IEEE Transactions on Medical Imaging, 14(4):669-680, December 1995

Bayesian Image Reconstruction in SPECT Using Higher Order Mechanical Models as Priors *

Soo-Jin Lee[#], Anand Rangarajan[§], and Gene Gindi[#¶]
[#]Department of Electrical Engineering, SUNY at Stony Brook, NY 11794-2350
[§]Department of Computer Science, Yale University, New Haven, CT 06520-2158
[¶]Department of Radiology, SUNY at Stony Brook, NY 11794-8460

Corresponding author: Gene Gindi
Department of Radiology
State University of New York
Stony Brook, NY 11794-8460
Phone: (516) 444-2539
E-mail: gindi@clio.rad.sunysb.edu

*S. J. Lee was supported in part by a Student Fellowship from the Education and Research Foundation of the Society of Nuclear Medicine. This work was supported in parts by the Stony Brook Foundation, the University Hospital Consortium, and NIH R01 NS32879 from the National Institute of Neurological Disorders and Stroke.

Abstract

While the ML-EM algorithm for reconstruction for emission tomography is unstable due to the ill-posed nature of the problem, Bayesian reconstruction methods overcome this instability by introducing prior information, often in the form of a spatial smoothness regularizer. More elaborate forms of smoothness constraints may be used to extend the role of the prior beyond that of a stabilizer in order to capture actual spatial information about the object. Previously proposed forms of such prior distributions were based on the assumption of a piecewise constant source distribution. Here, we propose an extension to a piecewise *linear* model – the weak plate – which is more expressive than the piecewise constant model. The weak plate prior not only preserves edges but also allows for piecewise ramplike regions in the reconstruction. Indeed, for our application in SPECT, such ramplike regions are observed in ground-truth source distributions in the form of primate autoradiographs of rCBF radionuclides. To incorporate the weak plate prior in a MAP approach, we model the prior as a Gibbs distribution and use a GEM formulation for the optimization. We compare quantitative performance of the ML-EM algorithm, a GEM algorithm with a prior favoring piecewise constant regions, and GEM algorithm with our weak plate prior. Pointwise and regional bias and variance of ensemble image reconstructions are used as indications of image quality. Our results show that the weak plate and membrane priors exhibit improved bias and variance relative to ML-EM techniques.

Key words: SPECT, Bayesian Reconstruction, Gibbs Priors, Weak Plate

1 Introduction

The maximum-likelihood (ML) approach using the expectation maximization (EM) algorithm has enjoyed continuing interest in emission tomography. However, due to the ill-posed nature of the reconstruction problem, the ML-EM method is known to be unstable and diverges in a least squares sense at higher iteration numbers. In contrast, Bayesian reconstruction methods in emission tomography are attractive in that they overcome this instability while retaining an important advantage of the ML-EM approach, notably a principled means of incorporating the imaging model. These methods extend ML approaches via maximum *a posteriori* (MAP) approaches.

Prior information serves as a regularizer for EM, but one might also view the use of priors as a means of incorporating actual, known information regarding the local spatial character of the source. Several such prior distributions have been proposed in the literature [1, 2, 3, 4, 5, 6, 7, 8]. In most of these proposals, the model for the object prior was based on assumptions of piecewise smoothness; here, piecewise smoothness corresponds to a notion of nearly constant, within undulations, patches in the object. These assumptions corresponded to the general observation, especially for brain imaging in single-photon emission computed tomography (SPECT) and positron emission tomography (PET), that various pharmaceuticals tend to concentrate somewhat uniformly in grey matter and in white matter and are absent from cerebro-spinal fluid (CSF), with a ratio of concentration approximately 4:1:0 for grey, white, and CSF, respectively [9, 10, 11]. Our own early work incorporated one such prior, a "weak membrane" model, to capture notions of piecewise smoothness [12, 8, 13, 14]. Most of the simulation experiments were done with the mathematical phantoms which were themselves piecewise constant. Reconstruction algorithms, including our own, that use this kind of piecewise constant assumption as prior perform better, according to most reasonable metrics, than those that make no such assumption, when tested on such phantoms. However, the self-consistent loop of piecewise constant prior and piecewise constant phantom leads one to question whether these good results generalize to a realistic clinical setting where the underlying (patient) source distribution may not be piecewise constant.

Here we argue for more expressive priors able to model more complicated forms than a piecewise constant source. To the extent that actual source distributions are more accurately modeled by these more expressive priors, we might expect more accurate reconstructions (in the sense of reduced bias and variance) and, hopefully, better reconstructions in a given task-specific sense. A natural question arises: what is the nature of a typical source structure? Preliminary evidence that the source structure is not piecewise constant is displayed in the form of autoradiograph source distributions of rhesus primates that have undergone testing for radiopharmaceutical development. Empirical evidence for this claim is illustrated for the case of regional cerebral blood flow (rCBF) SPECT in the form of a rhesus autoradiograph (Fig. 1) "ground truth" image of a radionuclide spatial distribution of a blood flow agent ^{99m}Tc-ECD [15]. Fig. 1 (and also Fig. 7) shows that many of the edge regions are not piecewise constant. In this case, the piecewise constant prior may lead to errors.

To overcome the above problems we propose a more accurate model, the *weak plate* prior. The essential feature of our prior model is that it favors piecewise *linear* ramplike regions. In fact, the weak plate is a part of a family of "mechanical" models – weak membrane (1st order), weak plate (2nd order), and weak quadric (3rd order) – in which a class of smoothness constraints derived from properties of ideal physical materials are used as models in the associated reconstruction problem [16]. Smoothness priors often correspond to a notion of fitting a member of a class of smooth functions to the data; "weak" constraints refer to the allowed inclusion of spatial discontinuities in the fit. Thus a weak membrane is a low order (1st derivative) spline allowed to have breaks (0th-order discontinuities) and a weak plate is the familiar (2nd-order) thin plate spline that is allowed to have breaks and creases (discontinuities in 0th and 1st derivatives). The weak plate prior is thus an extension of weak membrane to higher order.

This paper is concerned with the technical means of incorporating a piecewise linear priors, and with performance testing to validate possible advantages of a higher order prior. In Sec. 2 we discuss the mechanical models as priors in detail, then derive a Bayesian reconstruction model using the weak membrane and weak plate priors in Sec. 3. A MAP estimate for this Bayesian model is difficult to compute; in this section we present a deterministic annealing algorithm to address these computational difficulties. In Sec. 4 we present implementational details plus simulation results showing the advantages of our new prior in the sense of reduced bias and variance. We conclude with a discussion in Sec. 5, noting some limitations of this initial study.

2 Mechanical Models as Priors

In this section we discuss qualitatively the nature of the weak membrane (WM) and weak plate (WP) priors and defer to Sec. 3 their incorporation into a formal Bayesian framework. Because our priors take the form of the familiar Gibbs model, they may be expressed as energy functions.

Since our priors encourage the formation of smooth regions punctuated by a locus of discontinuities, we must include a mathematical means to model these discontinuities. Geman and Geman [17] introduced an unobservable line process l into the image model to preserve the discontinuities in the reconstructions, and versions of the line processes have been proposed for medical imaging in [18, 19, 8]. The line processes as proposed in [17] are binary variables (0 or 1) that act to suspend smoothness constraints at sites where they are turned on ($l = 1$). A horizontal line process at location (i, j), $l^h_{ij} = 1$, indicates a horizontal edge (discontinuity between pixels along the vertical direction) and a vertical line process, $l^v_{ij} = 1$, indicates a vertical edge (discontinuity along the horizontal direction). By embedding these line processes into the reconstruction space, the mechanical models for conventional smoothness constraints are extended to include *weak* continuity constraints [20].

We may write the energy for the WM prior as a function of the source, f, and the line processes, l, both defined on a two dimensional (2-D) lattice indexed by i, j. [21].

$$E(f, l) = \lambda \sum_{ij} \left[f_v^2(i,j)(1 - l^h_{ij}) + f_h^2(i,j)(1 - l^v_{ij}) \right] + \alpha \sum_{ij} \left(l^h_{ij} + l^v_{ij} \right). \tag{1}$$

In (1), f_v and f_h are the first order derivatives along the vertical and the horizontal directions, respectively, and are defined as $f_v(i,j) \stackrel{\text{def}}{=} f_{i+1,j} - f_{i,j}$ and $f_h(i,j) \stackrel{\text{def}}{=} f_{i,j+1} - f_{i,j}$. The binary variables l^v and l^h are vertical and horizontal line processes, respectively, and λ and α are positive hyperparameters. Note that if l is set to zero everywhere, the prior reduces to a familiar smoothness regularizing term penalizing the first squared derivatives. Of course, one could set $l = 1$ at all locations to minimize energy, but the last term penalizes the creation of the discontinuities, charging an amount α at each such site. Therefore, the two terms in (1) encourage smoothness except where discontinuities ($l_{ij} = 1$) occur. (As a physical analogy, the elastic membrane is said to be "broken" when $l_{ij} = 1$.) Note that spurious discontinuities in the reconstruction may be generated by noise. If α is close to zero there is little to discourage the formation of spurious discontinuities. But as α increases fewer discontinuities form, and for sufficiently large α the probability of spurious discontinuities becomes negligible. If α is too high, no discontinuities form and the prior again reduces to the simple smoothness constraint. The qualitative effect of the WM prior is to simultaneously detect the discontinuities and convolve the data with a smoothing filter only within continuous regions.

However, due to its nature in favoring piecewise constant reconstructions, the WM prior has the unfortunate effect of turning a ramp into a single step or stepped terraces depending on the filter scale as shown in Figs. 2(b) and (c). This is the "gradient limit" effect [21], which is a fundamental limitation of the WM. If the gradient of a ramp exceeds the limit, a discontinuity appears in its reconstruction. If the gradient is much greater than the limit, a solution with multiple breaks might have the smallest energy. The gradient limit is inversely proportional to the scale λ: the coarse scale (large λ) turns ramps into a few steps (Fig. 2(b)) and the fine scale (small λ) turns ramps into many steps (Fig. 2(c)). As λ is lowered, the ramp is better approximated, but the smoothing action of the prior becomes less effective. This disadvantage of the WM leads, in fact, to significant errors in the SPECT reconstructions as shown in our simulations.

While the WM prior favors the piecewise constant reconstructions, the WP prior favors the piecewise ramplike in addition to piecewise constant reconstructions [22]. Our corresponding energy function for the WP is [21]:

$$E(f,l) = \lambda \sum_{ij} \left\{ \left[f_{vv}^2(i,j) + 2f_{hv}^2(i,j) + f_{hh}^2(i,j) \right] (1 - l_{ij}) \right\} + \alpha \sum_{ij} l_{ij},$$

where our discrete approximations of the partial second derivatives are [23] $f_{vv}(i,j) \stackrel{\text{def}}{=} f_{i+1,j} - 2f_{i,j} + f_{i-1,j}$, $f_{hh}(i,j) \stackrel{\text{def}}{=} f_{i,j+1} - 2f_{i,j} + f_{i,j-1}$, $f_{hv}(i,j) \stackrel{\text{def}}{=} f_{i,j} - f_{i+1,j} - f_{i,j+1} + f_{i+1,j+1}$. Notice that this form has the line processes, l_{ij}, associated with the six nodes, $f_{i-1,j}$, $f_{i,j-1}$, $f_{i,j}$, $f_{i,j+1}$, $f_{i+1,j}$, and $f_{i+1,j+1}$. If l is set to zero everywhere, the prior reduces to the familiar thin plate spline smoothing functional. However, unlike the WM which encourages only the formation of piecewise constant regions, the WP encourages smoothness even in ramplike regions without incurring a penalty. That is, a "crease", a discontinuity in first derivative, will turn on the line process. Thus the discontinuities in the WP correspond to discontinuities in the source gradient in addition to those in the source itself. Of course, one might set $l = 1$ at all locations to minimize energy, but the last term penalizes the creation of the discontinuities, charging in amount of α each discontinuity. The interpretation [21] of line processes in the weak plate is a little different than in the WM; a crease appears as a contour along which the line processes are set, but a step appears as a thick contour of set line processes, two elements wide. This corresponds to the notion that, for a discrete signal, a step discontinuity may be interpreted as two adjacent creases.

3 A Bayesian Reconstruction Model with Weak Plate Prior

We model our problem on a 2-D finite discrete lattice indexed by i, j. Lowercase bold quantities denote 2-D vector fields and the corresponding lowercase, italicized quantities denote the elements of the vector field. Similarly, uppercase bold quantities denote 2-D random fields. Thus, $\Pr(\mathbf{F} = \mathbf{f})$ denotes the probability that the random field $\mathbf{F}$ takes the value $\mathbf{f}$ and f_{ij} denotes an element of $\mathbf{f}$ at the location (i, j). The informal notation in Sec. 2 is thus made precise.

We begin by formulating our reconstruction problem from Bayes' Theorem with the aid of the line processes:

$$\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g}) = \frac{\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})}{\Pr(\mathbf{G} = \mathbf{g})},$$

where $\mathbf{f}$, $\mathbf{l}$, and $\mathbf{g}$ are the source intensities, line processes, and projection data, respectively, and $\mathbf{F}$, $\mathbf{L}$, $\mathbf{G}$ are the associated 2-D random fields. In particular, the 2-D random field $\mathbf{F}$ represents the array of Poisson means of the Poisson emission probability density at each pixel. The term $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ is the prior probability for the fields $\mathbf{F}$ and $\mathbf{L}$, $\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ the likelihood probability, and $\Pr(\mathbf{G} = \mathbf{g})$ a constant. We estimate both the source field and the discontinuity field simultaneously by maximizing the *a posteriori* probability distribution, the MAP estimate $(\hat{\mathbf{f}}, \hat{\mathbf{l}})$:

$$\begin{aligned} (\hat{\mathbf{f}}, \hat{\mathbf{l}}) &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g}) \\ &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}), \end{aligned} \tag{2}$$

where the last equality derives from the fact that $\mathbf{G}$ is independent of $\mathbf{L}$.

3.1 Derivation of an Energy Function for MAP estimation

We model both the likelihood and the prior as Gibbs distributions. A Gibbs distribution is specified by an energy function E and the partition function Z (a suitable normalization).

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp(-E(\mathbf{y})). \tag{3}$$

Since the partition function is a normalization over all possible configurations of $\mathbf{Y}$ and not of $\mathbf{X}$, it is usually a function of $\mathbf{x}$. We describe the process by which an energy function can be obtained from the two terms in (2) constituting the posterior. Since the number of detected counts in SPECT is independently Poisson distributed, we model the likelihood as:

$$\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) = \prod_{t\theta} \frac{\bar{g}_{t\theta}^{g_{t\theta}} \exp(-\bar{g}_{t\theta})}{g_{t\theta}!}, \tag{4}$$

where $\bar{g}_{t\theta} \stackrel{\text{def}}{=} \sum_{ij} \mathcal{H}_{t\theta,ij} f_{ij}$. In (4), $g_{t\theta}$ is the number of detected counts in the detector bin indexed by t at angle θ, $\bar{g}_{t\theta}$ is the expected number of counts for a particular source $\mathbf{f}$, and $\mathcal{H}_{t\theta,ij}$ is the probability that a photon emitted from source location (i, j) hits detector bin t at angle θ. The major physical factors for SPECT, attenuation, scatter, and detector response, can be adequately modeled as linear effects and summarized by the probability matrix $\mathcal{H}_{t\theta,ij}$.

Using (3) and (4), the likelihood energy becomes

$$E_D(\mathbf{f}) = -\log \Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) = \sum_{t\theta} \left[-g_{t\theta} \log\left(\bar{g}_{t\theta}\right)\right] + \sum_{t\theta} \bar{g}_{t\theta},$$

where the term $\log(g_{t\theta}!)$ was dropped since it does not involve $\mathbf{f}$. The prior is defined over the source intensities and line processes. The corresponding prior energy is

$$E_P(\mathbf{f},\mathbf{l}) = -\log \Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})$$

which, for the cases of WP and WM, becomes

$$E_P^P(\mathbf{f},\mathbf{l}) = \lambda \sum_{ij} \left\{ \left[f_{vv}^2(i,j) + 2f_{hv}^2(i,j) + f_{hh}^2(i,j) \right] (1-l_{ij}) \right\} + \alpha \sum_{ij} l_{ij} \tag{5}$$

and

$$E_P^M(\mathbf{f},\mathbf{l}) = \lambda \sum_{ij} \left[f_v^2(i,j)(1-l_{ij}^h) + f_h^2(i,j)(1-l_{ij}^v) \right] + \alpha \sum_{ij} \left(l_{ij}^h + l_{ij}^v \right), \tag{6}$$

respectively. (Equations (5) and (6) repeat previous expressions, but with the understanding that $\mathbf{f}$ and $\mathbf{l}$ are now realizations of associated random fields.) We may thus write the following overall energies by simply taking the $-\log$ of the posterior to get

$$E^P(\mathbf{f},\mathbf{l}) \;=\; E_D(\mathbf{f}) + E_P^P(\mathbf{f},\mathbf{l}) \tag{7}$$

$$E^M(\mathbf{f},\mathbf{l}) \;=\; E_D(\mathbf{f}) + E_P^M(\mathbf{f},\mathbf{l}) \tag{8}$$

for WP and WM, respectively.

As it stands, the energy functions above are difficult to minimize, since they contain mixed binary and continuous variables. Notice that the energy can be minimized by setting

$$l_{ij} = \begin{cases} 0, & U_{ij} \le \frac{\alpha}{\lambda} \\ 1, & U_{ij} > \frac{\alpha}{\lambda} \end{cases}$$

for WP, and

$$l^v_{ij} = \begin{cases} 0, & V^h_{ij} \;\le\; \frac{\alpha}{\lambda} \\ 1, & V^h_{ij} \;>\; \frac{\alpha}{\lambda} \end{cases} \qquad\qquad l^h_{ij} = \begin{cases} 0, & V^v_{ij} \;\le\; \frac{\alpha}{\lambda} \\ 1, & V^v_{ij} \;>\; \frac{\alpha}{\lambda} \end{cases}$$

for WM, where $U_{ij} \stackrel{\text{def}}{=} f^2_{vv}(i,j) + 2f^2_{hv}(i,j) + f^2_{hh}(i,j)$ for WP, and $V^h_{ij} \stackrel{\text{def}}{=} f^2_h(i,j)$ and $V^v_{ij} \stackrel{\text{def}}{=} f^2_v(i,j)$ for WM. Using the above solution for the line process, we can eliminate the line processes in (5) and (6) and reduce the problem simply to a minimization over the source $\mathbf{f}$ alone. The new prior energies become

$$E^P_P(\mathbf{f}) = \sum_{ij} \phi^*\left(\sqrt{U_{ij}}\right), \tag{9}$$

where

$$\phi^*\left(\sqrt{U_{ij}}\right) = \begin{cases} \lambda U_{ij}, & \lambda U_{ij} \;\le\; \alpha \\ \alpha, & \lambda U_{ij} \;>\; \alpha. \end{cases} \tag{10}$$

for WP, and

$$E^M_P(\mathbf{f}) = \sum_{ij} \phi^*\left(\sqrt{V^v_{ij}}\right) + \sum_{ij} \phi^*\left(\sqrt{V^h_{ij}}\right) \tag{11}$$

for WM. The similarity of forms in (9) and (11) suggests a uniform notation $E_P(\mathbf{f}) = \sum_{ij} \phi^*\left(\sqrt{V_{ij}}\right)$ where V_{ij} is understood to equal U_{ij} for WP and either V^h_{ij} or V^v_{ij} for WM, with the summation of the two terms in (11) implicit for the WM case. This compact notation is quite convenient and illuminating in the subsequent development.

Figure 3(a) shows $\phi^*\left(\sqrt{V_{ij}}\right)$ as a "broken parabola". Though the global minima of (10) are the same as those of (5) and the use of binary line variables has been circumvented, there are problems associated with minimizing an energy function consisting of the likelihood and the broken parabola. The broken parabola makes the energy function non-convex and non-differentiable, thereby ruling out gradient-based descent methods.

3.2 A Continuation Method

In this subsection, we further refine the expression for the prior and present a deterministic annealing algorithm for the minimization of the energy function. The derivation here is skeletal as we make use of extensive reference to our previous work for the WM case [12, 13, 14]. We show that the broken parabola can be approached by a sequence of associated ϕ functions. The sequence approaches ϕ^* as a parameter β approaches infinity. The sequence can be obtained by first considering a transformation of the probability distributions to

$$\Pr_\beta(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{1}{Z(\mathbf{x}; \beta)} \exp(-\beta E(\mathbf{y})).$$

This new sequence of probability distributions, indexed by the real parameter ($\beta > 0$), is clearly related to the distributions in (3). As β approaches infinity, this new distribution becomes a highly peaked version of the distribution in (3). As far as MAP estimation is concerned, the location of the maximum is unchanged. In our deterministic annealing framework, β, identified as the inverse of a computational "temperature" ranges from low to high values according to an annealing schedule.

The broken parabola was obtained by eliminating the line process from the original objective function. The line process can also be "eliminated" from the posterior distribution $\Pr_\beta(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g})$. First, the marginal posterior distribution of the intensities is obtained by integrating out the line process from the Gibbs distribution. Since the prior does not involve the data $\mathbf{g}$, the partitioning of the posterior distribution can be written as

$$\Pr_\beta(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) = \Pr_\beta(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g})\Pr_\beta(\mathbf{L}=\mathbf{l}|\mathbf{F}=\mathbf{f}),$$

where $\Pr_\beta(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g})$ is the marginal posterior of $\mathbf{F}$. The energy function arising out of the marginal posterior of $\mathbf{F}$ is a function of $\mathbf{f}$ alone and not of $\mathbf{l}$. In [12, 14], we show that this function contains a smooth Gibbs prior that approaches the broken parabola as the parameter β is increased:

$$\begin{aligned}
\Pr_\beta(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g}) &= \sum_{\{l_{ij}\}} \Pr_\beta(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) \\
&= \frac{1}{Z_{MP}(\mathbf{g};\beta)} \exp\left[-\beta\left\{\sum_{t\theta}[(-g_{t\theta}\log(\bar{g}_{t\theta})+\bar{g}_{t\theta}] + \sum_{ij}\phi_\beta\left(\sqrt{V_{ij}}\right)\right\}\right],
\end{aligned}$$

where the notation $\sum_{\{l_{ij}\}}$ refers to summation over all possible realizations of the binary field $\mathbf{l}$, Z_{MP} is the appropriate normalization for the Marginal Posterior, and $\phi_\beta\left(\sqrt{V_{ij}}\right)$ is the smooth Gibbs prior defined by

$$\phi_\beta\left(\sqrt{V_{ij}}\right) \stackrel{\text{def}}{=} -\frac{1}{\beta}\log\left(\exp(-\beta\lambda V_{ij}) + \exp(-\beta\alpha)\right).$$

At large values of β, if $\alpha > \lambda V_{ij}$, the first term inside the $\log(\cdot)$ expression dominates and the entire expression reduces to λV_{ij}. If $\alpha < \lambda V_{ij}$, the second term inside the $\log(\cdot)$ expression dominates and the entire expression reduces to α. Inspection reveals that the ϕ_β function at large β is in fact the broken parabola of (10). At low values of β, the function is smooth, and, as can be seen from Fig. 3(b), approaches the broken parabola as $\beta \to \infty$. The overall energy function to be minimized at each β is thus

$$E(\mathbf{f};\beta) = E_D(\mathbf{f}) + E_{MP}(\mathbf{f};\beta), \tag{12}$$

where

$$E_{MP}(\mathbf{f};\beta) \stackrel{\text{def}}{=} \sum_{ij}\left[-\frac{1}{\beta}\log\left(\exp(-\beta\lambda V_{ij}) + \exp(-\beta\alpha)\right)\right].$$

Our original overall energy has thus been transformed in two stages to the overall energy defined above. Note that our notational conventions regarding V_{ij} apply directly to (12), so that both the WM and WP cases are represented by this equation.

Our task now reduces to finding the minimum at a given β of this energy function. The minimum is to be tracked as β is increased. Despite the absence of the line processes from our final energy function, they are implicitly contained in the broken parabola that is reached at limiting values of the control parameter β.

We summarize here our derivation [12, 13, 14] of a deterministic annealing algorithm to minimize the overall energy in (12). We utilize the familiar incomplete/complete data formulation of the Generalized Expectation Maximization (GEM) MAP approach [3]. The maximization step of the GEM-MAP algorithm in our case results in the minimization of

$$M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta) \stackrel{\text{def}}{=} -Q(\mathbf{f}|\hat{\mathbf{f}}^n) + E_{MP}(\mathbf{f};\beta), \tag{13}$$

where

$$-Q(\mathbf{f}|\hat{\mathbf{f}}^n) \stackrel{\text{def}}{=} \sum_{ij}\sum_{t\theta}\left[\mathcal{H}_{t\theta,ij}f_{ij} - g_{t\theta}\frac{\mathcal{H}_{t\theta,ij}\hat{f}^n_{i,j}}{\sum_{kl}\mathcal{H}_{t\theta,kl}\hat{f}^n_{kl}}\log(\mathcal{H}_{t\theta,ij}f_{ij})\right], \tag{14}$$

and $\hat{\mathbf{f}}^n$ is the estimate of $\mathbf{f}$ at the nth iteration. Note that the expectation step simply involves substituting $\hat{\mathbf{f}}^{n+1}$ for $\hat{\mathbf{f}}^{n}$ in $M(\mathbf{f}|\hat{\mathbf{f}}^{n})$. One could use an EM approach instead of GEM. However, unlike the EM approach, the GEM approach merely requires a monotonic increase in M-step without the requirement of maximization.

In this work, for the maximization step, we apply the method of Iterated Conditional Modes (ICM) [24], essentially a coordinate-wise descent method, to the above energy function $M(\mathbf{f}|\hat{\mathbf{f}}^{n};\beta)$. ICM is a logical choice for the optimization method since it can be used to obtain convenient closed form solutions for the update equation. Step sizes necessary for gradient descent methods are also avoided. ICM minimizes the energy function by a coordinate-wise descent procedure performed by minimizing the energy function with respect to f_{ij} while keeping all other source intensities fixed. Then a new location is chosen and the same method is repeated. After a full sweep of the lattice, the procedure is repeated until convergence criteria are met. Then β is increased and ICM is repeated.

Unfortunately, a closed-form solution for f_{ij} cannot be directly obtained by setting $\frac{\partial M(\mathbf{f}|\hat{\mathbf{f}}^{n};\beta)}{\partial f_{ij}} = 0$, as the differentiation results in an expression transcendental in f_{ij}. One possible solution to this problem is to consider the transcendental term to be a new, albeit dependent (on f_{ij}), variable and to descend on each such variable separately. In [12, 25, 26], we defined the new variables as

$$z_{ij} \stackrel{\text{def}}{=} \frac{1}{1+\exp(-\beta(\lambda V_{ij}-\alpha))} \tag{15}$$

and showed that z_{ij} can be identified as the mean value of the line process l_{ij} conditioned on $\mathbf{f}$ and $\mathbf{g}$ at each β. (Note that in equations that follow, the notation $z^{h}_{i,j}$ and $z^{v}_{i,j}$ for the case of WM corresponds to V_{ij} equal to V^{v}_{ij} or V^{h}_{ij}, respectively.)

With the new variables, $M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)$ becomes $M(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta)$. Now, we can perform ICM by keeping f_{ij} frozen while descending on z_{ij} and keeping z_{ij} frozen while descending on f_{ij}. As derived in [13, 14, 12], the update equations for f_{ij} are

$$f_{ij} = \frac{-(\sum_{t\theta}\mathcal{H}_{t\theta;ij} - 2\lambda X_3) + \sqrt{(\sum_{t\theta}\mathcal{H}_{t\theta;ij} - 2\lambda X_3)^2 + 8\lambda X_2 X_1}}{4\lambda X_2}, \tag{16}$$

where

$$X_1 \stackrel{\text{def}}{=} \sum_{t\theta} g_{t\theta}\frac{\mathcal{H}_{t\theta,ij}\hat{f}^n_{i,j}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}},$$

for WP,

$$X_2 \stackrel{\text{def}}{=} 10(1-z_{i,j}) + (1-z_{i,j+1}) + 3(1-z_{i,j-1}) + (1-z_{i+1,j}) + 3(1-z_{i-1,j}) + 2(1-z_{i-1,j-1})$$

$$\begin{aligned} X_3 \stackrel{\text{def}}{=} & \left\{2(f_{i,j+1}+f_{i,j-1}) + 2(f_{i+1,j}+f_{i-1,j}) + 2(f_{i,j+1}+f_{i+1,j}-f_{i+1,j+1})\right\}(1-z_{i,j}) \\ & +(2f_{i,j+1}-f_{i,j+2})(1-z_{i,j+1}) \\ & +\left\{(2f_{i,j-1}-f_{i,j-2}) + 2(f_{i,j-1}-f_{i+1,j-1}+f_{i+1,j})\right\}(1-z_{i,j-1}) \\ & +(2f_{i+1,j}-f_{i+2,j})(1-z_{i+1,j}) \\ & +\left\{(2f_{i-1,j}-f_{i-2,j}) + 2(f_{i-1,j}-f_{i-1,j+1}+f_{i,j+1})\right\}(1-z_{i-1,j}) \\ & +(2f_{i-1,j}+f_{i,j-1}-f_{i-1,j-1})(1-z_{i-1,j-1}), \end{aligned}$$

and for WM,

$$X_2 \stackrel{\text{def}}{=} (1-z^h_{i,j}) + (1-z^h_{i-1,j}) + (1-z^v_{i,j}) + (1-z^v_{i,j-1})$$

$$X_3 \stackrel{\text{def}}{=} f_{i+1,j}(1-z^h_{i,j}) + f_{i-1,j}(1-z^h_{i-1,j}) + f_{i,j+1}(1-z^v_{i,j}) + f_{i,j-1}(1-z^v_{i,j-1}).$$

The factors, $X2$ and $X3$, for the WP were not derived in [13, 14], but their derivation, although messy, follows in a manner similar to that of the WM case in [13, 14]. The interested reader may find the derivation in [25], available on request.

The M-step thus consists of two kinds of ICM updates. We first update $\mathbf{f}$ according to update equation (16) over the entire image and then update $\mathbf{z}$ according to (15) over the entire image as well. Once the new estimate $\hat{\mathbf{f}}^{n+1}$ has been obtained, a new M-step objective function is generated by substituting $\hat{\mathbf{f}}^{n+1}$ for $\hat{\mathbf{f}}^{n}$ to get $M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^{n+1}; \beta)$. Then the two-step ICM procedure is repeated. In this manner, the entire GEM procedure is carried out until convergence. Qualitatively, the update alternates between update of the (analog) line processes and the intensities.

To complete the definition of the algorithm, we specify two stopping criteria and an annealing schedule that specifies the rule for changing β at each epoch. For our simulations, we used doubling schedule for β. Iterations at a given β are performed until the relative energy change $\frac{E_n - E_{n-1}}{E_n - E_0}$, where n indexes iteration number within an epoch, is less than a threshold ϵ. The entire simulation is terminated when $z_{ij} \leq \tau$ or $z_{ij} \geq 1 - \tau$, where τ is chosen as a threshold. Since z corresponds [14] to the conditional expected value of the line process l, the approach of z to 0 or 1 corresponds to the approach of the ϕ function to sufficiently close distance from the broken parabola.

A special case occurs in the update of f_{ij} in (16): when all neighboring z values are sufficiently close to unity, X_2 approaches zero and the expression in (16) appears to diverge. However, in the limit as $z \rightarrow 1$ the expression, in fact, reduces to the standard EM update equation since the smoothing action of the prior is suspended at this location.

The strategy above was predicated on setting the partial derivative of $M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^{n}; \beta)$ with respect to f_{ij} to zero and solving for f_{ij}. This does not guarantee that the obtained solution is a minimum of $M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)$. In [14], a proof is given showing that the solution obtained is indeed a minimum.

In summary, energy minimization consists of minimizing, at a given β, the expression $M(\mathbf{f}|\hat{\mathbf{f}}^n; \beta)$ in (13) until the stopping criterion governed by threshold ϵ is reached. Then β is increased according to the annealing schedule. Iterations at different β proceed until the final stopping criterion specified by τ is attained.

4 Experiments and Results

We performed 2-D simulation studies with projection data from 128×128 phantoms, with 128 projection angles over 360° and 192 detector bins at each projection. Both detector bin and image pixel linear dimensions were 2 mm.

In this work we considered attenuation effects only. For attenuation correction, we used a method proposed by Liang *et al.* [27] that utilizes the Siddon's method [28] to recursively compute attenuation factors along each projection ray starting at the pixel closest to the detector. Though our projection scheme works even in the case of nonuniform attenuation, we considered only uniform attenuation (a good approximation for brain imaging) with an effective attenuation coefficient equal to 0.12 cm^{-1}. At each stage of our iterative reconstructions, then, compensation consisted of simply an attenuated forward projection. To generate projection data, we simply add Poisson noise to each of the attenuated projections of the phantom, using the noiseless projection value as the mean in a Poisson random number generator. The implementational details can be found in [29].

We tested the four reconstruction algorithms – GEM-MAP with the WM prior, GEM-MAP with the WP prior, and ML-EM with two different stopping rules. For convenience, we will refer to these simply as the WM, WP, EM-1, and EM-2 algorithms. The algorithms were tested on two digital phantoms, A and B. Phantom A (Fig. 4(a)) was derived from a digitized (256×256) rhesus monkey autoradiograph [15] of the rCBF SPECT agent ^{99m}Tc-ECD. We simply cropped the 256×256 original to a circular region and resampled into a 128×128 matrix. With this geometrical scaling and 2mm pixel size, the equivalent physical extent of each rCBF pattern was somewhat larger than that of a human brain. Effects of attenuation were thus sizable. The grey values of the autoradiograph were not gamma corrected for film response, but we scaled the grey level distribution of the phantom A to have an average contrast of 4:1 between grey matter and white matter. This was accomplished using average values in regions clearly belonging to grey or white matter. Phantom B (Fig. 4(b)) is a 2-D Hoffman brain phantom with activities of 4:1:0 in grey matter, white matter and CSF, respectively. (Note that the Hoffman phantom is piecewise constant.) The total number of detector counts for phantoms A and B were approximately 732K and 300K, respectively, where the number of counts for phantom A was determined by scaling the 300K Hoffman counts by the ratio (area phantom A / area phantom B). To evaluate and compare, in an ensemble sense, quantitative performance of the reconstruction algorithms, we generated two sets of 40 independent noise realizations of attenuated projection data for each phantom.

For the ML-EM reconstructions, we used two different stopping rules. For our EM-2 algorithm, we intended to utilize the χ^2 stopping rule proposed by Llacer and Veklerov

[30]. The χ^2 stopping rule terminates iterations when the following condition holds:

$$\chi^2 \stackrel{\text{def}}{=} \sum_{t\theta} \frac{\left(g_{t\theta} - \bar{g}_{t\theta}\right)^2}{\bar{g}_{t\theta}} \simeq M,$$

where M is the number of detector bins. Since our imaging system has 192×128 bins, the rule dictates that the ML-EM algorithm should be stopped when $\chi^2 = 24,576$, which for our simulations resulted in a typical number of iteration of around 5. At this point, the reconstructed images seemed to be far oversmoothed. We instead used $\chi^2 = 11,990$ for phantom A and $\chi^2 = 8,170$ for phantom B. These choices were made qualitatively based on the starting point for deterioration of the smoothness. The corresponding average number of iterations was 25 for both phantoms A and B. Our EM-1 algorithm adhered to a second stopping rule based on minimum root mean squared (RMS) error in the reconstruction. (We were interested in seeing the reconstruction stopped at the minimum RMS error though, of course, this rule is not applicable in the clinic.) With our simulations, the average number of iterations for the minimum RMS error averaged 16 for phantom A and 15 for phantom B. For our noise trials, we fixed the number of iterations at these average values.

A problem with Bayesian methods, including our WP and WM models, is that it is difficult to set hyperparameters (λ and α in our case) automatically. While this is an area of active research by us and others [31, 32], for the present work we set these values empirically. For the WM reconstructions, we first chose the parameter value of λ by considering the degree of smoothness in the reconstructed images; the λ was adjusted to smooth out the noise while keeping the edge structure. The gradient limit effects were also considered in finding the value of λ. Having set the λ, we adjusted the parameter α empirically to preserve grey/white matter boundaries. With simulations,

λ was set to 5.0 for both phantoms, and α was set to 0.18 for phantom A and 0.22 for phantom B. The annealing schedule ran through 12 to 16 values with a doubling at each new value of β starting from $\beta = 0.03125$. The initial value for the parameter β was chosen to be small enough to yield a very smooth initial reconstruction [14] but to approach the broken parabola in a reasonable amount of computation time. We terminated iterations at a given β when the relative energy change $\frac{E_n - E_{n-1}}{E_n - E_0}$ was less than the threshold $\tau = 0.05$ as described in Sec. 3. The entire simulation was terminated using a threshold value of $\epsilon = 0.1$ as described in Sec. 3. With these criteria for WM, there were approximately 5 to 20 iterations per each β for a total of around 200 iterations. For WP reconstructions λ was set to 5.0 for both phantoms, and α to 0.1 for phantom A and 0.2 for phantom B. Stopping criteria were identical to the WM case. The initial condition for $\mathbf{f}$ was a constant intensity of 1.0 for all four algorithms and $\mathbf{z}$ was initialized to 0.5 for both WM and WP algorithms.

Figures 5 and 6 show anecdotal reconstructed images from the four reconstruction algorithms for both phantoms. Several qualitative observations may be noted in Fig. 5. The WM reconstruction, Fig. 5(c), looks artificially patchy, a result not unexpected since WM tends to favor piecewise constant reconstructions. The gradient limit effects ("patchiness") may be reduced by increasing λ, but then a tradeoff results and the reconstructed images become excessively smooth. The WM attempts to create step edges even in the ramp regions, resulting in spurious discontinuities. By extension to a higher order model, the WP reconstructions avoid the artifacts of WM reconstructions as seen in Fig. 5(d). Comparison of WP to the EM reconstructions in Fig. 5 shows that WP reduces noise without introducing artifacts. While the visual improvements from EM-1 and EM-2 to WP are not stunning, close inspection reveals that the WP reconstruction, in fact, captures subtle aspects of the phantom in Fig. 5 that are missed by the EM-algorithms. These qualitative observations are supported by quantitative analyses below. Further qualitative observations are shown in Fig. 6, using the piecewise constant Hoffman phantom of Fig. 4(b). The WM reconstruction in Fig. 6(c) does not appear to be qualitatively stunning, but performs relatively well in a quantitative sense as described below. Again, the WP reconstruction of Fig. 6(d) suppress noise with no ill effects when compared to the EM reconstructions of Figs. 6(a) and (b).

Figure 7(a) shows two lines on phantom A along which profiles are displayed for the four algorithms. The profiles for line 2 in Figs. 7(e)-(h) are revealing. Both EM plots are noisy, the result for EM-1 (fewer iterations) is smoother than that for EM-2 but incurs a bias error. The WM reconstruction illustrates the tendency towards piecewise constant regions. Note especially in Fig. 7(g), the profile in the edge region around pixel 37. The WM has reconstructed a ramp as a step edge typifying the kinds of ensemble accuracy and precision errors reported below. On the other hand, the flat regions are, as expected, better approximated by WM than the EM algorithms. Comparison of the WP result in Fig. 7(h) shows that the noise errors typical of EM (Figs. 7(e) and (f)) and edge approximation and hot spot errors typical of WM Fig. 7(g) have both been corrected. The two ramp regions around pixel 40 are well reconstructed as is the smooth region around pixels 50-70. Figures. 7(b), (c) and (d) for line 1 show qualitatively similar results.

To evaluate the reconstructions quantitatively, we computed bias and standard deviation (STD) images as done in [31]. A bias image, b_{ij}, is defined as

$$b_{ij} \stackrel{\text{def}}{=} \frac{1}{K}\sum_{k=1}^{K}\left(\hat{f}_{ij}^{k} - f_{ij}\right),$$

where $\hat{f}_{ij}^{k}$ is the k^{th} reconstruction of phantom f at location (i,j) and the summation is over $K = 40$ independent noise trials. To display the bipolar bias image, an intermediate grey scale value of 128 out of 256 levels was used as zero bias. A standard deviation image, σ_{ij}, is defined as

$$\sigma_{ij} \stackrel{\text{def}}{=} \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(\hat{f}_{ij}^{k} - \bar{f}_{ij}\right)^2},$$

where $\bar{f}_{ij}$ is the mean of $\hat{f}_{ij}$ over the noise trials defined as $\bar{f}_{ij} \stackrel{\text{def}}{=} \frac{1}{K}\sum_{k=1}^{K}\hat{f}_{ij}^{k}$. Note that the same grey scale normalization is used for all images in a given figure to allow fair comparisons.

Figures 8 and 9 show our bias/STD images for phantoms A and B, respectively. In general, the bias images in Figs. 8 and 9 show a negative bias for high-signal (grey matter) regions and a positive bias for low signal (white matter) regions. Despite the unsatisfactory WM anecdotal images, the bias for WM (Figs. 8(b) and 9(b)) compares favorably with that of EM-1 and WP. The bias for WM appears less correlated with image intensity than for the case of WP and EM-1. More interesting results are seen in the case of STD images. The STD images for EM, Figs. 8(d) and 9(d) show noise growing with signal intensity as expected. The deficiency of the WM prior is illustrated in the STD images, Figs. 8(e) and 9(e). In both ramplike (Fig. 8) and steplike (Fig. 9) edge regions, the variance is high. This effect appears to be due to the fact that the

WM prior results in unstable estimates of edge location; in different noise realizations, the locations of continuity breaks can shift. However, in regions far from edges, the WM performs well in the sense of low variance. The WP prior produces excellent variance results as seen in Figs. 8(f) and 9(f). The relatively uniform STD images there indicate low variance both in ramp and step regions, as well as in constant regions. By themselves, the excellent variance results for WP are not significant, since low variance can be obtained simply by oversmoothing. When combined with good bias performance, however, the WP higher order prior indeed shows an advantage relative to EM and WM. Note also that the good performance of WP even for the Hoffman phantom illustrates that the weak plate prior adequately subsumes the case of piecewise flat image structure.

We also compared regional bias and variance for each reconstruction algorithm using two different measures. The first measure, designed to assess regional quantitation performance, is simply the bias and variance of a random variable $Z^{k,R} = \frac{1}{m}\sum_{ij\in R}\hat{f}_{ij}^k$ equal to the average count level in a region R comprised of m pixels. The appropriate regional definitions are then

$$\mathrm{B}_Z^R = \frac{1}{K}\sum_{k=1}^{K}(Z^{k,R} - Z^R)$$

for regional bias, and

$$\mathrm{STD}_Z^R = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(Z^{k,R} - \bar{Z}^R)^2}$$

for regional standard deviation, where $Z^R \stackrel{\text{def}}{=} \frac{1}{m}\sum_{ij\in R} f_{ij}$ and $\bar{Z}^R \stackrel{\text{def}}{=} \frac{1}{K}\sum_{k=1}^{K} Z^{k,R}$. Our second measure is intended to summarize the fidelity of a reconstruction estimate within a region. This measure is not a statistical bias and variance, but is a simple decomposition of RMS regional error into a sum of two terms obtained by suitable integration (over R) of b_{ij}^2 and σ_{ij}^2. The definitions are

$$\mathrm{B}_G^R = \sqrt{\frac{1}{m} \sum_{ij \in R} b_{ij}^2}$$

and

$$\mathrm{STD}_G^R = \sqrt{\frac{1}{m} \sum_{ij \in R} \sigma_{ij}^2},$$

where, as it turns out,

$$\left(\mathrm{B}_G^R\right)^2 + \left(\mathrm{STD}_G^R\right)^2 = \left(\mathrm{RMS}_G^R\right)^2,$$

and $\mathrm{RMS}_G^R = \sqrt{\frac{1}{m} \sum_{ij} r_{ij}^2}$ is the usual expression for regional RMS error with r_{ij} the pointwise RMS error. (This definition has a convenient interpretation: If $\mathbf{f}$, the true object within R, and $\hat{\mathbf{f}}^k$, the K noise realizations are m-dimensional vectors, then B_G^R, a measure of bias, is the length of the vector from $\bar{\mathbf{f}}$ to $\mathbf{f}$, and STD_G^R is a scalar measure of the spread of the $\hat{\mathbf{f}}^k$ about its mean $\bar{\mathbf{f}}$.)

Figure 10(a) shows 17 circular regions, each of 21 pixels, superposed on the phantom, and Figs. 10(b)-(e) show bias B_Z^R - variance STD_Z^R plots for each algorithm. Inspection of these shows that as we increase the number of iterations from EM-1 to EM-2, bias improves at the expense of variance, as expected. Comparison of the results for WP and WM shows WP typically does much better in terms of STD than WP in regions 1 - 7 containing edge structure, and does slightly better in bias in these regions. On the other hand, in constant regions (8,10,11,16,17), WM has improved STD, but not bias. Overall, the WP prior leads to excellent STD behavior. At our choice of iteration number, the WP also outperforms EM in bias, though at high iteration number (~ 200), EM bias is comparable. At these high iterations, EM STD becomes exceedingly high, however.

Figure 11(a) shows a second set of regions chosen to surround interesting region structure: regions 9, 10, 11, 12, and 15 enclose interesting edge and ramplike structures, while other regions enclose relatively flat regions of high and low signal intensity. In Figs. 11(b)-(e), we display B_G^R and STD_G^R. The plots again show an observed bias-variance tradeoff with increasing iteration number for EM. The WP prior has nearly identical bias as WM, but lower STD, especially in edge regions. Comparison to the pointwise bias/STD images suggests that the higher WM STD is due to a mislocalization error of an edge by a WM prior since the WM pointwise STD is especially high in the vicinities of high gradient. Comparison of Figs. 11(d) and (e) shows that for the (boxed) regions enclosing significant edge structure, the WM STD is much higher than that for WP, but for (circled) flat regions, WM compares favorably. Comparison of Figs. 11(b), (c), (d), and (e) shows that for comparable performance in bias, the WP exhibits superior performance in variance in regions including edge structure.

5 Discussion and Conclusions

We have considered a new, higher order mechanical model as a prior. Regional bias and variance analyses show that the WP exhibits both small bias and variance in most of the regions that incorporate edge structure, as well as in smooth regions. A reasonable conclusion is that for comparable performance to EM and WM in bias, WP yields reduced variance in all types of regions. Although the WP prior favors piecewise linear ramplike regions, the results from the Hoffman phantom show that the WP works even on the piecewise constant regions by stabilizing edge locations. In contrast, the WM prior produces inaccurate localization of estimated edges yielding the large variance in the edge regions, though it accurately reconstructs the piecewise constant regions.

Despite the encouraging results, any enthusiasm for the use for WP models should be tempered by the fact that the behavior of these priors with respect to hyperparameters (λ and α) and annealing schedule is not yet completely quantified. In particular, the sensitivity of our bias/variance metrics to these parameters needs to be better understood. Our results are predicated on our choices for EM stopping rules, and on the asymptotic behavior of our WM and WP algorithms. The WP and WM reconstruction methods could be made more practical by a systematic way of determining parameters λ, α, and the annealing schedule for β. Our own initial efforts for [32, 33] parameter determination utilize maximum likelihood methods with training data, and other efforts [34] have been reported for determining λ.

We note that, for our simulations, the total computing time for the WP and WM reconstruction algorithms is about 8 times as long as the EM reconstructions. In fact, though the WP and WM algorithms appear complex, they incur only a slightly greater computational cost per iteration than EM. As EM inevitably becomes practical, so will this form of GEM algorithm. For example, EM reconstruction times of 30 sec/plane for unoptimized code on a conventional workstation are reported in [35], and these times can be dramatically reduced with parallel processing.

References

[1] S. Geman and D. E. McClure, "Statistical Methods for Tomographic Image Reconstruction", *Bulletin of the International Statistical Institute*, LII-4, pp. 5–21, 1987.

[2] E. Levitan and G. T. Herman, "A Maximum *A Posteriori* Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography", *IEEE Trans. on Medical Imaging*, MI-6, pp. 185–192, 1987.

[3] T. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Bayesian Reconstruction for Poisson Data Using Gibbs Priors", *IEEE Trans. on Medical Imaging*, MI-8(2), pp. 194–202, June 1989.

[4] V. E. Johnson, W. H. Wong, X. Hu, and C. Chen, "Bayesian Reconstruction of PET Images Using Gibbs priors", In D. A. Ortendahl and J. Llacer, editors, *Information Processing in Medical Imaging*, pp. 15–28, Wiley-Liss, 1989.

[5] P. J. Green, "Bayesian Reconstructions from Emission Tomography Data Using a Modified EM Algorithm", *IEEE Trans. on Medical Imaging*, MI-9(1), pp. 84–93, March 1990.

[6] K. Lange, "Convergence of EM Image Reconstruction Algorithms with Gibbs Smoothing", *IEEE Trans. on Medical Imaging*, MI-9(4), pp. 439–446, December 1990.

[7] D. S. Lalush and B. M. W. Tsui, "Simulation Evaluation of Gibbs Prior Distributions for Use in Maximum *a posteriori* SPECT Reconstructions", *IEEE Trans. on Medical Imaging*, MI-11, pp. 267–275, 1992.

[8] G. Gindi, M. Lee, A. Rangarajan, and G. Zubal, "Bayesian Reconstruction of Functional Images Using Anatomical Information as Priors", *IEEE Trans. on Medical Imaging*, 12, pp. 670–680, Dec. 1993.

[9] S. C. Huang, M. E. Phelps, E. F. Hoffman, K. Sideris, C. J. Selin, and D. E. Kuhl, "Noninvasive Determination of Local Cerebral Metabolic Rate of Glucose in Man", *Am. J. Physiol.*, 238, pp. E69–E82, 1980.

[10] P. Davies and A. Verth, "Regional Distribution of Muscarinic Acetylcholine Receptor in Normal and Alzheimer's - Type Dementia Brains", *Brain Res.*, 138, pp. 385–392, 1978.

[11] E. J. Hoffman, P. D. Cutler, W. M. Digby, and J. C. Mazziotta, "3-D Phantom To Simulate Cerebral Blood Flow and Metabolic Images for PET", *IEEE Trans. Nuclear Sci*, 37, pp. 616–620, 1990.

[12] A. Rangarajan, M. Lee, G. Zubal, and G. Gindi, "A Continuation Method for Emission Tomography", In *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume II, pp. 1204–1207, October 1992.

[13] M. Lee, A. Rangarajan, G. Zubal, and G. Gindi, "A Continuation Method for Emission Tomography", *IEEE Trans. on Nuclear Science*, 40, pp. 2049–2058, Dec. 1993.

[14] G. Gindi, A. Rangarajan, M. Lee, P. J. Hong, and G. Zubal, "Bayesian Reconstruction for Emission Tomography via Deterministic Annealing", In H. Barrett and A. Gmitro, editors, *Information Processing in Medical Imaging*, pp. 322–338, Springer–Verlag, 1993.

[15] R. C. Walovitch, T. C. Hill, S. T. Garrity, E. H. Cheesman, B. A. Burgess, D. H. O'Leary, A. D. Watson, M. V. Ganey, R. A. Morgan, and S. J. Williams, "Characterization of Technetium-99m-L,L-ECD for Brain Perfusion Imaging, Part 1: Pharmacology of Technetium-99m ECD in Nonhuman Primates", *Journal of Nuclear Medicine*, 30, pp. 1892–1901, 1989.

[16] D. Geman and G. Reynolds, "Constrained Restoration and the Recovery of Discontinuities", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-14(3), pp. 367–383, March 1992.

[17] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6(6), pp. 721–741, November 1984.

[18] V. E. Johnson, W. H. Wong, X. Hu, and C.-T. Chen, "Image Restoration Using Gibbs Priors: Boundary Modeling, Treatment of Blurring, and Selection of Hyperparameter", *IEEE Trans. Patt. Anal. Mach. Intell.*, 13(5), pp. 413–425, May 1991.

[19] R. Leahy and X. Yan, "Incorporation of Anatomical MR Data for Improved Functional Imaging with PET", In A. C. F. Colchester and D. J. Hawkes, editors, *Information Processing in Medical Imaging*, pp. 105–120, Springer–Verlag, 1991.

[20] G. E. Hinton, *Relaxation and Its Role In Vision*, PhD thesis, University of Edinburgh, 1977.

[21] A. Blake and A. Zisserman, *Visual Reconstruction*, Artificial Intelligence, MIT Press, Cambridge, MA, 1987.

[22] S. J. Lee, A. Rangarajan, and G. Gindi, "Weak Plate Mechanical Models in Bayesian Reconstruction for Emission Tomography", In *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume 3, pp. 1533–1537, November 1993.

[23] W. E. L. Grimson, *From Surfaces to Images*, The MIT press, Cambridge, MA, 1981.

[24] J. Besag, "On the Statistical Analysis of Dirty Pictures", *Journal of the Royal Statistical Society,* Series B, 48(3), pp. 259–302, 1986.

[25] S. J. Lee, A. Rangarajan, and G. Gindi, "Bayesian Image Reconstruction in SPECT Using Mechanical Models as Priors", Technical Report MIPL-TR-9402, Medical Image Processing Laboratory, Department of Radiology, SUNY at Stony Brook, July 1994.

[26] M. Lee, *Bayesian Reconstruction In Emission Tomography Using Gibbs Priors*, PhD thesis, Yale University, New Haven, CT, 1994.

[27] Z. Liang, T. G. Turkington, D. R. Gilland, R. J. Jaszczak, and R. E. Coleman, "Simultaneous Compensation for Attenuation, Scatter and Detector Response for SPECT Reconstruction in Three Dimensions", *Phys. Med. Biol.*, 37, pp. 587–603, 1991.

[28] R. Siddon, "Fast Calculation of the Exact Radiological Path for a 3D CT Array", *Med. Phys.*, 12, pp. 252–255, 1985.

[29] S. J. Lee, *Bayesian Image Reconstruction in Emission Computed Tomography Using Mechanical Models as Priors*, PhD thesis, State University of New York at Stony Brook, Stony Brook, NY, 1995.

[30] J. Llacer and E. Veklerov, "Feasible Images and Practical Stopping Rules for Iterative Algorithms in Emission Tomography", *IEEE Trans. on Medical Imaging*, MI-8, pp. 186–193, 1989.

[31] Z. Zhou, R. M. Leahy, and E. U. Mumcuoglu, "A Comparative Study of the Effects of Using Anatomical Priors in PET Reconstruction", In *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume 3, pp. 1749–1753, November 1993.

[32] S. J. Lee, G. R. Gindi, I. G. Zubal, and A. Rangarajan, "Using Ground-Truth Data to Design Priors in Bayesian SPECT Reconstruction", In Y. Bizais, C. Barillot, and R. D. Paola, editors, *Information Processing in Medical Imaging*, pp. 27–38, Kluwer Academic Publishers, 1995.

[33] G. Gindi and A. Rangarajan, "What Can SPECT Reconstruction Learn from Autoradiography", In *IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume 4, pp. 1715–1719, November 1994.

[34] Z. Zhou, R. M. Leahy, and E. U. Mumcuoglu, "Maximum Likelihood Hyperparameter Estimation for Gibbs Priors with Applications to PET", In Y. Bizais, C. Barillot, and R. D. Paola, editors, *Information Processing in Medical Imaging*, pp. 39–51, Kluwer Academic Publishers, 1995.

[35] J. Llacer, E. Veklerov, L. R. Baxter, S. T. Grafton, L. K. Griffeth, R. A. Hawkins, C. K. Hoh, J. C. Mazziotta, E. J. Hoffman, and C. E. Metz, "Results of a Clinical Receiver Operating Characteristic Study Comparing Filtered Backprojection and Maximum Likelihood Estimator Images in FDG PET Studies", *J. Nucl. Med.*, 34, pp. 1198–1203, 1993.

Figure Captions

Figure 1: (a) Autoradiograph from rhesus monkey shows distribution of radionuclide from an rCBF study using a ^{99m}Tc-ECD radiopharmaceutical. Areas corresponding to white matter, cortical, and grey-matter structures are evident. (b) A surface plot of a section of the autoradiograph.

Figure 2: Reconstruction of a ramp phantom illustrates the behavior of the mechanical priors. (a) ML-EM reconstruction. Effects of noise are evident. (b) GEM-MAP reconstruction using weak membrane prior. Flat regions are well recovered, but linear regions become terraced. (c) Same as previous case but with a different value of the parameter λ. (d) GEM-MAP reconstruction using weak plate prior. Both flat and linear regions are well recovered.

Figure 3: (a) The "broken parabola" $\phi^*(x)$. (b) The plots show the approximating function $\phi_\beta(x)$ for several values of β, illustrating that $\lim_{\beta\to\infty}\phi_\beta(x) = \phi^*(x)$.

Figure 4: Phantoms used in the experiments. (a) Phantom A. (b) Phantom B.

Figure 5: Anecdotal reconstructions for Phantom A. (a),(b),(c),(d): EM-1, EM-2, WM and WP reconstructions, respectively.

Figure 6: Anecdotal reconstructions for Phantom B. (a),(b),(c),(d): EM-1, EM-2, WM and WP reconstructions, respectively.

Figure 7: Profile plots for an anecdotal reconstruction of phantom A. The dotted line indicates the phantom, and the solid line the reconstruction. (a) Phantom, with profile lines shown. (b) Profile for line 1, EM-1. (c) Profile for line 1, WM. (d) Profile for

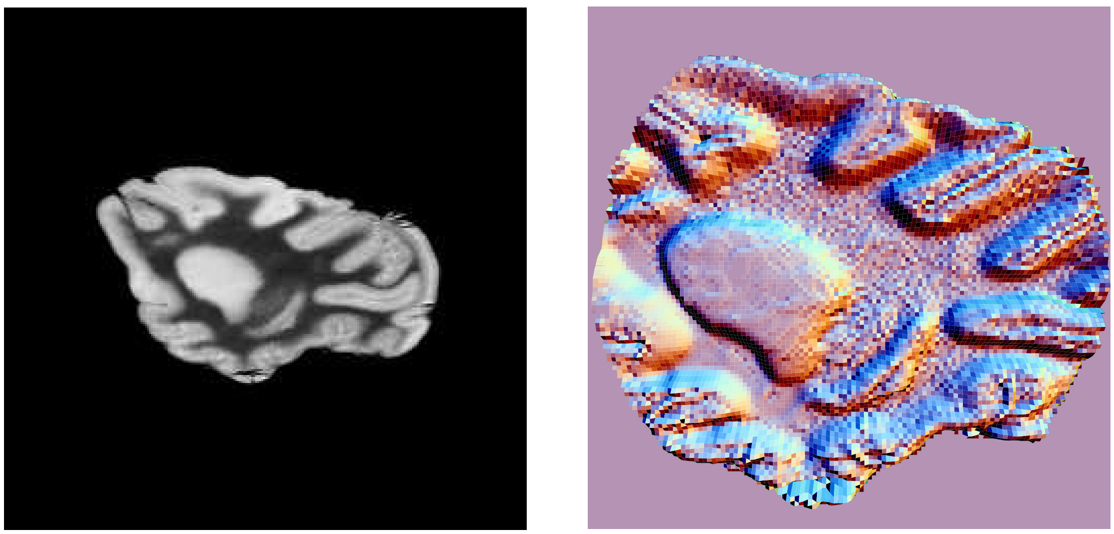
line 1, WP. (e) Profile for line 2, EM-1. (f) Profile for line 2, EM-2. (g) Profile for line 2, WM. (h) Profile for line 2, WP.

Figure 8: Pointwise bias-standard deviation (STD) images for phantom A obtained from 40 independent noise realizations. The bias images are bipolar, with a value of zero displayed as an intermediate grey, with darker/lighter regions corresponding to negative/positive bias. Grey scale is same for all figures. (a) Bias EM-2. (b) Bias WM. (c) Bias WP. (d) STD EM-2. (e) STD WM. (f) STD WP.

Figure 9: Pointwise bias-standard deviation images for Phantom B obtained for 40 independent noise realizations. Grey scale is same for all figures. Descriptions for images (a)-(f) follow those in Fig. 8.

Figure 10: Regional bias (B_Z^R) - standard deviation (STD_Z^R) plots for Phantom A. (a) The phantom with region borders superposed. In (b)-(e), the abscissa indexes the region number, and plotted points with error bars indicate bias $\pm$ one standard deviation. (b) Plots for EM-1 reconstructions. (c) Plots for EM-2. (d) Plots for WM. (e) Plots for WP.

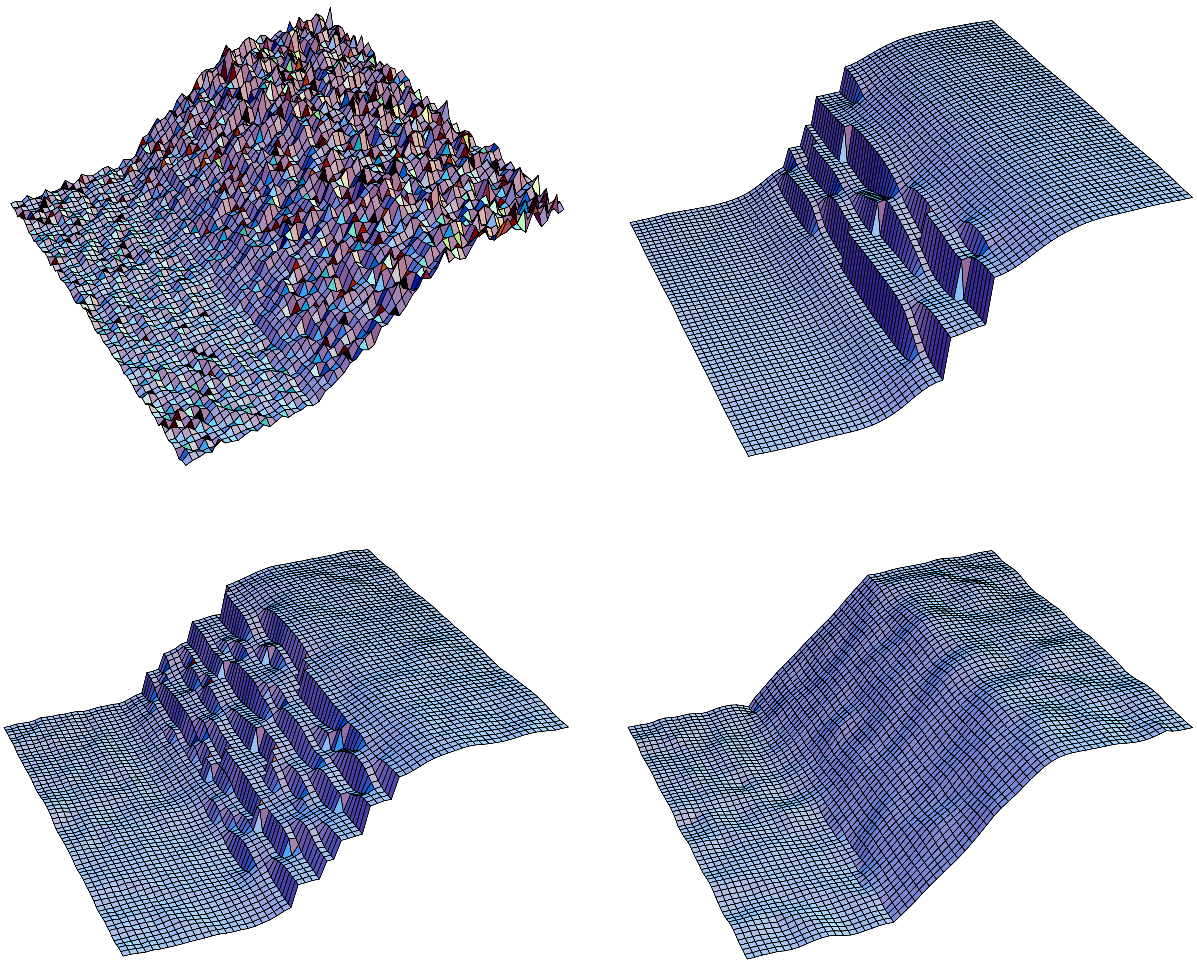
Figure 11: Regional measures B_G^R and STD_G^R for Phantom A. (a) The phantom with region borders superposed. In (b)-(e), the abscissa indexes the region number, and plotted points with error bars indicate $\mathrm{B}_G^R \pm \mathrm{STD}_G^R$. (b) Plots for EM-1 reconstructions. (c) Plots for EM-2. (d) Plots for WM. (e) Plots for WP.



(a) **(b)**

Figure 1:

(a) (b)



(c) (d)

Figure 2:

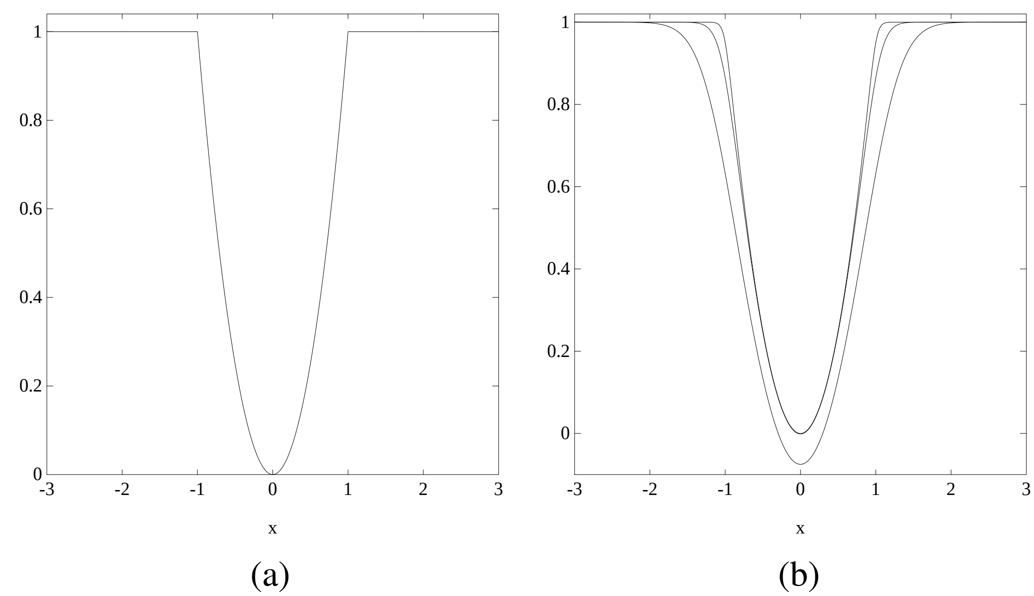
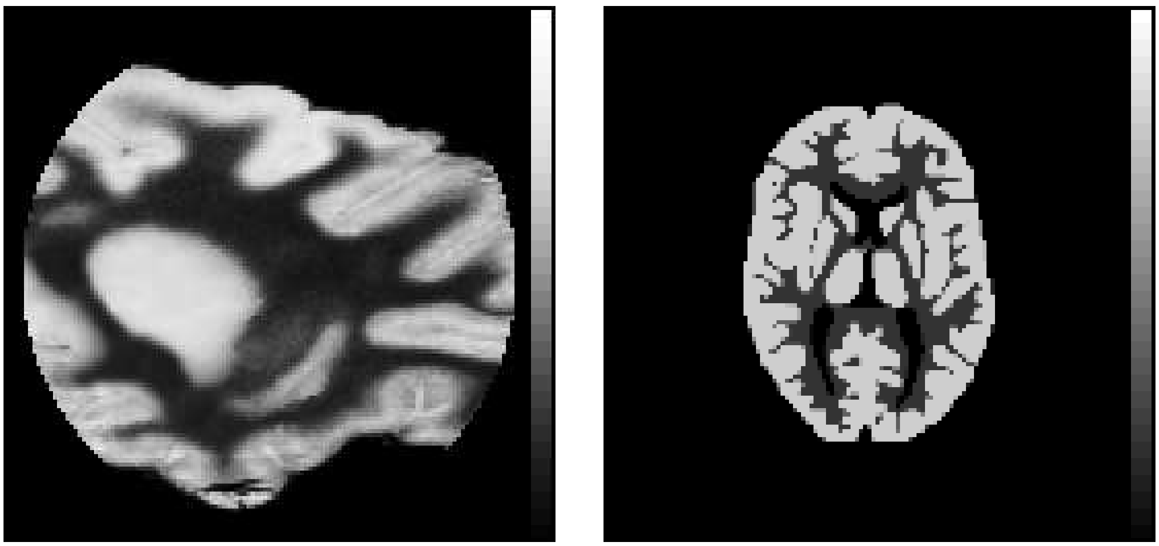


Figure 3:



(a) (b)

Figure 4:

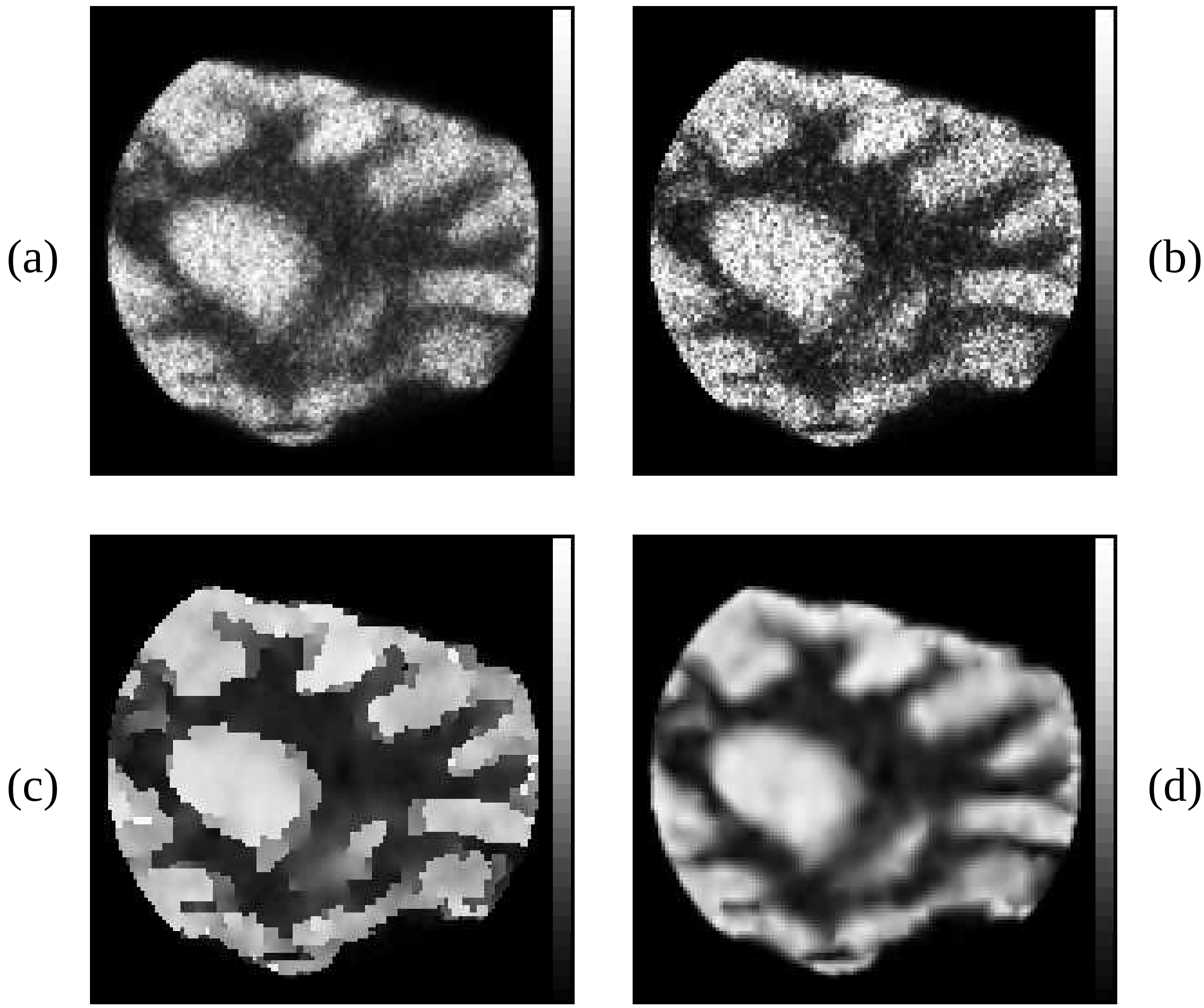


Figure 5:

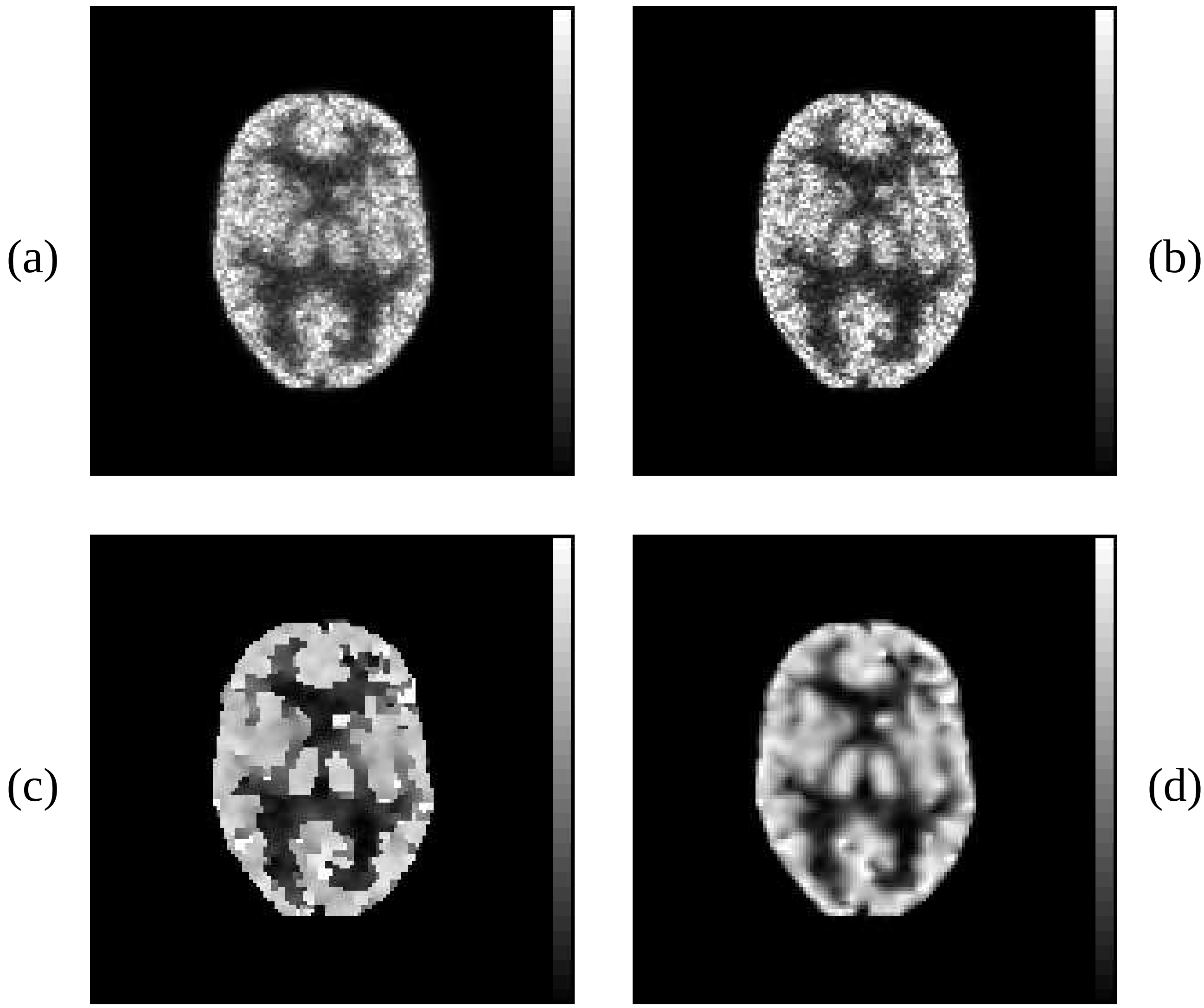


Figure 6:

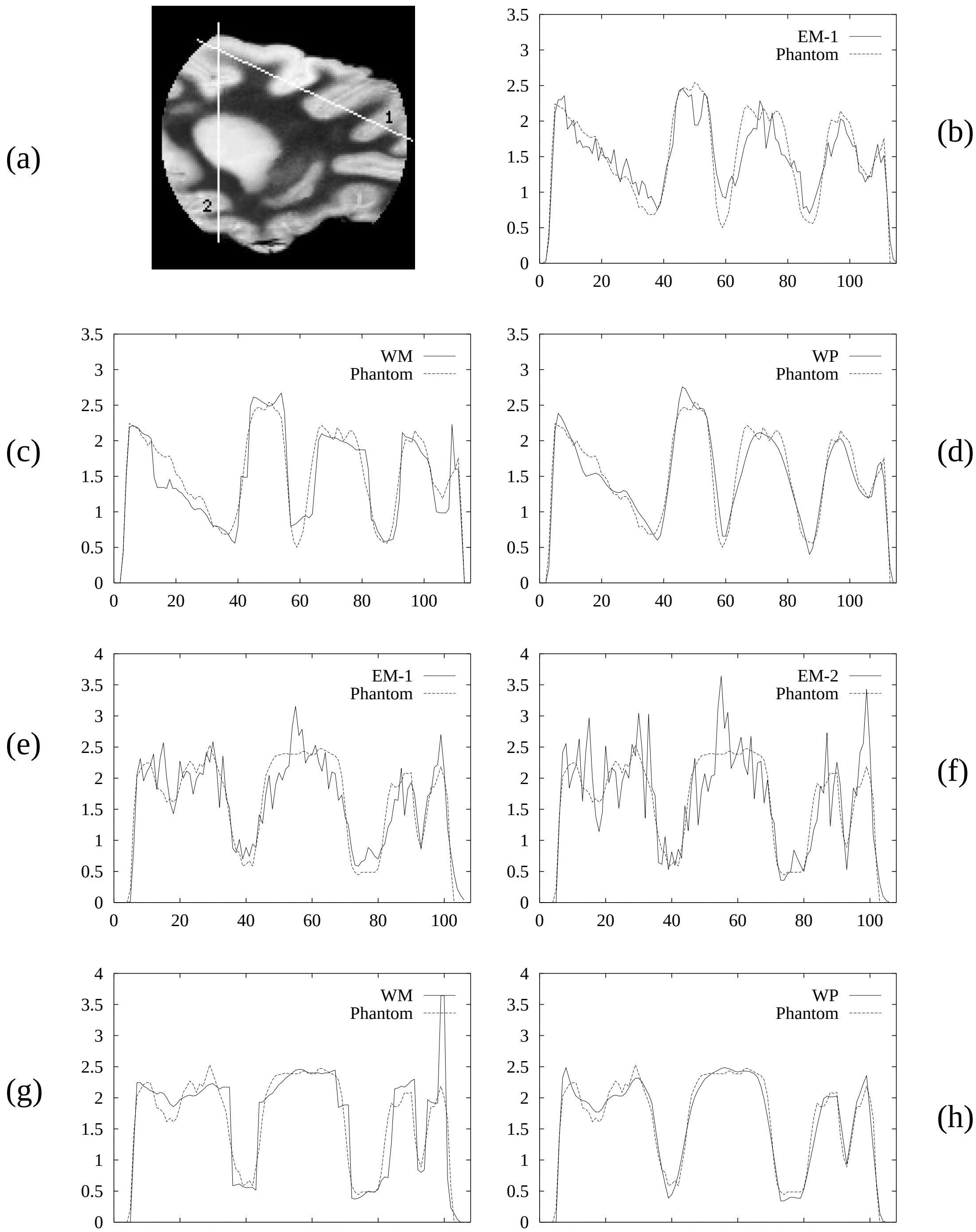

(a)
1
2
(b)
EM-1
Phantom
(c)
WM
Phantom
(d)
WP
Phantom
(e)
EM-1
Phantom
(f)
EM-2
Phantom
(g)
WM
Phantom
(h)
WP
Phantom

Figure 7:

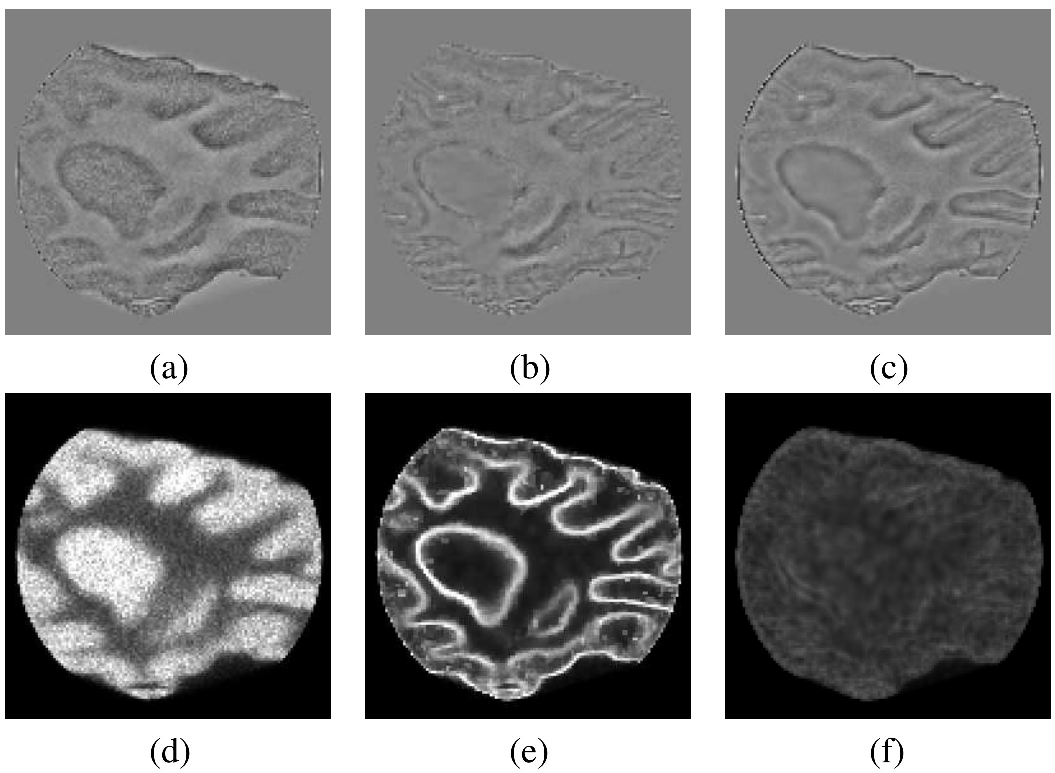
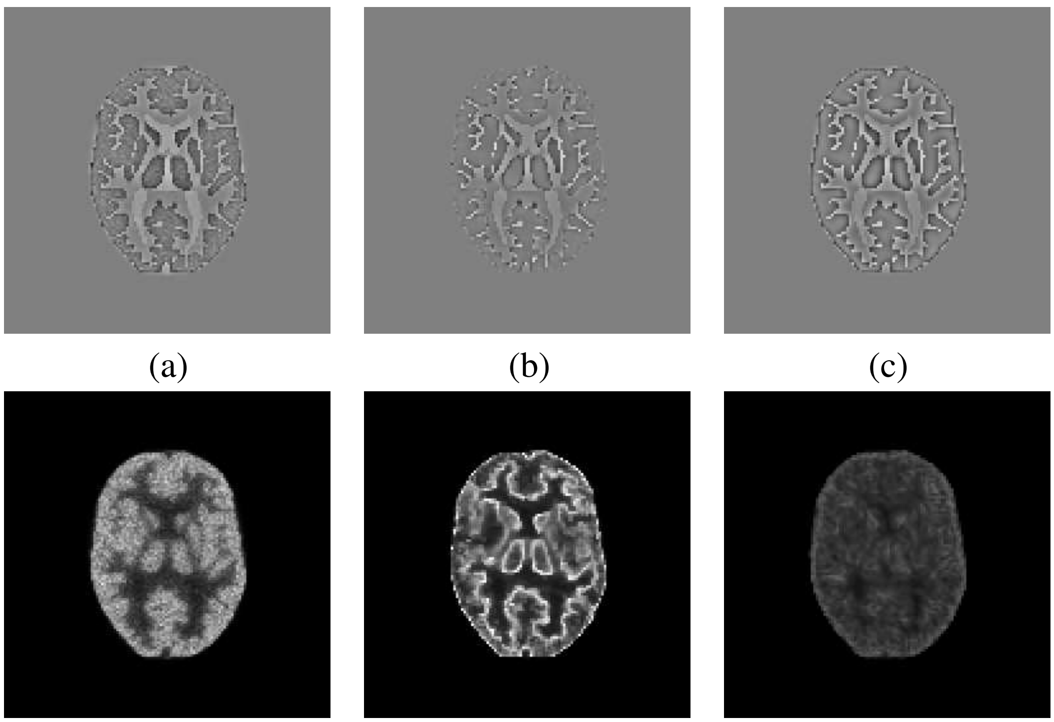


Figure 8:



(a) (b) (c)

(d) (e) (f)

Figure 9:

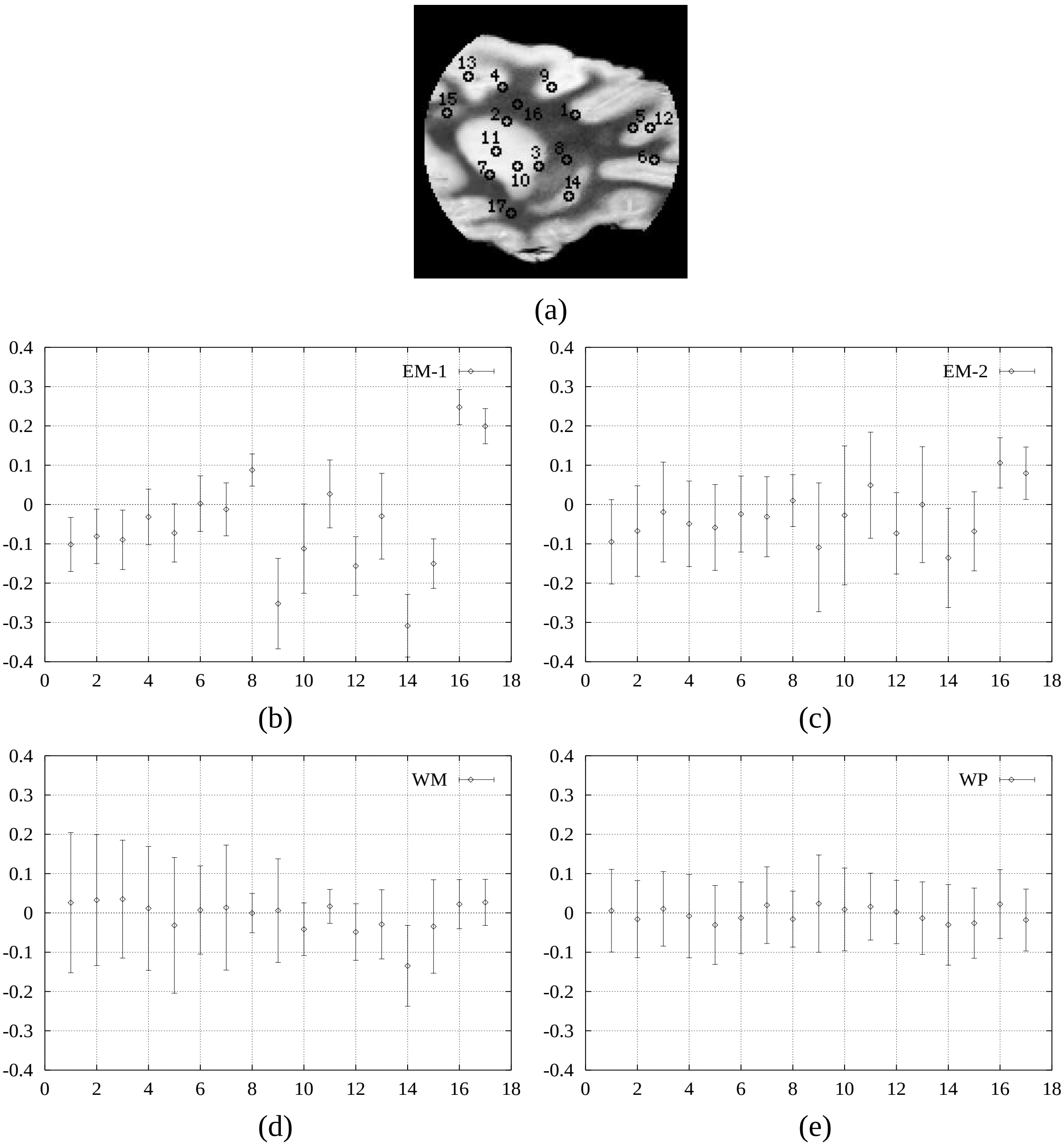

13
4
9
15
2
16
1
5
12
11
3
8
6
7
10
14
17
(a)
EM-1
EM-2
WM
WP
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0 2 4 6 8 10 12 14 16 18
(b)
(c)
(d)
(e)

Figure 10:

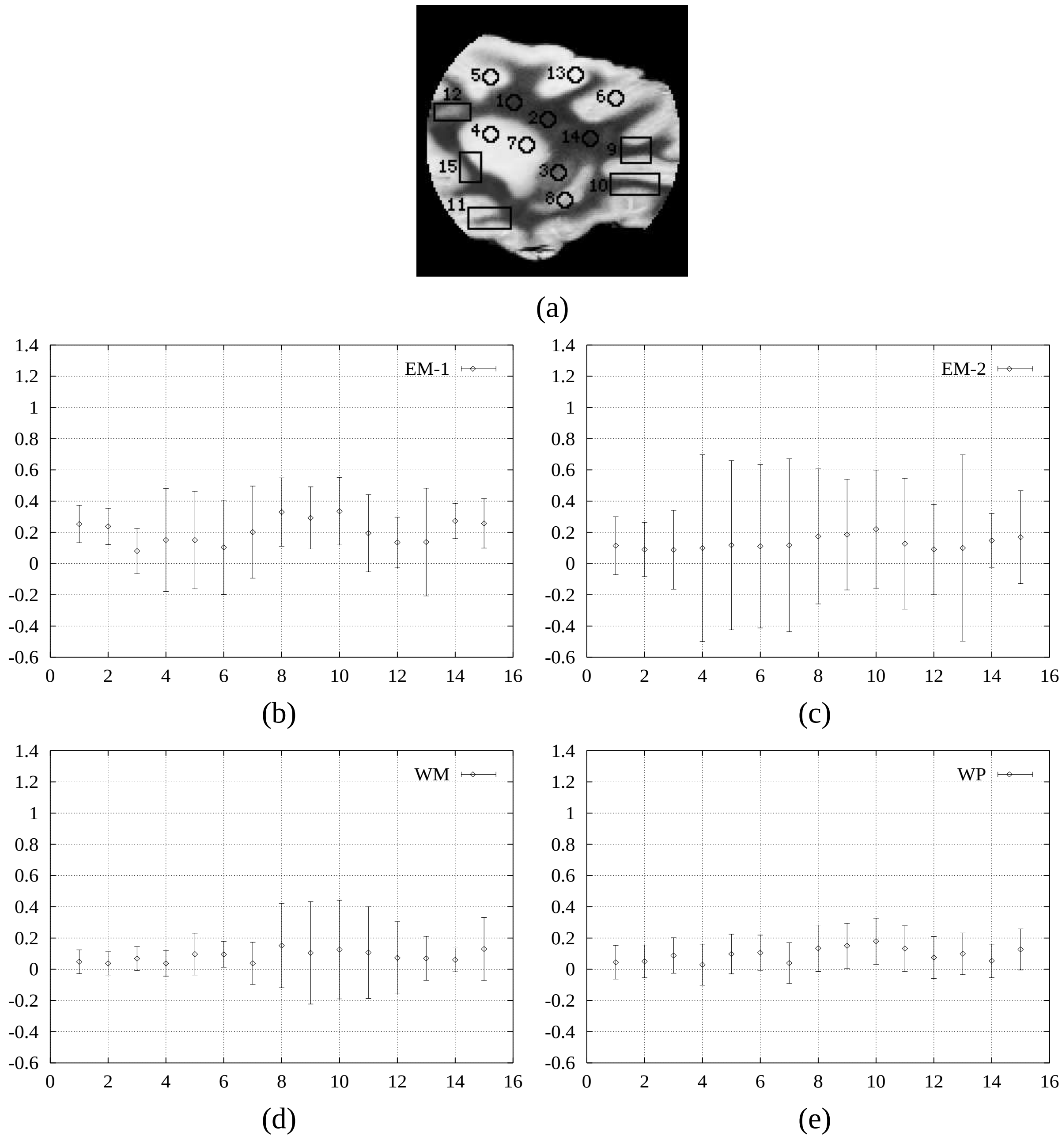


Figure 11: