal to noise ratio than the ML estimate.

## 7. Algorithms for 4D image reconstruction

An iterative image reconstruction algorithm seeks to improve a given parameter vector estimate  $\boldsymbol{\theta}^{(k)}$  such that when the estimate is used to generate a 4D image  $\mathbf{f}^{(k)}$  and subsequently mapped by  $\mathbf{A}$  to give a model of the mean data  $\mathbf{q}^{(k)}$ ,  $\mathbf{q}^{(k)}$  agrees in a better way with the measured data  $\mathbf{m}$ . For Poisson data this better agreement is specified by obtaining a new estimate  $\boldsymbol{\theta}^{(k+1)}$  which reduces the KL distance between  $\mathbf{m}$  and  $\mathbf{q}^{(k)}$ , as was seen in the previous section. Now, since the parameters being estimated are often non-linear for 4D reconstruction, the algorithms require a more comprehensive coverage for the 4D case, as most 3D reconstruction algorithms for emission tomography are limited to a completely linear relationship between  $\boldsymbol{\theta}$  and  $\mathbf{q}$ .

### 7.1. ML and MAP estimation for linear parameters

We nonetheless start with the case of linear parameter estimation, as this corresponds to using any of the temporal basis functions previously given in table 1, representing a significant

portion of the 4D literature. For these cases, conventional 3D iterative image reconstruction algorithms can be readily adapted, just by modifying the linear forward model to account for the choice of spatiotemporal basis functions. One of the most popular 3D iterative image reconstruction algorithms in PET and SPECT is the expectation maximization (EM) algorithm (Dempster *et al* 1977, Shepp and Vardi 1982) used to find ML estimates of  $\theta$ . Its popularity can be understood due to the extreme simplicity of its implementation and its robust performance, often delivering visually appealing images if not run to the point of convergence (see for example Angelis *et al* (2011a)). For 3D reconstruction, with the parameter vector  $\theta$  corresponding directly to 3D image voxel values,  $\mathbf{f}(\theta) = \theta$ , the ML-EM algorithm is given by:

$$\theta^{(k+1)} = \frac{\theta^{(k)}}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{q}^{(k)}} \quad (36)$$

where vector division and multiplication is taken to be component-wise. The model of the mean accounts for a model or estimate of the mean scatter and random events, contained in  $\rho$ :

$$\mathbf{q}^{(k)} = \mathbf{A}\theta^{(k)} + \rho \quad (37)$$

Practical implementation of (36) is relatively straightforward. A current 3D image estimate  $\theta^{(k)}$  is forward projected by  $\mathbf{A}$  to give  $\mathbf{q}^{(k)}$ , the ratio of the measured data  $\mathbf{m}$  to  $\mathbf{q}^{(k)}$  is backprojected by  $\mathbf{A}^T$  to give an image which is used, after first dividing by the sensitivity image  $\mathbf{A}^T \mathbf{1}$  (a backprojection of unit data), to multiplicatively update  $\theta^{(k)}$ . This sequence of steps provides a new estimate  $\theta^{(k+1)}$ .

For the 4D case with linear parameters to estimate (such as all the cases shown in table 1), equation (36) can be adapted in an obvious way:

$$\theta^{(k+1)} = \frac{\theta^{(k)}}{\mathbf{B}^T \mathbf{A}^T \mathbf{1}} \mathbf{B}^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{q}^{(k)}} \quad (38)$$

with

$$\mathbf{q}^{(k)} = \mathbf{A}\mathbf{B}\theta^{(k)} + \rho \quad (39)$$

Hence the main difference for 4D reconstruction with linear parameters compared to 3D reconstruction is the use of a matrix  $\mathbf{B} = \{b_{v_t, v_p}\}_{V_T \times V_P}$  containing the time-sampled temporal basis functions for every voxel volume. The size of matrix  $\mathbf{B}$  is considerable, but it is nearly always simplified by assuming the temporal basis functions to be identical for every voxel, such that  $b_{v_t, v_p} = b_{t, p}$ . The measured data vector in equation (38) now contains elements for every time point ( $\mathbf{m} = \{m_{it}\}$ ), and hence the forward model is required to generate time-dependent estimates of  $\mathbf{q}$ . This 4D MLEM approach has been widely used, starting in the late 1990s with (Matthews *et al* 1997), and continuing with many others more than a decade later (Tsoumpas *et al* 2008b, Hong and Reader 2008). All the cases in the literature shown in table 1 can be used with equation (38). However, the method can be computationally demanding: if the temporal basis functions cover all time frames, then the entire dynamic data is implicated in each and every image update, and furthermore convergence can be extremely slow for non-orthogonal basis functions, such as the two Patlak temporal basis functions.

In contrast, the nested EM method, proposed by Wang and Qi (2010) and based essentially on the original work of Carson and Lange (1985), is more computationally straightforward to implement. It consists of two far less demanding stages which separate the tomographic part of the problem from the image-space based part of the problem. First, the current estimate of the parameters  $\theta^{(k)}$  is used to form the current 4D image according to:

$$\mathbf{f}(\theta^{(k)}) = \mathbf{B}\theta^{(k)} \quad (40)$$

Then a provisional EM update for each and every 3D image frame is obtained:

$$\mathbf{f}_{EM}^{(k)} = \frac{\mathbf{f}(\boldsymbol{\theta}^{(k)})}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A} \mathbf{f}(\boldsymbol{\theta}^{(k)}) + \boldsymbol{\rho}} \quad (41)$$

where it is important to note that  $\mathbf{B}$  does not appear in (41), and the comparison to the regular 3D ML-EM update equation (36) should be clear. Equation (41) accounts for the tomographic system model  $\mathbf{A}$  and the measured data  $\mathbf{m}$  to deliver a provisional 4D image update  $\mathbf{f}_{EM}$ . Finally, a simple image-space update can be carried out to obtain a new estimate of the parameters:

$$\boldsymbol{\theta}^{(k+1)} = \frac{\boldsymbol{\theta}^{(k)}}{\mathbf{B}^T \mathbf{1}} \mathbf{B}^T \frac{\mathbf{f}_{EM}^{(k)}}{\mathbf{f}(\boldsymbol{\theta}^{(k)})} \quad (42)$$

Equation (42) can be viewed as an image-space EM algorithm: the noisy data are now given by  $\mathbf{f}_{EM}$  (where the  $\mathbf{f}_{EM}$  image accounts for the tomographic data and the system model), and the ‘model of the mean’ now only needs to be concerned about the image-space model, without concern about the tomographic component. This greatly aids implementation.

Not only do equations (41) and (42) separate the image model from the tomographic model, they are also much more memory efficient and practical to implement than the fully 4D version of MLEM (equation (38)). Only a time series of images, rather than a full time-series of projection data, needs to be handled in memory for finding the coefficients of the temporal basis functions, and remarkably the results are equivalent to the more demanding equation (38), under the modest requirement that the system matrix  $\mathbf{A}$  be time-invariant. Proof of this equivalence can be seen by substituting (41) into (42):

$$\boldsymbol{\theta}^{(k+1)} = \frac{\boldsymbol{\theta}^{(k)}}{\mathbf{B}^T \mathbf{1}} \mathbf{B}^T \frac{1}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{q}^{(k)}} = \frac{\boldsymbol{\theta}^{(k)}}{\mathbf{B}^T \mathbf{1} \mathbf{A}^T \mathbf{1}} \mathbf{B}^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{q}^{(k)}} \quad (43)$$

which requires  $\mathbf{A}$  to be the same for all time frames in order to equal (38). The image-space update equation (42) can be applied more than once, in order to provide acceleration if desired.

Not only can the EM algorithm be easily adapted to a 4D version, but also least-squares methods are just as readily adapted, such as for example the simultaneous algebraic reconstruction technique (SART) (Andersen and Kak 1984). As for EM, one simply uses a 4D system model rather than a 3D one:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \lambda^{(k)} \frac{\mathbf{1}}{\mathbf{B}^T \mathbf{A}^T \mathbf{1}} \mathbf{B}^T \mathbf{A}^T \left[ \frac{\mathbf{m} - \mathbf{A} \mathbf{B} \boldsymbol{\theta}^{(k)}}{\mathbf{A} \mathbf{B} \mathbf{1}} \right] \quad (44)$$

where  $\lambda$  is an optional relaxation parameter ( $\lambda=1$  in the original version). SART can offer potential benefits as it does not have the non-negativity constraint of the EM algorithm, a constraint which can be viewed both as beneficial (reducing noise in low count regions) and detrimental (introducing bias in low count regions). Nonetheless, due to the popularity of EM, even versions of EM which permit negative values have been proposed in the literature. The two main examples are NEG-ML (Nuyts *et al* 2002, Grezes-Besset *et al* 2007) and AB-EM (Byrne 1998, Verhaeghe and Reader 2010), which Rahmim *et al* picked up specifically for use in 4D reconstruction (Rahmim *et al* 2012) so as to permit necessary negative values when using the Logan basis functions for reversibly binding radiotracer kinetics.

Of course, in addition to the above examples, other 3D reconstruction algorithms have also been modified for 4D in a similar fashion, including the preconditioned conjugate gradient algorithm (PCG) (Wang *et al* 2008, Rakvongthai *et al* 2013), for accelerated convergence. In

fact the nested EM algorithm itself has also been further accelerated through use of conjugate gradients (Wang and Qi 2010).

It is interesting to note at this point that, in the same spirit as the nested EM algorithm, MAP estimation has also been previously proposed as an image-space operation, providing parameter updates expressed in terms of the EM update image  $\mathbf{f}_{EM}$ . A popular and simple example of this, though not theoretically robust, is the one step late (OSL) EM MAP algorithm (Green 1990):

$$\boldsymbol{\theta}^{(k+1)} = \frac{\mathbf{f}_{EM}^{(k)}}{\mathbf{1} + \frac{\beta}{s} U'(\boldsymbol{\theta}^{(k)})} \quad (45)$$

where  $U'$  is the partial derivative of the energy function evaluated at the current estimate of the parameters (hence OSL, with a well-known choice being the so-called median root prior (Alenius and Ruotsalainen 1997)), and  $\mathbf{s}$  is the sensitivity image defined by a backprojection of unit data,  $\mathbf{A}^T \mathbf{1}$ . This OSL method has itself been applied to fully 4D reconstruction (Tsoumpas *et al* 2013).

Many other MAP estimation methods implicitly or explicitly use the EM update image to achieve updates of the parameters via image space. Examples include the very early method of Levitan and Herman (1987), generalized EM (Hebert and Leahy 1989) and the more robust approach of DePierro (1995). All these approaches use image-space updates which require  $\mathbf{s}$  and  $\mathbf{f}_{EM}$  but without need of the tomographic data  $\mathbf{m}$ , and hence one can increasingly appreciate the great utility of the EM update image.

## 7.2. Non-linear parameters

Since non-linear parameter estimation in 3D emission tomography image reconstruction has been largely overlooked, there have been no 3D reconstruction algorithms ready and waiting for simple adaptation to 4D non-linear parameter estimation. Specialized algorithms have needed to be developed. Given the necessity of these developments for 4D reconstruction, there is now in fact scope for transfer of methodology in the opposite direction: 4D non-linear estimation methods (to be described below) can serve to increase the generality of 3D image reconstruction algorithms for emission tomography. In fact this has already started for the example of resolution modeling with conventional linear parameter estimation, where exploitation of a 4D method has allowed advances to be made in a nested 3D method (Angelis *et al* 2013).

In the following coverage one key methodology proposed for PET will be focused on, as in principle it can be applied to *any* of the TAC models previously described, whether linear or non-linear, and also applies to temporal, spatial or spatiotemporal parameters. The method is essentially a nested EM approach, but with an image-space update that accommodates non-linear parameters. The principles of the method date back to 1985, to an algorithm proposed by Carson and Lange named 'EMPIRA' (the EM parametric image reconstruction algorithm, (Carson and Lange 1985)). However, no concrete results were shown in that work, nor in the slightly earlier pioneering work for compartmental kinetic parameter estimation from dynamic time of flight PET data (Snyder 1984). An implementation of the original 1984 work of Snyder was explored more recently (Schottlander *et al* 2006), but unfortunately without use of any tomographic system model and hence again no concrete results. Returning to EMPIRA, others, e.g. Chiao *et al* (1994), did recognize the potential of the EMPIRA approach, even using the method to estimate ROI boundaries. The principles of the 1985 EMPIRA method have, 25 years later, been independently proposed afresh, two years apart, by two key research groups in this area: Matthews *et al* (2010) and Wang and Qi (2012), the latter group referring to their method as OTEM (optimization transfer EM). Curiously a similar situation occurred with the

widespread EM algorithm for 2D and 3D reconstruction, which was proposed independently by two groups (Shepp and Vardi 1982) and (Lange and Carson 1984), also two years apart!

Whilst the first appearance of the method about to be presented with demonstrated concrete results was in 2010, an identical (but not statistically-derived) nested EM image-space fitting approach was nonetheless proposed even earlier in 2006 (Reader *et al* 2006b).

The method proceeds in two key stages, with the two stages forming a cycle which is iterated.

*Stage 1: An EM-type update of the images from the raw measured data.* Starting with a given estimate of the parameters  $\theta^{(k)}$ , which in the current context are taken as kinetic parameters (but of course, this is not a requirement), the corresponding time series of 3D images  $\mathbf{f}(\theta^{(k)})$  (e.g. a TAC for every voxel) is created from those parameters. The PET or SPECT forward model  $\mathbf{A}$  is then applied to generate the estimate of the mean measured data,  $\mathbf{A}\mathbf{f}(\theta^{(k)})$ , for use in a conventional EM-type update:

$$\mathbf{f}_{EM}^{(k)} = \frac{\mathbf{f}(\theta^{(k)})}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\mathbf{f}(\theta^{(k)}) + \rho} \quad (46)$$

Equation (46) is as simple to implement as the well-established MLEM algorithm applied individually to each 3D image frame to generate an intermediate dynamic (4D) image update. In the original EMPIRA proposal of Lange and Carson in 1985, this update step is in effect obtained by combining equations (8) and (10) in that article (where also a time dependent system matrix  $\mathbf{A}$  is permitted).

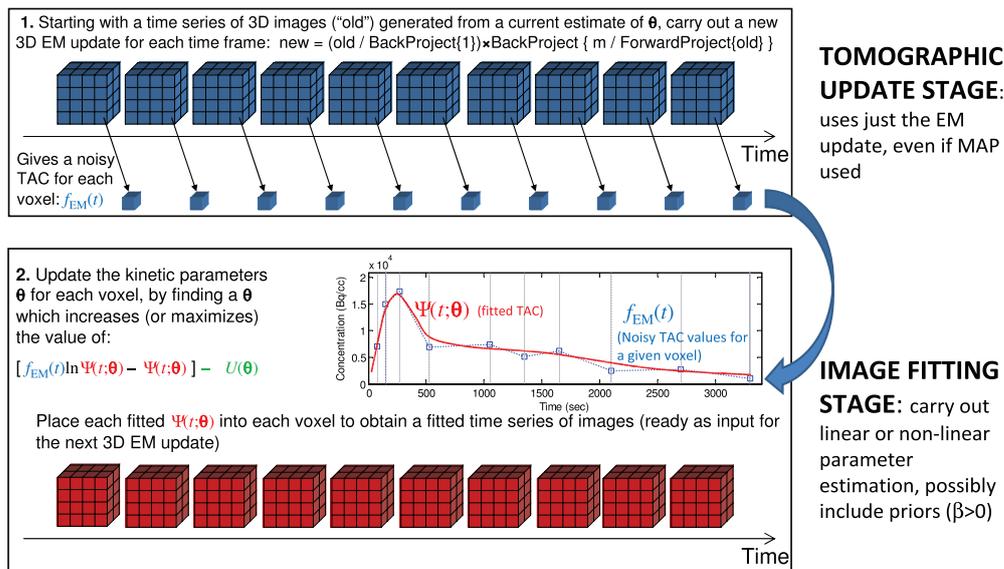
*Stage 2: Image space fitting of the parameters.* Stage two involves updating the parameters,  $\theta^{(k+1)}$ , by finding parameter values which fit the EM update image series obtained from equation (46). Just as the regular MLEM algorithm seeks a maximum of the Poisson log-likelihood, equivalent to minimizing the KL distance, so also a sensitivity-image weighted KL distance is now used in the *image space fitting*. We will keep this as a maximization problem, given by:

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}} \sum_{vt} \left[ s \left( \mathbf{f}_{EM}^{(k)} \ln \mathbf{f}(\theta) - \mathbf{f}(\theta) \right) \right]_{vt} \quad (47)$$

where the maximization considers all  $VT$  samples in the discrete 3D image time series. The link with the previously described KL distance in (28) and (29) is clear, but with the substitution of the EM-type dynamic updated image  $\mathbf{f}_{EM}$  in place of the measured data and the non-linear modeled  $\mathbf{f}$  in place of the linearly modeled  $\mathbf{q}$ . Equation (47) corresponds to equation (11) in Carson and Lange's 1985 article, and it does not necessarily have to be maximized, so long as an increase is however achieved. The link with the nested EM algorithm for linear parameter estimation is clear: equation (41) corresponds to (46), and the image-space EM update (42) increases the objective function of (47).

Having carried out stage 2, stage 1 is then repeated using the new parameter estimates. Note that due to the use of non-linear parameters, convergence to a local maximum only is assured. Figure 6 illustrates the 2 stage process.

These two stages, which are essentially identical in three independent publications from 1985, 2010 and 2012, constitute a powerful framework for estimation of non-linear parameters in emission tomography, whether 2D, 3D, 4D or of even higher dimensions. Having identified the common framework, it is worth mentioning again some useful distinctions in the independent proposals: i) the Lange and Carson (EMPIRA) 1985 derivation explicitly accommodates a time-dependent system matrix, ii) the Wang and Qi 2012 derivation (OTEM) explicitly includes a MAP (rather than just ML) objective function, and iii) the Matthews *et al* 2010 proposal suggests an approximate least squares approach to solving (47) as will be described below.



**Figure 6.** Illustration of one iteration of the two stage process for fully 4D or direct parametric image reconstruction in emission tomography, based on the methods of Carson and Lange (1985), Matthews *et al* (2010) and Wang and Qi (2012). In stage two, the maximizing of the negative KL distance should be weighted by the time-dependent sensitivity images, but this is not necessary for TAC fitting at the voxel level with a time-invariant system matrix.

Finally, if regularization is required, then this can be incorporated in the image-space fitting via a modification as follows:

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}} \sum_{vt} \left[ s \left( \mathbf{f}_{EM}^{(k)} \ln \mathbf{f}(\theta) - \mathbf{f}(\theta) \right) \right]_{vt} - \beta U(\theta) \tag{48}$$

which is effectively a non-linear generalization of the linear-modeled generalized EM for MAP estimation (Hebert and Leahy 1989).

### 7.3. Approaches to implementation

We can now review various implementations for stage 2, as this constitutes the more challenging step, with stage 1 involving nothing more complicated than a series of conventional 3D EM image updates.

**7.3.1. Linear model.** Reverting back briefly to the use of linear  $\theta$  parameters: in this case the objective of equation (47) can of course be maximized through iterative use of an EM-type algorithm, with its linear forward model. This approach actually corresponds to the nested EM algorithm. One can proceed with as many or as few image-space EM updates as desired, recalling that equation (47) does not necessarily have to be maximized, before updating again from the raw measured tomographic emission data (equation (46)). So the task specified by equation (47) is achieved by:

$$\theta^{(k+1)} = \frac{\theta^{(k)}}{\mathbf{B}^T \mathbf{1}} \mathbf{B}^T \frac{\mathbf{f}_{EM}^{(k)}}{\mathbf{B} \theta^{(k)}} \tag{49}$$

where the general model  $\mathbf{f}(\boldsymbol{\theta})$  has been replaced by an explicitly linear model, matrix  $\mathbf{B}$  operating on the parameter vector  $\boldsymbol{\theta}$ .

**7.3.2. Non-linear models.** For non-linear parameters  $\boldsymbol{\theta}$ , which is the primary concern here for kinetic parameter estimation, a number of options have now been proposed and practically implemented in the literature for the stage 2 optimization.

The first approach is to use a weighted least squares approximation (Matthews *et al* 2010). This is motivated by examining the form of the objective function (47) – it corresponds to minimizing the KL distance measure which also naturally arises from the Poisson log-likelihood, as was shown in equations (27) to (29). Given that a Poisson distribution for a sufficiently large mean value can be well approximated by a Gaussian (with matching mean and variance) then one could, as an alternative to the KL distance, use a Gaussian log-likelihood. This corresponds to a least squares distance, weighted by the modeled mean. However, the modeled mean depends on the very estimate being sought, so instead the estimate from the previous iteration is used for the weights in the weighted least squares (WLS) objective (hence it can be referred to as ‘one step late’):

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{vt} \left[ \frac{\mathbf{s}(\mathbf{f}_{EM}^{(k)} - \mathbf{f}(\boldsymbol{\theta}))^2}{\mathbf{f}(\boldsymbol{\theta}^{(k)})} \right]_{vt} \quad (50)$$

where, as usual, vector multiplication (including squaring) and division is understood to be component-wise. The benefit of this approach is that existing non-linear weighted least squares algorithms can be used to solve (50), as an approximate alternative to solving (47). For example the generalized linear least squares (GLLS) method has been successfully applied (Angelis *et al* 2014, Kotasidis *et al* 2012). An alternative approximation would be to use  $\mathbf{f}_{EM}^{(k)}$  instead of  $\mathbf{f}(\boldsymbol{\theta}^{(k)})$  for the weights.

In contrast, Wang and Qi tackle the stage 2 non-linear optimization of equation (47) directly, using a modified Levenberg Marquardt algorithm (Marquardt 1963, Wang and Qi 2012) and other earlier work, picking up the proposal of EMPIRA, also used the method of Marquardt (Chiao *et al* 1994).

Of course, many other algorithm options exist for the problem of maximizing equation (47), with for example a Newton-Raphson approach as originally suggested for this problem by Carson and Lange in 1985 (Carson and Lange 1985), and very recent work on 4D reconstruction has proposed using an EM-inspired preconditioned gradient ascent approach (Rahmim *et al* 2014).

It is interesting to note that prior to their 2012 OTEM work, Wang and Qi had previously proposed another algorithm (Wang and Qi 2009a), which used paraboloidal surrogate functions for the penalized log-likelihood. Since their paraboloidal surrogates method was surpassed by their OTEM method (which is identical in principle to the method previously proposed by Matthews *et al* and covered above), we do not consider it further here. An excellent review of OTEM compared to the paraboloidal surrogate method is given in Wang and Qi (2012), where it is also compared with the one step late approximation shown in equation (50). A further review is available in their review article (Wang and Qi 2013).

We will now move on to consider a derivation of the above two stage approach to non-linear parameter estimation for PET and SPECT, showing how the non-linear image-space update arises naturally from a so-called complete data framework, which will be covered in the next section. Before doing so, we briefly note here that there are other recent examples of algorithms which also handle the non-linear parameter estimation problem through a complete data description, also using EM. Specifically, the very first and original work of Snyder (Snyder 1984), the work of Yan *et al* (2012) and more recently that of Su *et al* (2013). One particular choice of complete data for the EM framework, and how it results in the image-space objective function given by (47), will be presented next.

#### 7.4. Derivation of the EM approach for non-linear parameters

The aforementioned algorithms rely heavily on the EM framework. Given the scarcity of derivations of the EM algorithm in the literature for both PET and SPECT, it is beneficial for this review to provide not only a derivation of the EM framework for the 4D case, but also for the case of non-linear parameters. The derivation given here is in the original spirit of the EM algorithm, which has its origins in an intuitive approach for problems in areas such as genetics, and which was subsequently shown to have theoretically appealing properties (Dempster *et al* 1977, Laird 2010).

In essence one first *designs* a set of data which would make the parameter estimation problem very simple indeed. These data are referred to as complete data, and there are often many ways in which these data can be formulated, according even to the creativity of the researcher. Good examples in the field of 4D image reconstruction for the case of kinetic parameter estimation can be seen in the work of Schottlander *et al* (2006) (who followed the approach of Snyder (1984)) and Yan *et al* (2012) (who devised a novel formulation of complete data). The key point is to formulate a complete data set that makes the problem very simple to solve. The EM approach then alternates between two key steps: expectation, and then maximization. The expectation step involves generating a complete dataset (formally the conditional expectation of the complete data), based on a current estimate of the parameters ( $\theta^{(k)}$ ) and based on the actual measured data ( $\mathbf{m}$ ). The maximization step involves using this generated complete data set (which, remember, has been deliberately designed to make the optimization very simple), to find a new estimate of the parameters which optimizes the desired objective (e.g. which maximizes the likelihood). With the new estimate of the parameters, the expectation step can then be applied again to create new complete data, the objective function maximized again, and so on. Note that this EM framework is a special case of a more general approach known as optimization transfer. In optimization transfer one seeks to optimize a simpler (surrogate) objective function, which when optimized yields a parameter estimate which is closer to maximizing (or minimizing) the actually desired objective function. In this context, the EM method uses the conditional expectation of the complete data Poisson log-likelihood as a surrogate function for the measured data Poisson log-likelihood, as will be seen below.

The reason for needing a method like EM is that a closed form solution for finding the maximum of the Poisson log-likelihood (for the case of estimating many parameters  $\theta$  from many measurements  $\mathbf{m}$ ) is not possible for any realistic choice of the system matrix  $\mathbf{A}$ . So, how can we make it easier to find the maximum of the Poisson log-likelihood, when our parameters to estimate are the mean number of emissions from each voxel for each time frame, and all we have is the number of counts in each sinogram bin for each time frame? Well it turns out that the maximum likelihood estimate can be found in one very simple step if we were to be in possession of a Poisson-distributed complete data set,  $\mathbf{z} = \{z_{ivt}\}$ , which indicates exactly how many counts ( $z$ ) detected in sinogram bin  $i$  came from voxel  $v$  at time index  $t$ . If we had a dataset like that, then it is easy to show that maximizing the Poisson log-likelihood is achieved by just summing the appropriate counts for each voxel, and then normalizing for the number of contributions using the sensitivity image.

Imagine for now that we do have access to such a complete data set (we will deal with the fact that don't have this data later on). First it is useful to relate the designed complete data to the actually measured data  $m_{it}$  (which can now be called 'incomplete data'). Since  $z_{ivt}$  is like having counts in a single sinogram bin  $i$  with individual labels indicating which voxel  $v$  they came from, we can relate them back to  $m_{it}$  just by summing over all the voxels which contributed to sinogram bin  $i$  at time frame  $t$ :

$$m_{it} = \sum_v z_{ivt} \quad (51)$$

It then follows that the mean of  $m_{it}$ , which we model by equation (21) (here we will ignore  $\rho$ ), is

$$\bar{m}_{it} = q_{it}(\boldsymbol{\theta}) = \sum_v a_{it,v} f_{vt}(\boldsymbol{\theta}) = \sum_v \bar{z}_{ivt} \quad (52)$$

and so the model of the mean of a complete data element  $z_{ivt}$  is given by:

$$\bar{z}_{ivt} = a_{it,v} f_{vt}(\boldsymbol{\theta}) \quad (53)$$

Now, we will re-write the Poisson posterior probability (previously given in equation (32) for the incomplete data  $m_{it}$ ) using the complete data:

$$\lambda^{(MAP)}(\boldsymbol{\theta} | \mathbf{z}) = \prod_{t=1}^T \prod_{v=1}^V \prod_{i=1}^I \frac{\exp[-a_{it,v} f_{vt}(\boldsymbol{\theta})] [a_{it,v} f_{vt}(\boldsymbol{\theta})]^{z_{ivt}} \exp[-\beta U(\boldsymbol{\theta})]}{z_{ivt}! Z} \quad (54)$$

The log-posterior for the complete data is therefore (ignoring the constant terms which will not affect the subsequent maximization with respect to  $\boldsymbol{\theta}$ ):

$$\ln \lambda^{(MAP)}(\boldsymbol{\theta} | \mathbf{z}) = \sum_{t=1}^T \sum_{v=1}^V \sum_{i=1}^I (z_{ivt} \ln [a_{it,v} f_{vt}(\boldsymbol{\theta})] - a_{it,v} f_{vt}(\boldsymbol{\theta})) - \beta U(\boldsymbol{\theta}) \quad (55)$$

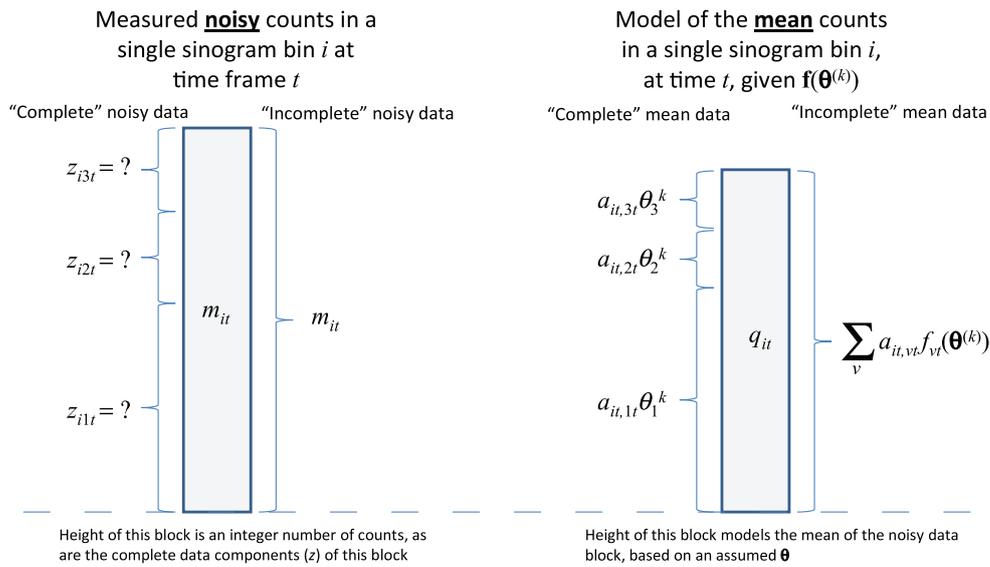
At this point we note that in the majority of the literature a linear model is assumed for  $f_{vt}(\boldsymbol{\theta})$ , and equation (55) is therefore relatively easy to maximize for a given element of  $\boldsymbol{\theta}$  by setting the partial derivative equal to zero and solving. When  $\beta = 0$  the solution is very simple (which was the whole point of using complete data), but for non-zero  $\beta$  either a quadratic needs to be solved (e.g. Levitan and Herman 1987, Depierro 1995) or else a previous estimate of  $\boldsymbol{\theta}$  needs to be used in  $U$  (e.g. the one step late method (Green 1990)).

However, before going further, we recall that we don't actually have the desired complete data  $\mathbf{z}$ . This is where the EM method comes in. In the expectation step we generate an estimate of the complete data, based on the measured data  $\mathbf{m}$  and a current estimate of the parameters  $\boldsymbol{\theta}^{(k)}$ . In essence, by claiming to know already the parameters  $\boldsymbol{\theta}$ , and having the data  $\mathbf{m}$ , it is straightforward to find what the corresponding elements of  $\mathbf{z}$  would have to be. Specifically, we can come up intuitively with the following ratio (see the explanation in figure 7), which requires that the complete-data fractionation of the incomplete data be equal for both the noisy components (left hand side of the below equation) and the modeled mean components (right hand side of the below equation):

$$\frac{z_{ivt}(\boldsymbol{\theta}^{(k)})}{m_{it}} = \frac{a_{it,v} f_{vt}(\boldsymbol{\theta}^{(k)})}{\sum_v a_{it,v} f_{vt}(\boldsymbol{\theta}^{(k)})} \quad (56)$$

As figure 7 demonstrates, by knowing the decomposition of the mean of each element  $q_{it}$  of the incomplete data, we can apply exactly the same fractionation to each element  $m_{it}$  of the noisy incomplete data to obtain an estimate of  $z_{ivt}$ :

$$z_{ivt}(\boldsymbol{\theta}^{(k)}) = \frac{a_{it,v} f_{vt}(\boldsymbol{\theta}^{(k)})}{\sum_v a_{it,v} f_{vt}(\boldsymbol{\theta}^{(k)})} m_{it} \quad (57)$$



**Figure 7.** Graphical explanation of how to obtain expected complete data  $z$  (for the case of just 3 voxels), given an example measured number of counts  $m$  for a bin  $i$ , and under the assumption that the estimate of the parameters,  $\theta^{(k)}$ , is correct. The complete data decomposes the measured incomplete data into its constituent parts, indicating which voxel  $v$  gave rise to which counts in bin  $i$ . By examining the fractionation of the model mean counts,  $q_{it}$ , one can apply the very same fractionation to the actual measured counts to obtain an expected, usually non-integer,  $z$ .

Alternatively, equation (57) can be seen as taking the noisy number of counts  $m_{it}$  for a sinogram bin  $i$  at time  $t$ , and dividing these counts amongst all the possible voxels which could have contributed to  $m_{it}$ . These intuitive perspectives on forming an estimated complete dataset are entirely consistent with the intuitive origins of EM. However, formally, (57) is the conditional expectation of the complete data, given  $\theta^{(k)}$  and  $\mathbf{m}$ , and it should be noted that (57), being a conditional expectation, is unlikely to deliver an integer value for a given  $z_{ivt}$ .

Now we have all the components for EM. The conditional expectation of the complete data for the current parameter estimate  $\theta^{(k)}$  (equation (57), (E-step)) as well as a forward model that can model the mean of the complete data (equation (53)) for any choice of  $\theta$ . Hence, parameters  $\theta$  for the forward model are sought so as to maximise the objective function of equation (55) (M-step), which is relatively easy to do with complete data.

By substituting the conditional expectation of the complete data (57) into our complete-data objective function (55) we obtain:

$$\ln \lambda^{(MAP)}(\theta|\mathbf{z}) = \sum_{t=1}^T \sum_{v=1}^V \left( \ln [f_{vt}(\theta)] \sum_{i=1}^I z_{ivt}(\theta^{(k)}) + \sum_{i=1}^I z_{ivt}(\theta^{(k)}) \ln [a_{it,vt}] - f_{vt}(\theta) \sum_{i=1}^I a_{it,vt} \right) - \beta U(\theta) \quad (58)$$

It is helpful to note, in terms of the terminology used in the literature, that equation (58) can be considered either as i) the conditional expectation of the complete data log-likelihood (log-posterior probability for  $\beta > 0$ ), or ii) the complete data log-likelihood (posterior probability) using the conditional expectation of the complete data given  $\theta^{(k)}$  and  $\mathbf{m}$ . Noting that the second

term in the brackets on the right hand side of (58) is independent of  $\theta$  (and so is a constant that will not affect the value of  $\theta$  which maximizes (58)) and factoring out the time-dependent sensitivity images  $\left(\sum_{i=1}^I a_{it,vt}\right)$  gives:

$$\ln \lambda^{(MAP)}(\theta | \mathbf{z}) = \sum_{t=1}^T \sum_{v=1}^V \left( \sum_{i=1}^I a_{it,vt} \left( \ln [f_{vt}(\theta)] \frac{\sum_{i=1}^I z_{ivt}(\theta^{(k)})}{\sum_{i=1}^I a_{it,vt}} - f_{vt}(\theta) \right) \right) - \beta U(\theta) + const. \tag{59}$$

Equation (59) contains a ratio term which will be recognizable as the very well-known EM image reconstruction update, using the current image estimate  $\mathbf{f}$  generated from  $\theta^{(k)}$  (whether via a linear or non-linear model):

$$\frac{\sum_i z_{ivt}(\theta^{(k)})}{\sum_{i=1}^I a_{it,vt}} = \frac{f_{vt}(\theta^{(k)})}{\sum_{i=1}^I a_{it,vt}} \sum_i a_{it,vt} \frac{m_{it}}{\sum_v a_{it,vt} f_{vt}(\theta^{(k)})} \tag{60}$$

where equation (57) has been used to obtain the right hand side of (60). Given the centrality of (60) to much of the 3D image reconstruction literature, and hence its widespread adoption in both PET and SPECT image reconstruction, we will specially label it as the EM update image:

$$f_{vt}^{(EM)}(\theta^{(k)}) = \frac{f_{vt}(\theta^{(k)})}{\sum_{i=1}^I a_{it,vt}} \sum_i a_{it,vt} \frac{m_{it}}{\sum_v a_{it,vt} f_{vt}(\theta^{(k)})} \tag{61}$$

where (EM) has been placed in superscript to allow subscript indices to be shown, but the image is the same as the vector  $\mathbf{f}_{EM}^{(k)}$ . Equation (61) provides a time series of 3D EM image updates, and the regular time-independent 3D EM update is found for the special case of discarding the time frame index  $t$ . For most cases in the literature the image  $f_{vt}(\theta)$  is given directly by the contents of  $\theta$ , but importantly in this formulation  $f_{vt}(\theta)$  could be that simple or indeed a far more involved function of non-linear parameters  $\theta$ . Hence we can now rewrite equation (59) as:

$$\ln \lambda^{(MAP)}(\theta | \mathbf{z}) = \sum_{t=1}^T \sum_{v=1}^V \left( \sum_{i=1}^I a_{it,vt} [f_{vt}^{(EM)}(\theta^{(k)}) \ln [f_{vt}(\theta)] - f_{vt}(\theta)] \right) - \beta U(\theta) + const. \tag{62}$$

This indicates that we can obtain an improved parameter estimate  $\theta^{(k+1)}$  by maximizing (62)

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}} \sum_{t=1}^T \sum_{v=1}^V s_{vt} [f_{vt}^{(EM)}(\theta^{(k)}) \ln [f_{vt}(\theta)] - f_{vt}(\theta)] - \beta U(\theta) \tag{63}$$

which is none other than the previous equation (48), where we have identified the time-dependent sensitivity images as:

$$s_{vt} = \sum_{i=1}^I a_{it,vt} \tag{64}$$

To within a constant, the function being maximized in equation (63) is the conditional expectation of the dynamic complete data Poisson log likelihood (or log posterior if  $\beta > 0$ ). It can also be

viewed as a surrogate function in optimization transfer terminology- optimizing (63) is simpler than directly optimizing the incomplete data version of the Poisson log posterior probability. Furthermore, the function being maximized in (63) is recognizable, by comparison with the previous equations (28) and (29), as none other than the negative of the KL distance between the modeled image  $f_{vt}(\theta)$  and the regular tomographic EM update image  $f^{(EM)}_{vt}$ . This is an extremely important result, that has in effect been shown a number of times by many researchers in the literature already, dating back to Carson in 1985. As such, equation (63) is a very general and powerful formulation for 4D / direct image reconstruction in PET and SPECT imaging, unifying much of the EM literature in both 3D and 4D PET, as demonstrated in table 4.

The following important observations about the generalized equation (63) (with (61) or (46)) and links to the reconstruction literature need to be made:

- 1) Both ML and MAP estimation are covered by equations (63) and (61), for a linear or non-linear parameterization of the time-series of images  $\mathbf{f}(\theta)$ , accommodating also a time-variant sensitivity image. Carson's EMPIRA accommodates time-dependent sensitivity images (Carson and Lange 1985), which can be required for SPECT.
- 2) When  $\beta = 0$  and when  $\mathbf{f}$  is the most simple linear case of  $\mathbf{f}(\theta^{(k)}) = \theta^{(k)}$ , then  $\mathbf{f}^{EM}$  is the one step maximizer  $\theta^{(k+1)}$  of (63), meaning that just regular MLEM is carried out for each time frame, as would be expected. When  $\beta > 0$ , the generalized EM MAP formulations developed in the 1980s and 1990s are obtained (see Qi and Leahy (2006)).
- 3) When  $\beta = 0$  and a general linear model of  $\mathbf{f}(\theta)$  is used, such as  $\mathbf{f}(\theta^{(k)}) = \mathbf{B}\theta^{(k)}$ , then the EM algorithm can iteratively be applied to progressively and monotonically decrease the KL distance of (63). This is the nested EM method for 4D (Wang and Qi 2010) or 3D (Angelis *et al* 2013).
- 4) Equation (63) is *completely free from any tomographic aspect*: just a time series of sensitivity images and a time series of regular 3D EM updates are needed. This has considerable practical significance. The system model, the huge time-series of Poisson distributed acquired data, along with the estimates of scatter and randoms, do not even appear in equation (63), since they are all handled by the regular EM update of (46). This conveniently isolates any non-linear aspects, and any regularization (MAP) requirement, into *just an image-space fitting process* requiring minimization of a KL distance.

### 7.5. Model-specific algorithms

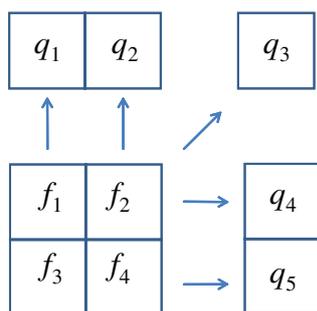
The aforementioned approach based on an EM framework is a relatively straightforward and adaptable method in the literature for non-linear parameter estimation in emission tomography. Nonetheless, other methods have been proposed (notably Kamasak *et al* (2005), Yan *et al* (2012), and more recently Su *et al* (2013)), but with the caveat that they have been presented in model specific form—for specific choices of the non-linear parameters  $\theta$ . The method of Yan *et al* relies on a 1 tissue compartment model, and the method of Kamasak *et al* is presented for a 2 tissue compartment model and uses a coordinate descent method, although the method should be equally applicable to any non-linear model. Excellent reviews of these methods are already given in (Wang and Qi 2013).

### 7.6. Early methods developed for SPECT

In terms of practical implementations for estimating non-linear parameters directly from SPECT projection data, pioneering work was in fact carried out in the 1990s (Zeng *et al* 1995,

**Table 4.** Summary of methods in the literature, from 3D to 4D, which all use the framework presented by equation (63) in conjunction with the conventional EM update image (61).

Image series model $f_{vt}(\theta^{(k)})$	EM image formed by	Method and reference	Use of (63) (conditional expectation of complete data log likelihood or posterior)
$\mathbf{f}(\theta^{(k)}) = \theta^{(k)}$	$\mathbf{f}^{EM} = \frac{\theta^{(k)}}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A} \theta^{(k)}}$	Standard 3D or independent frame 4D ML-EM ( $\beta=0$ ) Generalized EM ( $\beta > 0$ ) (Hebert and Leahy 1989)	Maximize (63) directly by differentiating and setting to zero: $\theta^{(k+1)} = \mathbf{f}^{EM}$ Maximization of (63) not necessary, but just an increase by coordinate ascent
$\mathbf{f}(\theta^{(k)}) = \mathbf{B}\theta^{(k)}$	$\mathbf{f}^{EM} = \frac{\theta^{(k)}}{\mathbf{B}^T \mathbf{A}^T \mathbf{1}} \mathbf{B}^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A} \mathbf{B} \theta^{(k)}}$	OSL-EM (Green 1990)	Maximize (63) by $\theta^{(k+1)} = \frac{\mathbf{f}^{EM}}{\mathbf{1} + \frac{\beta}{s} U^T(\theta^{(k)})}$
$\mathbf{f}(\theta^{(k)}) = \mathbf{H}\theta^{(k)}$ where $\mathbf{H}$ is a matrix modeling resolution (e.g. a convolution)	$\mathbf{f}^{EM} = \frac{\mathbf{H}\theta^{(k)}}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A} \mathbf{H} \theta^{(k)}}$	EM-MAP (Leviton and Herman 1987, Depierro 1995)	Maximize by finding the root of a quadratic function of $\theta$
General linear $f_{vt}(\theta) = [\mathbf{B}\theta]_{vt}$	$\mathbf{f}^{EM} = \frac{\mathbf{B}\theta^{(k)}}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A} \mathbf{B} \theta^{(k)}}$	4D MLEM (Matthews <i>et al</i> 1997)	Maximize (63) directly by differentiating and setting to zero: $\theta^{(k+1)} = \mathbf{f}^{EM}$
Any non-linear model $f_{vt}(\theta)$	$\mathbf{f}^{EM} = \frac{\mathbf{f}(\theta^{(k)})}{\mathbf{A}^T \mathbf{1}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A} \mathbf{f}(\theta^{(k)})}$	3D nested MLEM, (Angelis <i>et al</i> 2013)	Do one or more image space EM iterations to increase (63)
		4D nested EM (Wang and Qi 2010)	$\theta^{(k+1)} = \frac{\theta^{(k)}}{\mathbf{H}^T \mathbf{1}} \mathbf{H}^T \frac{\mathbf{f}^{EM}}{\mathbf{H} \theta^{(k)}}$
		One step late weighted least squares (Matthews <i>et al</i> 2010)	Do one or more image space EM iterations to increase (63) $\theta^{(k+1)} = \frac{\mathbf{f}^{EM}}{\mathbf{B}^T \mathbf{1}} \mathbf{B} \theta^{(k)}$
			Carry out a weighted least squares fit in image space to approximately maximize (63)
			$\theta^{(k+1)} = \operatorname{argmin}_{\theta} \sum_{vt} \left[ \frac{\mathbf{s}(\mathbf{f}^{EM} - \mathbf{f}(\theta))^2}{\mathbf{f}^{EM}} \right]_{vt}$
		EMPIRA (Carson and Lange 1985)	Newton Raphson suggested for maximizing (63)
		OTEM (Wang and Qi 2012)	Modified Levenberg Marquardt proposed for maximization of (63)



**Figure 8.** Spatial relation between the 4 pixel values  $f_1, f_2, f_3, f_4$  and the 5 projection bins  $q_1, q_2, q_3, q_4$  and  $q_5$ , assuming a simple line integral model. Note that each projection bin contains 26 time frames, and the overall goal is to estimate kinetic parameters in the 4 pixel locations.

Huesman *et al* 1998, Reutter *et al* 1998, Gullberg *et al* 1999). In these early approaches, the one tissue compartmental model is used to describe the activity concentrations. The model of the projection data therefore depends on the kinetic parameters ( $F_v, K_1$  and  $k_2$ ) for every  $j$ -th basis function. To reduce the dimensionality only a limited number of spatial basis functions, typically 4, were used (effectively corresponding to ROIs of different tissue regions). For example 13 parameters (4 regions  $\times$  3 kinetic parameters, and a background scaling factor) were estimated from the projection data in Zeng *et al* (1995), Huesman *et al* (1998), Reutter *et al* (1998). The weighted sum of square differences between modeled and measured data is then minimized using standard optimization strategies (e.g. the Newton-Raphson method was used in Reutter *et al* (1998)). However these approaches have not been applied to large scale problems (e.g. voxel-level / parametric methods) and represent rather specially developed cases for SPECT. A model with two exponentials was used by Limber *et al* (1995), for direct parametric reconstruction into a very limited number of pixels.

## 8. Simple simulation examples

In this section a simple simulation example is presented to illustrate the benefits of the 4D reconstruction method reviewed in the previous section. The simplified imaging system (figure 8) consists of just 5 projection data bins ( $I=5$ ) and there are 26 time frames ( $T=26$ ). The spatial basis functions chosen are 4 pixels ( $J=V=4$ ), with 26 time frames of varying duration to cover 3450 s. For conventional reconstruction (with a post-reconstruction estimation of kinetic parameters) this corresponds to a TAC defined by 26 parameters estimated for each pixel (so  $P=26$ ). For direct reconstruction this will be reduced, as explained below, to finding just 3 parameters ( $P=3$ ) to define the TAC at each pixel. To simulate idealized data, two different system matrices were considered for relating the 4 pixels to the 5 data bins, a sparse matrix  $A_S$  and a denser matrix  $A_D$ :

$$A_S = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad A_D = \begin{bmatrix} 1 & 0.5 & 1 & 0.5 \\ 0.5 & 1 & 0.5 & 1 \\ 0.5 & 1 & 1 & 0.5 \\ 1 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 1 \end{bmatrix} \quad (65)$$

Note these  $I \times J$  system matrices are used for each time frame of data, so the system is assumed to not change with time. With the more densely populated matrix  $A_D$  there will be far more significant spatial noise correlations in the reconstructed image, which is potentially a key benefit of direct reconstruction methods as these correlations are usually ignored in post-reconstruction kinetic analysis. Four ground truth TACs were generated, one for each pixel, based on the SRTM, with choices of  $(R_1, k_2, BP)$  given by: (0.87, 0.27, 3.72), (0.90, 0.35, 2.52), (0.84, 0.17, 1.8) and (0.80, 0.3, 4.02), respectively. The reference region TAC was based on one obtained from a real [ $^{11}\text{C}$ ]raclopride scan, and the first set of SRTM parameters (0.87, 0.27, 3.72) was likewise based on a fit to the striatal region of a real [ $^{11}\text{C}$ ]raclopride scan.

Four main simulations were then performed, corresponding to the use of the two different matrices  $A_S$  and  $A_D$ , and the use of two noise levels (low noise data, so that the reconstructed pixel TACs resemble ROI TACs from a real full size dynamic 3D reconstruction, and high noise data, so that the reconstructed TACs resemble single pixel TACs from a full size reconstruction, figure 9). For each simulation the ground truth TACs were forward projected by the system matrix, and Poisson noise was introduced. A total of 500 realizations of noisy data were considered for each case. The data were then reconstructed into 4 pixels of 26 time frames using i) a conventional approach (considering up to a total of 160 iterations of MLEM, followed by post-reconstruction kinetic fitting to find the binding potential, BP, for each pixel) and using ii) the direct reconstruction approach (equations (46) and (47) given in the previous section). The kinetic fitting for both the post-reconstruction approach as well as the direct approach used a 1 gigabyte dictionary of TACs, generated for a range of sample  $R_1, k_2$  and BP values. A given pixel TAC  $\mathbf{f}_j = \{f_{jt}, t=1 \dots T\}$  was fitted by considering each possible modeled TAC  $\mathbf{g}$  from the dictionary, simply evaluating the objective function for each  $\mathbf{g}$ . So the fitted values for pixel  $j$  are found according to the maximum value of this objective:

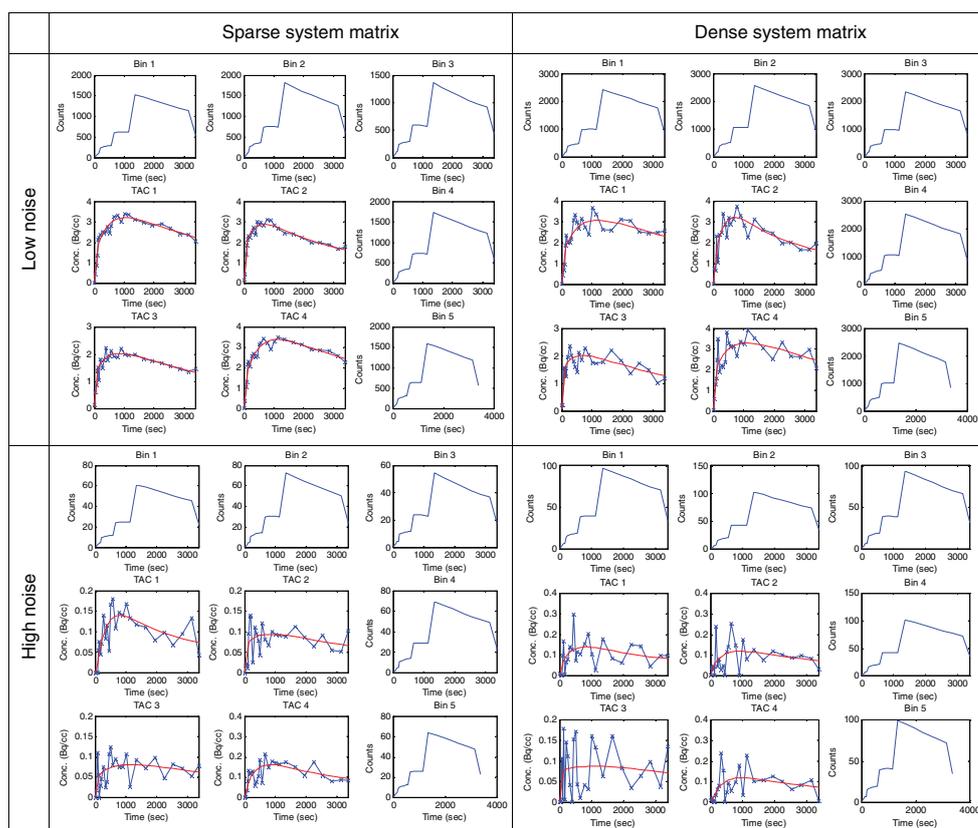
$$\begin{bmatrix} R_1^{FIT} \\ k_2^{FIT} \\ BP^{FIT} \end{bmatrix}_j = \underset{R_1, k_2, BP}{\operatorname{argmax}} \sum_{t=1}^T (f_{jt} \ln g_{jt}(R_1, k_2, BP) - g_{jt}(R_1, k_2, BP)) \quad (66)$$

However, whilst this dictionary approach is fine for single post-reconstruction kinetic fitting, it was not found suitable for the direct reconstruction method, as very small changes in the pixel TACs result in identical dictionary fits, meaning that the direct algorithm can no longer effectively update the TACs and the kinetic parameter estimates. For this reason a line search in each of the three kinetic parameters was also performed, within the range of the dictionary precision, so as to permit subtle changes in the fit parameters.

Explicitly then, the direct reconstruction proceeded as follows:

1. Uniform image initialization (all TACs set to 1 for all pixels)
2. One regular EM-type update of the TAC values (independent frame EM updates) based on the measured data (equation (61))
3. Fitting of the TACs based on the objective function in equation (66) (equation (63) with  $\beta = 0$ , and time-invariant sensitivity)
4. Using the fitted TACs as the new TAC values for the 4 pixels
5. Looping back to step 2

Figure 10 shows the root mean square error in the BP estimates for all pixels, for the four main simulations. As might be expected, the direct reconstruction method offers increasing levels of error reduction as noise in the data increases (where it is important to note that for a fixed count level, use of a denser system matrix increases noise in the reconstructed images, and so direct reconstruction offers benefits for this case also). An interesting point to note is that

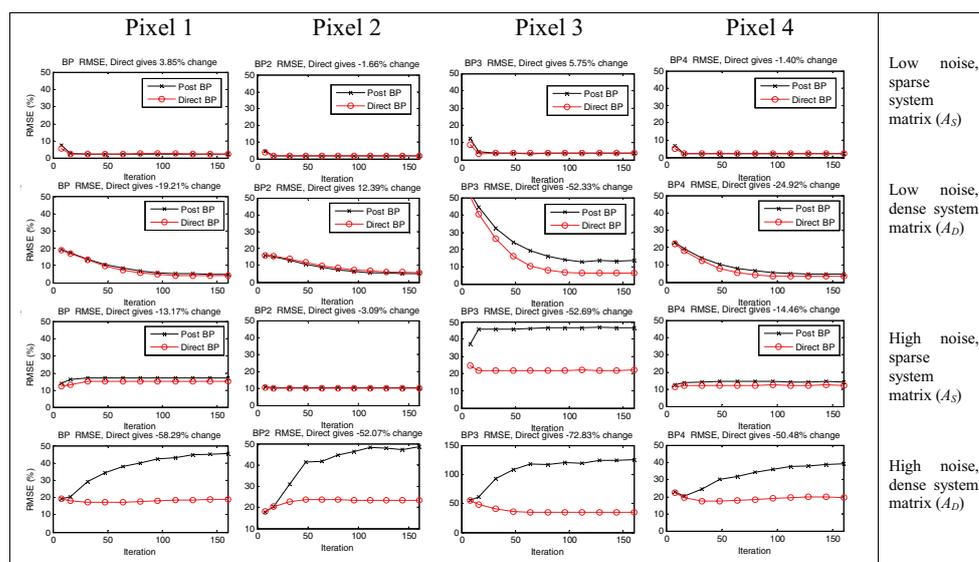


**Figure 9.** Example reconstructed TACs (joined crosses) in the four pixels and example fits (continuous solid line, based on equation (66)) for the two noise levels considered and for the two system matrices considered. Note that the TACs in the pixels are corrected for frame length and calibrated to give Bq/mL, but the raw counts in the five projection bins are not corrected for frame length.

for low noise (and low spatial noise correlations) the direct method essentially matches the conventional method, when using the same objective function for the kinetic fitting.

## 9. Performance benefits of 4D: brief summary of results from the literature

We now consider the benefits of 4D reconstruction compared to conventional 3D methodology as reported in a variety of implementations from the literature. The literature is diverse in terms of the simulation studies performed to assess these methods, in terms of different levels of statistical quality of the emission data (number of counts in the list-mode or projection data), the particular radiotracer and distribution considered, the voxel/pixel/ROI sizes, the particular parameters estimated and the algorithm used. Despite this diversity and all these differences, most research articles do nonetheless have the common goal of comparing fully 4D or direct kinetic parameter estimation with the conventional method of independent-frame reconstruction followed by post-reconstruction kinetic modeling. In this section we present typically reported levels of reduction in error (root mean square error (RMSE), mean absolute error (MAE)) or the typical levels of reduction in standard deviation (SD) for a given level of



**Figure 10.** Root mean square error (RMSE) for BP estimates in the four pixels, as a function of iteration for MLEM with post-reconstruction kinetic modeling (Post BP) and for the direct MLEM method (Direct BP). The four main simulation cases are shown (low noise and high noise, low (sparse matrix) and high (dense matrix) levels of spatial correlations in the reconstructions). 500 realizations were generated for each simulation.

bias compared to the case of conventional methodology. It is of course important to note that these different metrics reflect different performance parameters (RMSE and MAE encompass both noise and bias, whereas SD indicates only noise). Therefore the performance differences between 3D and 4D reconstruction presented in this section are not necessarily directly comparable between publications, but do nonetheless allow relative claims of improvement to be compared.

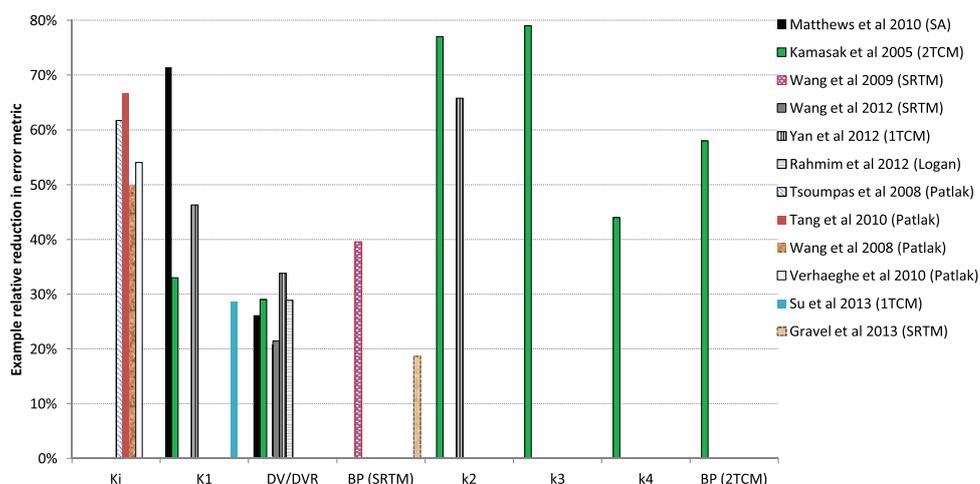
Figure 11 summarizes the reported benefits for the PET literature. The results include linear parameter estimation (Tsoumpas *et al* 2008b, Wang *et al* 2008, Matthews *et al* 2010, Tang *et al* 2010, Verhaeghe *et al* 2010, Rahmim *et al* 2012) and non-linear parameter estimation (Kamasak *et al* 2005, Wang and Qi 2009b, Wang and Qi 2012, Yan *et al* 2012, Su *et al* 2013, Gravel and Reader 2013). The work by Matthews *et al* (2010) considers simulated [ $^{11}\text{C}$ ]diprenorphine brain data and reports on RMSE reduction for the macro-parameters  $K_1$  and DV estimated with direct reconstruction using a spectral analysis model. The performance improvement of direct reconstruction compared to conventional indirect reconstruction is compared after 400 iterations in the figure. (Kamasak *et al* 2005) reports on direct parametric reconstruction using a full two-tissue compartmental model. Simulated [ $^{11}\text{C}$ ]raclopride brain data is presented where the direct reconstruction provides a reduction in normalized RMSE in the parameters  $K_1$ - $k_4$  as well as in BP and DV. The errors in BP and DV estimated from the full two-tissue compartmental model were even lower compared to using the linear Logan graphical model simplification applied after reconstruction. (Wang and Qi 2009b) considers the simplified reference model and SD versus bias trade off curves are shown for simulated brain data with the tracer [ $^{11}\text{C}$ ]DASB, so figure 11 reports a typical SD reduction at fixed bias. In (Wang and Qi 2012) the spatial background standard deviation versus ROI mean value is shown for the DV ratio (DVR) using the SRTM for real brain data with the tracer [ $^{11}\text{C}$ ]SCH23390.

In Yan *et al* (2012) the one-tissue compartmental model is used for direct 4D parametric reconstruction of simulated data with kinetics corresponding to the tracer [ $^{11}\text{C}$ ]P943, and the averaged reduction in the coefficient of variation (COV) for  $K_1$ ,  $k_2$  and DV is reported in figure 11. The greater COV reduction was found for the lowest count levels, as would be expected. Note also that their reported noise reductions were in some cases accompanied by a simultaneous bias reduction. Rahmim *et al* (2012) report COV reduction for simulated brain data using the Logan linear model for [ $^{11}\text{C}$ ]raclopride. Results of direct parametric reconstruction of the  $K_i$  macro-parameter using the linear Patlak model and simulated brain data for [ $^{18}\text{F}$ ]FDG can be found in Tsoumpas *et al* (2008b) (averaged reduction in RMSE is shown in figure 11), Tang *et al* (2010) (reduction in normalized standard deviation is shown) and Wang *et al* (2008) (reduction in standard deviation is shown in figure 11). The reduction in MAE for post-reconstruction Patlak graphical analysis after 4D reconstruction using spectral analysis basis functions for simulated brain [ $^{18}\text{F}$ ]FDG data can be found in Verhaeghe *et al* (2009). Finally figure 11 also includes results for heart imaging with [ $^{13}\text{N}$ ]NH<sub>3</sub> (Su *et al* 2013) and also results from simulated [ $^{11}\text{C}$ ]raclopride data for the HRRT brain PET scanner (Gravel and Reader 2013).

In summary, these broad-ranging evaluations report error reductions in the range of 20% to 80%, in all cases showing at least some benefit arising from 4D or direct reconstruction. The benefits are of course largely dependent on the noise level and noise correlations in the reconstructed data under consideration in each case, as in fact benefits can become negligible in the presence of high count data (as indicated in the simple simulations in the previous section) or for large ROIs. To complement the summary of improvements shown in figure 11, figure 12 shows example images of the visual improvement in image quality obtained with 4D reconstruction compared to conventional post-reconstruction modeling. Finally, figure 13 demonstrates the progressive advantages of improved modeling for parametric  $K_i$  images: from regular 3D OSEM, to inclusion of resolution modeling, to introducing data-derived temporal basis functions (Reader *et al* 2006a) and then finally with the use of data-derived spectral-analysis basis functions proposed in Reader *et al* (2007).

## 10. Conclusions and future directions

Fully 4D and direct parametric image reconstruction methods can reduce the impact of noisy PET or SPECT emission data on end-point parameter estimates. The techniques presented in this review emphasize the direct use of the raw PET or SPECT data, for which the primary advantage is that the noise model is well known to be Poisson, contrary to the case of post-reconstruction parameter estimation methods which often approximate the unknown noise distribution of the images being fitted. The direct methods not only use more accurate noise modeling, but also implicitly account for spatial noise correlations and impose the temporal modeling constraints throughout the iterative reconstruction procedure. Furthermore, the methods are not limited to kinetic parameter estimation, but in fact virtually any post-reconstruction parameter estimation task can be converted to a direct parameter estimation task, thereby correctly modeling the noise in the raw data. Conceptually, direct linear or non-linear parameter estimation is not difficult to implement, thanks to the two stage process now proposed by multiple researchers in the field. This nested methodology conveniently separates the linear tomographic problem from the potentially non-linear image-space modeling problem. Provided the KL distance measure is used as the objective function for the image-space modeling (i.e. for the fitting of the parameterized model to a current 3D EM image update), then the parameter estimation can be appropriately performed with full account for



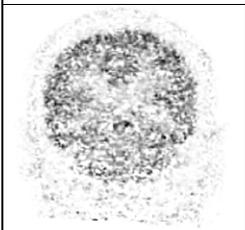
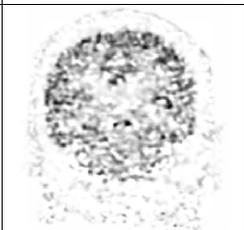
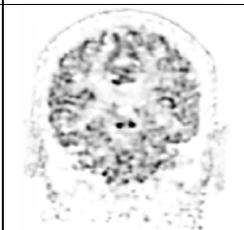
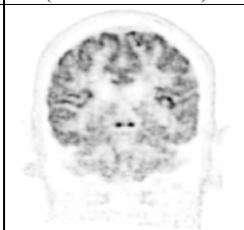
**Figure 11.** Summary of reported relative benefits of 4D or direct reconstruction compared to conventional independent-frame reconstruction with post-fitting of the images. Results from 12 papers in terms of reduction in standard deviation (for a given bias) or the reduction in root mean square error (or broadly comparable error measure) are shown, including linear and non-linear parameter estimation methods.

<i>Patlak intercept: c</i> (Verhaeghe <i>et al</i> 2010)		<i>Radiotracer influx: K<sub>i</sub></i> (Tsoumpas <i>et al</i> 2008b)	
3D	4D	3D	4D

**Figure 12.** Example parametric images for FDG PET imaging from the literature using simulated data. Left: parametric images for the Patlak intercept, *c*, when using conventional 3D OSEM reconstruction with 8 iterations (16 subsets) and then post-fitting and when using data-derived spectral analysis basis functions for 4D OSEM reconstruction with 40 iterations and then post-fitting. Right: direct 4D reconstruction using Patlak basis functions for FDG (Tsoumpas *et al* 2008b), shown for the same number of OSEM iterations (480 iterations, 12 subsets) for both 3D and 4D.

the Poisson-distributed emission data. Of course though, this parameter fitting does need to be carried out for each and every EM update, rather than just once in the case of post-reconstruction fitting.

To handle list-mode data, rather than sinogram or projection data, the nested method presented in this review would need to use a list-mode EM algorithm for the 3D image frames update in stage 1 of the method (for example the early work on list-mode EM by (Parra and Barrett 1998) and (Reader *et al* 1998)) rather than a projection-data based EM algorithm (the example shown in figure 6 of this review). The future will see not only direct use of raw PET

3D OSEM	3D OSEM+PSF	4D OSEM+PSF (data-derived temporal basis functions) (Reader <i>et al</i> 2006a)	4D OSEM+PSF (data-derived spectral analysis temporal basis functions) (Reader <i>et al</i> 2007)
			

**Figure 13.** Progressive visual improvement in post-fitted parametric image quality (for the tracer influx  $K_i$ ) arising from improved spatial and then improved temporal modeling for real FDG PET data. HRRT list-mode data from a 60 min [ $^{18}\text{F}$ ]FDG (184 MBq) study of a probable Alzheimer's patient was reconstructed and then post-fitted for  $K_i$  at the voxel level for four methods, using a population input function. From left to right: conventional independent frames 3D OSEM, 3D OSEM with resolution modeling (point spread function (PSF) modeling), 4D OSEM with basis functions estimated from the data, and finally 4D OSEM using a data-derived surrogate input function with spectral analysis basis functions. The core components of the dynamic EM reconstruction algorithm used for the HRRT list-mode data (with scatter and randoms corrections, attenuation and normalization, resolution modeling kernel and subsets) are detailed in previous work (Reader *et al* 2006a). The last two columns correspond to using the two cases previously presented in the penultimate row of table 1.

and SPECT data, whether in projection data format or list-mode format, but probably also direct use of raw simultaneously-acquired MR data. This will allow dual-modality tracers and contrast agents to be used in a single scanning session, combining both data streams in algorithms to estimate directly the parameters of interest. The reconstruction will therefore be able to incorporate information on subject anatomy and motion, along with cardiac and respiratory motion, generating 5D or 6D images. There is even scope for estimating the motion from the acquired data, including motion fields amongst the parameters to be estimated during image reconstruction. Also, with the noise reduction achieved by direct methods, it is conceivable that these methods can aid signal separation for more ambitious multi-radiotracer PET and/or SPECT studies, which when further combined with MR contrast agents and even multi-parametric MR imaging (e.g. Ma *et al* (2013)) could provide interesting potential for near simultaneous imaging of a useful range of physiological and anatomical parameters.

However some possible concerns with direct parameter estimation methods for PET and SPECT have been identified. Firstly, if a model is inappropriate then errors can occur which exceed those of a post-reconstruction estimation approach, as in 4D and direct reconstruction the errors propagate to other regions in the image throughout the iterative reconstruction procedure. Nonetheless, there is promising work indicating this can be accounted for through use of residual modeling terms (Kotasidis *et al* 2014). A further concern might be the assessment of the fit quality of a given kinetic model: often in post-reconstruction kinetic parameter estimation one has estimates of the fit quality, but direct methods do not implicitly deliver this information. The work of Kamasak *et al* (2014) has started to address this issue, essentially exploiting a backprojection of the difference between the measured sinograms and the kinetically modeled sinograms. Given these possible concerns, the significance of the end point impact on actual studies will also be necessary to evaluate in order to see if the benefits

outweigh any possible drawbacks. Recent work has indicated that the benefits under certain conditions (e.g. [ $^{11}\text{C}$ ]raclopride imaging (Gravel and Reader 2013)) are relatively modest compared to other study types. Yet even if the end point benefits of these 4D and direct estimation methods are not always substantial (there is clear variation in the level of benefit reported in the previous figure 11), in general provided the modeling is accurate, they can only serve to help, since they represent the theoretically correct way of performing many tasks which are currently carried out on a post-reconstruction basis. They can even apply, for example, to such apparently straightforward tasks as post-reconstruction image registration (Gravel *et al* 2013).

## Acknowledgments

This review and research was undertaken, in part, thanks to funding from the Canada Research Chairs program. The authors gratefully acknowledge the use of time activity curve data supplied by Paul Gravel for the simulation section, and previous informal discussions with Julian Matthews regarding the expectation maximization algorithm. Finally, the authors acknowledge Karl Herholz and Matthew Jones (University of Manchester, Wolfson Molecular Imaging Centre) for the raw PET list-mode data acquired with the HRRT used for figure 13.

## References

- Alenius S and Ruotsalainen U 1997 Bayesian image reconstruction for emission tomography based on median root prior *Eur. J. Nucl. Med.* **24** 258–65
- Andersen A H and Kak A C 1984 Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm *Ultrason. Imag.* **6** 81–94
- Angelis G I *et al* 2011a The performance of monotonic and new non-monotonic gradient ascent reconstruction algorithms for high-resolution neuroreceptor PET imaging *Phys. Med. Biol.* **56** 3895–917
- Angelis G I *et al* 2011b Convergence optimization of parametric MLEM reconstruction for estimation of Patlak plot parameters *Comput. Med. Imag. Graph.* **35** 407–16
- Angelis G I, Matthews J C, Kotasidis F A, Markiewicz P J, Lionheart W R and Reader A J 2014 Evaluation of a direct 4D reconstruction method using generalised linear least squares for estimating nonlinear micro-parametric maps *Ann. Nucl. Med.* **1**–14
- Angelis G I *et al* 2013 Acceleration of image-based resolution modelling reconstruction using an expectation maximization nested algorithm *Phys. Med. Biol.* **58** 5061–83
- Barrett H H and Myers K J 2004 *Foundations of Image Science* (New York: Wiley)
- Bauschke H H *et al* 1999 An EM algorithm for dynamic SPECT *IEEE Trans. Med. Imag.* **18** 252–61
- Bentourkia M and Zaidi H 2007 Tracer kinetic modeling in PET *PET Clinics* **2** 267–77
- Blinder S *et al* 2003 Fully 4D dynamic image reconstruction by nonlinear constrained programming *IEEE Nuclear Science Symp. Conf. Record.* **5** 3161–65
- Bousse A *et al* 2012 Monotonic algorithm for joint entropy-based anatomical priors in parametric PET image reconstruction *IEEE Nuclear Science Symp. and Medical Imaging Conf. Record (NSS/MIC)* **3918**–24
- Bowsher J E *et al* 1996 Bayesian reconstruction and use of anatomical a priori information for emission tomography *IEEE Trans. Med. Imag.* **15** 673–86
- Byrne C L 1998 Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods *IEEE Trans. Image Process.* **7** 100–09
- Carson R E 1986 A maximum-likelihood method for region-of-interest evaluation in emission tomography *J. Comput. Assist. Tomogr.* **10** 654–63
- Carson R E and Lange K 1985 A statistical model for positron emission tomography—the EM parametric image reconstruction algorithm—comment *J. Am. Stat. Assoc.* **80** 20–2
- Celler A *et al* 2000 Performance of the dynamic single photon emission computed tomography (dSPECT) method for decreasing or increasing activity changes *Phys. Med. Biol.* **45** 3525–43

- Chiao P C *et al* 1994 Model-based estimation for dynamic cardiac studies using ECT *IEEE Trans. Med. Imag.* **13** 217–26
- Comtat C *et al* 1998 Fast reconstruction of 3D PET data with accurate statistical modeling *IEEE Trans. Nucl. Sci.* **45** 1083–89
- Cunningham V J and Jones T 1993 Spectral-analysis of dynamic PET studies *J. Cereb. Blood Flow Metab.* **13** 15–23
- de Jong H W A M *et al* 2007 Performance evaluation of the ECAT HRRT: an LSO-LYSO double layer high resolution high sensitivity scanner *Phys. Med. Biol.* **52** 1505–26
- Defrise M, Clack R and Townsend D W 1995 Image-reconstruction from truncated 2-dimensional parallel projections *Inverse Problems* **11** 287–313
- Delso G *et al* 2011 Performance measurements of the siemens mMR integrated whole-body PET/MR scanner *J. Nucl. Med.* **52** 1914–22
- Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc. Ser. B-Methodol.* **39** 1–38 ([www.jstor.org/stable/2984875](http://www.jstor.org/stable/2984875))
- Depierro A R 1995 A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography *IEEE Trans. Med. Imag.* **14** 132–37
- Fahey F H 2002 Data acquisition in PET imaging *J. Nucl. Med. Technol.* **30** 39–49 (PMID:12055275)
- Farncombe T *et al* 2001 The incorporation of organ uptake into dynamic (SPECT dSPECT) image reconstruction *IEEE Trans. Nucl. Sci.* **48** 3–9
- Gjedde A 1995 Origins of the patlak plot *Nucl. Med. Commun.* **16** 979–80
- Gravel P, Verhaeghe J and Reader A J 2013 3D PET image reconstruction including both motion correction and registration directly into an MR or stereotaxic spatial atlas *Phys. Med. Biol.* **58** 105–26
- Gravel P and Reader A J 2013 Direct 4D PET MLEM reconstruction of parametric images using the simplified reference tissue model with the basis function method *IEEE and Medical Imaging Conf. Nuclear Science Symp. (NSS/MIC)* pp 1–7
- Gravier E *et al* 2006 Fully 4D motion-compensated reconstruction of cardiac SPECT images *Phys. Med. Biol.* **51** 4603–19
- Green P J 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm *IEEE Trans. Med. Imag.* **9** 84–93
- Grezes-Besset L *et al* 2007 Simulation-based evaluation of NEG-ML iterative reconstruction of low count PET data *IEEE Nuclear Science Symp. Conf. Record* **4** 3009–14
- Grotus N *et al* 2009 Fully 4D list-mode reconstruction applied to respiratory-gated PET scans *Phys. Med. Biol.* **54** 1705–21
- Gullberg G T, Huesman R H, Di Bella E V, Zeng G L, Reutter B W, Christian P E and Foresti S A 1999 Dynamic cardiac single-photon emission computed tomography *Nuclear Cardiology: State of the Art and Future Directions* ed B L Zaret and G A Beller (St Louis, MO: Mosby) pp137–87
- Gullberg G T *et al* 2010 Dynamic single photon emission computed tomography—basic principles and cardiac applications *Phys. Med. Biol.* **55** R111–R91
- Gunn R N, Gunn S R and Cunningham V J 2001 Positron emission tomography compartmental models *J. Cereb. Blood Flow Metab.* **21** 635–52
- Gunn R N *et al* 1997 Parametric imaging of ligand-receptor binding in PET using a simplified reference region model *Neuroimage* **6** 279–87
- Hebber E *et al* 1997 Direct estimation of dynamic parameters in SPECT tomography *IEEE Trans. Nucl. Sci.* **44** 2425–30
- Hebert T and Leahy R 1989 A generalized EM algorithm for 3-d bayesian reconstruction from poisson data using Gibbs priors *IEEE Trans. Med. Imag.* **8** 194–202
- Herholz K, Herscovitch P and Heiss W-D 2004 *Neuro PET: PET in Neuroscience and Clinical Neurology* (Berlin: Springer)
- Hong I and Reader A J 2008 Ultrafast 4D PET image reconstruction with user-definable temporal basis functions *Nucl. Sci. Symp. Conf. Record* pp 5475–8
- Huesman R H *et al* 1998 Kinetic parameter estimation from SPECT cone-beam projection measurements *Phys. Med. Biol.* **43** 973–82
- Innis R B *et al* 2007 Consensus nomenclature for *in vivo* imaging of reversibly binding radioligands *J. Cereb. Blood Flow Metab.* **27** 1533–39
- Jacobson M W and Fessler J A 2006 Joint estimation of respiratory motion and activity in 4D PET using CT side information 3rd *IEEE Int. Symp. on Biomedical Imaging: Nano to Macro* **1–3** 275–78

- Jones T and Rabiner E A 2012 The development past achievements and future directions of brain PET *J. Cereb. Blood Flow Metab.* **32** 1426–54
- Judenhofer M S *et al* 2007 PET/MR images acquired with a compact MR-compatible PET detector in a 7-T magnet *Radiology* **244** 807–14
- Kadrmas D J and Gullberg G T 2001 4D maximum *a posteriori* reconstruction in dynamic SPECT using a compartmental model-based prior *Phys. Med. Biol.* **46** 1553–74
- Kamasak M E *et al* 2005 Direct reconstruction of kinetic parameter images from dynamic PET data *IEEE Trans. Med. Imag.* **24** 636–50
- Kamasak M E *et al* 2014 Quality and precision of parametric images created from PET sinogram data by direct reconstruction: proof of concept *IEEE Trans. Med. Imag.* **33** 695–707
- Kotasidis F A *et al* 2014 Application of adaptive kinetic modeling for bias propagation reduction in direct 4D image reconstruction *Phys. Med. Biol.* **59** 6061–84
- Kotasidis F A *et al* 2012b Direct parametric reconstruction for dynamic [F-18]-FDG PET/CT imaging in the body *IEEE Nuclear Science Symp. and Medical Imaging Conf. Record (NSS/MIC)* 3383–86
- Kotasidis F A *et al* 2011 Single scan parameterization of space-variant point spread functions in image space via a printed array: the impact for two PET/CT scanners *Phys. Med. Biol.* **56** 2917–42
- Laird N M 2010 The EM algorithm in genetics genomics and public health *Stat. Sci.* **25** 450–57
- Lammertsma A A and Hume S P 1996 Simplified reference tissue model for PET receptor studies *Neuroimage* **4** 153–58
- Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306–16
- Leahy R M and Qi J Y 2000 Statistical approaches in quantitative positron emission tomography *Stat. Comput.* **10** 147–65
- Lee T S, Segars W P and Tsui B M W 2005 Study of parameters characterizing space-time Gibbs priors for 4D MAP-RBI-EM in gated myocardial perfusion SPECT *IEEE Nuclear Science Symp. Conf. Record.* **4** 2124–8
- Levitani E and Herman G T 1987 A maximum *a posteriori* probability expectation maximization algorithm for image-reconstruction in emission tomography *IEEE Trans. Med. Imag.* **6** 185–92
- Lewitt R M 1992 Alternatives to voxels for image representation in iterative reconstruction algorithms *Phys. Med. Biol.* **37** 705–16
- Lewitt R M and Matej S 2003 Overview of methods for image reconstruction from projections in emission computed tomography *Proc. IEEE* **91** 1588–611
- Li Q Z *et al* 2007 A fast fully 4-D incremental gradient reconstruction algorithm for list mode PET data *IEEE Trans. Med. Imag.* **26** 58–67
- Li T F *et al* 2006 Model-based image reconstruction for four-dimensional PET *Med. Phys.* **33** 1288–98
- Limber M A *et al* 1995 Direct reconstruction of functional parameters for dynamic spect *IEEE Trans. Nucl. Sci.* **42** 1249–56
- Logan J 2000 Graphical analysis of PET data applied to reversible and irreversible tracers *Nucl. Med. Biol.* **27** 661–70
- Ma D *et al* 2013 Magnetic resonance fingerprinting *Nature* **495** 187–92
- Maltz J S 2000 Direct recovery of regional tracer kinetics from temporally inconsistent dynamic ECT projections using dimension-reduced time-activity basis *Phys. Med. Biol.* **45** 3413–29
- Maltz J S 2001 Optimal time-activity basis selection for exponential spectral analysis: application to the solution of large dynamic emission tomographic reconstruction problems *IEEE Trans. Nucl. Sci.* **48** 1452–64
- Markiewicz P J *et al* 2005 An advanced analytic method incorporating the geometrical properties of scanner and radiation emissions into the system model for the true component of 3D PET data *IEEE Nuclear Science Symp. Conf. Record.* 42310–4
- Markiewicz P J *et al* 2007 High accuracy multiple scatter modelling for 3D whole body PET *Phys. Med. Biol.* **52** 829–47
- Marquardt D W 1963 An algorithm for least-squares estimation of nonlinear parameters *J. Soc. Indust. Appl. Math.* **11** 431–41
- Matthews J C *et al* 1997 The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction *Phys. Med. Biol.* **42** 1155–73
- Matthews J C *et al* 2010 Direct reconstruction of parametric images using any spatiotemporal 4D image based model and maximum likelihood expectation maximisation *IEEE Nuclear Science Symp. Conf. Record (NSS/MIC)* 2435–41

- Mc Ardle B A *et al* 2012 Does rubidium-82 PET have superior accuracy to SPECT perfusion imaging for the diagnosis of obstructive coronary disease? A systematic review and meta-analysis *J. Am. College Cardiol.* **60** 1828–37
- McLennan A and Brady S M 2009 Dynamic PET reconstruction algorithms using empirical mode decomposition regularisation *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro* **871–74**
- McLennan A and Brady M 2010 Direct pharmacokinetic parameter estimation using weighted least squares *Proc. SPIE Medical Imaging: Physics of Medical Imaging* 762224
- Meikle S R *et al* 1998 Parametric image reconstruction using spectral analysis of PET projection data *Phys. Med. Biol.* **43** 651–66
- Muehllehner G and Karp J S 2006 Positron emission tomography *Phys. Med. Biol.* **51** R117–R37
- Nichols T E, Qi J Y and Leahy R M 1999 Continuous time dynamic PET imaging using list mode data *Information Processing in Medical Imaging Lecture Notes in Computer Science*: **1613** 98–111
- Nichols T E *et al* 2002 Spatiotemporal reconstruction of list-mode PET data *IEEE Trans. Med. Imag.* **21** 396–404
- Nuyts J, Stroobants S, Dupont P, Vleugels S, Flamen P and Mortelmans L 2002 Reducing loss of image quality because of the attenuation artifact in uncorrected PET whole-body images *J. Nucl. Med.* **43** 1054–62
- Panin V Y *et al* 2006 Fully 3-D PET reconstruction with system matrix derived from point source measurements *IEEE Trans. Med. Imag.* **25** 907–21
- Parra L and Barrett H H 1998 List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET *IEEE Trans. Med. Imag.* **17** 228–35
- Patlak C S, Blasberg R G and Fenstermacher J D 1983 Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data *J. Cereb. Blood Flow Metab.* **3** 1–7
- Qi J Y and Leahy R M 2006 Iterative reconstruction techniques in emission computed tomography *Phys. Med. Biol.* **51** R541–R78
- Rahmim A, Tang J and Zaidi H 2009 Four-dimensional (4D) image reconstruction strategies in dynamic PET: beyond conventional independent frame reconstruction *Med. Phys.* **36** 3654–70
- Rahmim A, Qi J and Sossi V 2013 Resolution modeling in PET imaging: theory practice benefits and pitfalls *Med. Phys.* **40** 064301
- Rahmim A *et al* 2012 Direct 4D parametric imaging for linearized models of reversibly binding PET tracers using generalized AB-EM reconstruction *Phys. Med. Biol.* **57** 733–55
- Rahmim A, Tang J and Mohy-ud-Din H 2014 Direct 4D parametric imaging in dynamic myocardial perfusion PET *Front. Biomed. Technol.* **1** 4–13
- Rakvongthai Y *et al* 2013 Direct reconstruction of cardiac PET kinetic parametric images using a preconditioned conjugate gradient approach *Med. Phys.* **40** 102501
- Razifar P *et al* 2006 Volume-wise application of principal component analysis on masked dynamic PET data in sinogram domain *IEEE Trans. Nucl. Sci.* **53** 2759–68
- Reader A J and Zaidi H 2007 Advances in PET image reconstruction *PET Clinics* **2** 173–90
- Reader A J *et al* 1998 Fast accurate iterative reconstruction for low-statistics positron volume imaging *Phys. Med. Biol.* **43** 835–46
- Reader A J *et al* 2003 EM algorithm system modeling by image-space techniques for PET reconstruction *IEEE Trans. Nucl. Sci.* **50** 1392–97
- Reader A J *et al* 2006a Joint estimation of dynamic PET images and temporal basis functions using fully 4D ML-EM *Phys. Med. Biol.* **51** 5455–74
- Reader A J, Matthews J C, Sureav F C, Comtat C, Trébossen R and Buvat I 2006b Iterative kinetic parameter estimation within fully 4D PET image reconstruction *IEEE Nuclear Science Symp. Conf. Record* **3** 1752–6
- Reader A J *et al* 2007 Fully 4D image reconstruction by estimation of an input function and spectral coefficients *IEEE Nuclear Science Symp. Conf. Record* **5** 3260–7
- Reilhac A *et al* 2008 Simulation-based evaluation of OSEM iterative reconstruction methods in dynamic brain PET studies *Neuroimage* **39** 359–68
- Reutter B W, Gullberg G T and Huesman R H 1998 Kinetic parameter estimation from attenuated SPECT projection measurements *IEEE Trans. Nucl. Sci.* **45** 3007–13
- Reutter B W, Gullberg G T and Huesman R H 2000 Direct least-squares estimation of spatiotemporal distributions from dynamic SPECT projections using a spatial segmentation and temporal B-splines *IEEE Trans. Med. Imag.* **19** 434–50

- Reutter B W *et al* 2005 Improved quantitation of dynamic SPECT via fully 4-D joint estimation of compartmental models and blood input function directly from projections *IEEE Nuclear Science Symp. Conf. Record* **4** 2337–41
- Schirmmacher R *et al* 2013 Design of brain imaging agents for positron emission tomography: do large bioconjugates provide an opportunity for *in vivo* brain imaging? *future Med. Chem.* **5** 1621–34
- Schlemmer H P W *et al* 2008 Simultaneous MR/PET imaging of the human brain: feasibility study *Radiology* **248** 1028–35
- Schottlander D *et al* 2006 Preliminary evaluation of a kinetic parameter estimator with application to direct parametric reconstruction *3rd IEEE Int. Symp. on Biomedical Imaging: Nano to Macro* 932–35
- Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imag.* **1** 113–22
- Siddon R L 1985 Fast calculation of the exact radiological path for a 3-dimensional CT array *Med. Phys.* **12** 252–55
- Sitek A, Bella D and Gullberg C T 1999 Factor analysis of dynamic structures in dynamic SPECT imaging using maximum entropy *IEEE Trans. Nucl. Sci.* **46** 2227–32
- Sitek A, Di Bella E and Gullberg C T 2000 Factor analysis with a priori knowledge—application in dynamic cardiac SPECT *Phys. Med. Biol.* **45** 2619–38
- Sitek A, Gullberg G T, Di Bella E and Cellar A 2001 Reconstruction of dynamic renal tomographic data acquired by slow rotation *J. Nucl. Med.* **42** 1704–12
- Snyder D L 1984 Parameter-estimation for dynamic studies in emission-tomography systems having list-mode data *IEEE Trans. Nucl. Sci.* **31** 925–31
- Snyder D L and Miller M I 1985 The use of sieves to stabilize images produced with the EM algorithm for emission tomography *IEEE Trans. Nucl. Sci.* **32** 3864–72
- Su K H, Yen T C and Fang Y H 2013 A novel approach for direct reconstruction of parametric images for myocardial blood flow from PET imaging *Med. Phys.* **40** 102505
- Tang J *et al* 2008 Direct 4D reconstruction of parametric images incorporating anato-functional joint entropy *IEEE Nuclear Science Symp. Conf. Record* 5471–4
- Tang J *et al* 2010 Direct 4D reconstruction of parametric images incorporating anato-functional joint entropy *Phys. Med. Biol.* **55** 4261–72
- Tsoumpas C, Turkheimer F E and Thielemans K 2008a A survey of approaches for direct parametric image reconstruction in emission tomography *Med. Phys.* **35** 3963–71
- Tsoumpas C, Turkheimer F E and Thielemans K 2008b Study of direct and indirect parametric estimation methods of linear models in dynamic positron emission tomography *Med. Phys.* **35** 1299–309
- Tsoumpas C *et al* 2013 The effect of regularization in motion compensated PET image reconstruction: a realistic numerical 4D simulation study *Phys. Med. Biol.* **58** 1759–73
- Turkheimer F E *et al* 2003 A linear wavelet filter for parametric imaging with dynamic PET *IEEE Trans. Med. Imag.* **22** 289–301
- Verhaeghe J and Reader A J 2010 AB-OSEM reconstruction for improved Patlak kinetic parameter estimation: a simulation study *Phys. Med. Biol.* **55** 6739–57
- Verhaeghe J and Reader A J 2013 Simultaneous water activation and glucose metabolic rate imaging with PET *Phys. Med. Biol.* **58** 393–411
- Verhaeghe J *et al* 2007 Reconstruction for gated dynamic cardiac PET imaging using a tensor product spline basis *IEEE Trans. Nucl. Sci.* **54** 80–91
- Verhaeghe J *et al* 2008 Dynamic PET reconstruction using wavelet regularization with adapted basis functions *IEEE Trans. Med. Imag.* **27** 943–59
- Verhaeghe J *et al* 2009 Motion-compensated fully 4D PET reconstruction using PET data supersets *IEEE Nuclear Science Symp. Conf. Record (NSS/MIC)* 3000–04
- Verhaeghe J *et al* 2010 Motion compensation for fully 4D PET reconstruction using PET superset data *Phys. Med. Biol.* **55** 4063–82
- Wang G and Qi J 2010 Acceleration of the direct reconstruction of linear parametric images using nested algorithms *Phys. Med. Biol.* **55** 1505–17
- Wang G and Qi J 2012 An optimization transfer algorithm for nonlinear parametric image reconstruction from dynamic PET data *IEEE Trans. Med. Imag.* **31** 1977–88
- Wang G, Fu L and Qi J 2008 Maximum *a posteriori* reconstruction of the patlak parametric image from sinograms in dynamic PET *Phys. Med. Biol.* **53** 593–604
- Wang G and Qi J 2009a Generalized algorithms for direct reconstruction of parametric images from dynamic PET data *IEEE Trans. Med. Imag.* **28** 1717–26

- Wang G and Qi J 2009b Direct reconstruction of PET receptor binding parametric images using a simplified reference tissue model *Proc. SPIE Medical Imaging* **7258** 72580V
- Wang G and Qi J 2013 Direct estimation of kinetic parametric images for dynamic PET *Theranostics* **3** 802–15
- Watson C C 2009 Signal-to-noise ratio equalized filtered back-projection for positron emission tomography *IEEE Trans. Nucl. Sci.* **56** 108–15
- Wernick M N, Infusino E J and Milosevic M 1999 Fast spatio-temporal image reconstruction for dynamic PET *IEEE Trans. Med. Imag.* **18** 185–95
- Yan J H, Planeta-Wilson B and Carson R E 2008 Direct 4D list mode parametric reconstruction for PET with a novel EM algorithm *IEEE Nuclear Science Symp. and Medical Imaging Conf. Record* **3625–8**
- Yan J H, Planeta-Wilson B and Carson R E 2012 Direct 4-D PET list mode parametric reconstruction with a novel EM algorithm *IEEE Trans. Med. Imag.* **31** 2213–23
- Yan J *et al* 2009 Initial evaluation of direct 4D parametric reconstruction with human PET data *IEEE Nuclear Science Symp. Conf. Record (NSS/MIC) IEEE* **2503–6**
- Yetik I S and Qi J Y 2006 Direct estimation of kinetic parameters from the sinogram with an unknown blood function *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro* **1–3** 295–8
- Zeng G L, Gullberg G T and Huesman R H 1995 Using linear time-invariant system theory to estimate kinetic parameters directly from projection measurements *IEEE Trans. Nucl. Sci.* **42** 2339–46
- Zeng G L *et al* 1991 3-Dimensional iterative reconstruction algorithms with attenuation and geometric point response correction *IEEE Trans. Nucl. Sci.* **38** 693–702
- Zhao H and Reader A J 2003 Fast ray-tracing technique to calculate line integral paths in voxel arrays *IEEE Nuclear Science Symp. Conf. Record* **4** 2808–12