LA-UR-99-0883
Approved for public release;
distribution is unlimited.

Title: **Bayesian estimation of regularization parameters for deformable surface models**

Author(s): Gregory S. Cunningham, Andre Lehovich, and Kenneth M. Hanson

Submitted to:

http://lib-www.lanl.gov/la-pubs/00818453.pdf

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

FORM 836 (10/96)

Medical Imaging: Image Processing, K.M. Hanson, ed., *SPIE Proc.* **3661**, pp. 562-573, 1999.

Bayesian estimation of regularization parameters for deformable surface models

Gregory S. Cunningham [a], Andre Lehovich [b], and Kenneth M. Hanson [a]

[a] Los Alamos National Laboratory, MS P940, Los Alamos, NM

[b] University of Arizona, Dept. of Applied Mathematics, Tucson, AZ

ABSTRACT

In this article we build on our past attempts to reconstruct a 3D, time-varying bolus of radiotracer from first-pass data obtained by the dynamic SPECT imager, FASTSPECT, built by the University of Arizona. The object imaged is a CardioWest total artificial heart. The bolus is entirely contained in one ventricle and its associated inlet and outlet tubes. The model for the radiotracer distribution at a given time is a closed surface parameterized by 482 vertices that are connected to make 960 triangles, with nonuniform intensity variations of radiotracer allowed inside the surface on a voxel-to-voxel basis. The total curvature of the surface is minimized through the use of a weighted prior in the Bayesian framework, as is the weighted norm of the gradient of the voxellated grid. MAP estimates for the vertices, interior intensity voxels and background count level are produced. The strength of the priors, or hyperparameters, are determined by maximizing the probability of the data given the hyperparameters, called the evidence. The evidence is calculated by first assuming that the posterior is approximately normal in the values of the vertices and voxels, and then by evaluating the integral of the multi-dimensional normal distribution. This integral (which requires evaluating the determinant of a covariance matrix) is computed by applying a recent algorithm from Bai *et. al.* that calculates the needed determinant efficiently. We demonstrate that the radiotracer is highly inhomogenous in early time frames, as suspected in earlier reconstruction attempts that assumed a uniform intensity of radiotracer within the closed surface, and that the optimal choice of hyperparameters is substantially different for different time frames.

Keywords: Bayesian inference, hyperparameters, deformable models, dynamic cardiac SPECT

1. INTRODUCTION

The FASTSPECT imaging system[1], developed at the University of Arizona, has been used for first-pass tomographic imaging of the time-varying distribution of a bolus of Tc-99m pertechnetate radiotracer infused into a CardiacWest Total Artifical Heart. The FASTSPECT machine simultaneously provides 24 pinhole views of the bolus distribution evolving in time, and is unique in its ability to perform this type of dynamic imaging. The goal in obtaining first-pass tomographic data is to demonstrate that clinically important measures of heart function, such as ejection fraction and wall motion, can be quantitatively estimated without having to gate and average over many cardiac cycles, an approach necessarily utilized by single- or dual-head cardiac SPECT systems. If ejection fraction and wall motion can be estimated from first-pass data during the first few cardiac cycles, then later cycles can be used to estimate myocardial perfusion, another important indicator of heart function. If successful, the FASTSPECT approach would mean that a single, relatively cheap instrument could perform multiple diagnostic tests of cardiac function with a single bolus of radiotracer. This type of capability would be clinically valuable and affordable for use in emergency rooms across the country to do initial assessment of cardiac patients.

A traditional approach to reconstruction of the 24-view tomographic data might employ the EM method to produce the maximum likelihood estimate of the activity in each voxel of the volume being imaged. A voxel-based reconstruction can be further processed by manual or automated segmentation to yield an estimate of an isosurface of the radiotracer distribution. If the radiotracer distribution is homogeneously mixed throughout a ventricular chamber of interest, then the time-sequenced estimated isosurface yields an estimate of the ventricular volume as a function of time, and hence ejection fraction, but it also provides a great deal more information of potential clinical value, since the entire interior surface of the ventricle is revealed. It is the dynamic information about the interior surface of the ventricle that distinguishes FASTSPECT and its potential follow-up (a clinical device) from other cardiac SPECT implementations, wherein time-averaging over different cardiac cycles is needed to provide similar information.

The principal attraction of the traditional approach to tomographic reconstruction and segmentation is that it is fast, and provides nearly optimal solutions when used in medical imaging modalities, e.g., PET and MRI, where the number of views and SNR are quite high. However, for data with very low SNR and including only a limited number of views, the traditional

approach will perform poorly at estimating the surface that defines the spatial extent of the radiotracer distribution. This is due to the fact that the non-parametric, voxel-based model used in the tomographic reconstruction is underconstrained. Surface estimates performed in this way will be unacceptably noisy for the type of data analyzed in this article. For such sparse and noisy data, an approach that directly estimates the shape parameters of a time-varying surface from the raw projection data is advantageous, although potentially very time-consuming.

We have discussed the direct estimation approach in previous work, where we formulated a Bayesian estimation problem for first-pass tomographic imaging using FASTSPECT that directly estimates the time-varying (x,y,z) components of the vertices of a triangulated surface within which it is assumed the bolus is uniformly mixed in the blood pool[2-4]. In that work, we identified several deficiencies in our approach, namely 1) the inadequacy of the assumption that the radiotracer is uniformly mixed in the blood pool within the bounding surface, 2) the ad hoc nature of the L-curve approach for selection of hyperparameters, and 3) the inability to allow high-curvature kinks to form at the tube/ventricle joins. We have addressed all three of these concerns in the last year, but the focus of this article is on items 1) and 2). We allow voxel-to-voxel variations in the intensity of the radiotracer within the bounding surface, and penalize the roughness of the variations with a weighted prior that is the squared norm of the numerical gradient of the voxellated grid. We explore the feasibility of maximizing the evidence over the hyperparameters as opposed to selecting them by finding the "corner of the L-surface". We have solved the problem of allowing kinks to form at tube/ventricle joins by deweighting the prior in user-defined regions of the 3D volume, but we have not tested this capability in the context of 3D models that admit spatially-varying intensity distributions within the bounding surface, which is the focus of this article, and so we will not report any more on that capability herein.

The rest of the article is organized as follows: first we give a brief overview of FASTSPECT and the data we analyzed. Next, we formulate the Bayesian estimation problem, motivate the use of evidence, and give a brief overview of the Bai *et. al*. algorithm for calculating the determinant of large matrices. Finally, we present results and draw some conclusions.

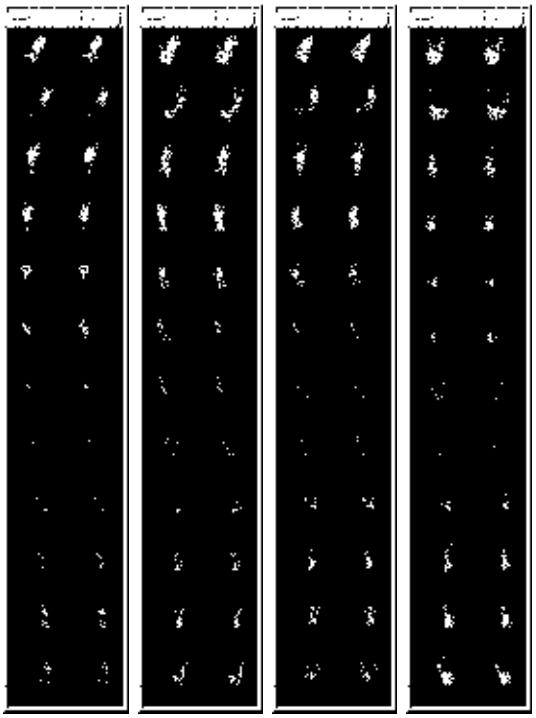
2. THE DATA

2.1. FASTSPECT

FASTSPECT is a dynamic SPECT imager that has been used for brain, heart and bone imaging[1]. Two circular arrays with a total of 24 pinhole apertures are arranged on a hemispherical dome that is roughly 35 cm in diameter. The hemisphere surrounds the volume of interest. Each pinhole is mapped to an Anger detector, and an estimate of the position of each detected photon must lie on a 64x64 uniformly-binned image grid. If a reliable estimate of the position of a photon cannot be determined, then the photon hit is discarded. Pinholes of various diameter can be inserted into the dome surrounding the object volume; 4.0 mm diameter pinholes were used to generate the data analyzed in this article. The system is characterized by a matrix, **H,** that is measured by passing a small volume element of radiotracer throughout the volume being imaged, and measuring the response of every detector pixel to that source, producing an enormous amount of information, even when compressed to take advantage of the sparsity of the matrix (150 MB of disk space after compression; 300 MB after partial uncompression in memory to speed up the calculation). The system matrix used in this article was obtained by passing a $[5\text{mm}]^3$ volume element through a 43x57x39 grid. The system matrix is noisy, since only a finite number of counts are obtained for each location of the source. Given enough patience and time, though, this noise could presumably be made as low as is needed. Note that attenuation through the dome is included in the measurement of **H**. If information is available concerning attenuating material between the radiotracer distribution and the pinholes, it can also be incorporated into **H**, and this was done for the **H** used to analyze the data discussed in this article[2].

2.2. The imaged object and raw data

The object that was imaged is a CardioWest Total Artificial Heart. Only the left and right ventricles, each about 120 ml in volume, were used. The lungs were simulated using bottles filled with water and styrofoam beads. A 20 mCi bolus of Tc-99m pertechnetate was injected into the input tube of the right ventricle at a site outside the field of view about 17 cm from the input valve of the right ventricle. The 20 mCi is first placed into 0.5 mL of a syringe and injected into a very small tube that leads to the input tube of the right ventricle. The bolus is eventually “flushed” into the input tube using a volume of water that is greater than the volume enclosed by the small tube. This process is thought to parallel the method of injection for humans, in which the injection is followed by a saline flush. Ultimately, the fluid flows into and out of a Donovan mock circulatory system, which is out of the field of view.



a) b) c) d)

Figure 1. Raw data for frames a) 46 b) 51 c) 56 and d) 61.

We present analysis results for four of the 150 50-msec frames that were available (frames 46,51,56,61) during which time the bolus just starts to enter the right ventricle from the input tube, mixes into the blood pool in the ventricle, and just starts to get pushed out the output tube. Frame 46 (Figure 1a) contained a total of 4935 counts, of which a few percent appear to be inconsistent with the assumption that the radiotracer distribution is contained within the right ventricle and its associated input and output tubes. We assume that these counts are from photons that were scattered but still accepted, although other explanations may be possible. We refer to these counts as "background" counts, and will model and estimate this background as a single constant, different in value for each time frame, but the same for all detector pixels at a given time.

3. THE BAYESIAN ESTIMATION PROBLEM

We have implemented a general tool for Bayesian estimation, in the context of image analysis using geometric models, that we call the Bayes Inference Engine (BIE)[5]. We break the Bayesian estimation problem into three parts: the object model, the measurement model, and the probability model. In the BIE, the user constructs a graphical program that transforms object and measurement system parameters into predicted data. The predicted data are compared with real data to produce a minus-log-likelihood function which is combined with prior information to produce a minus-log-posterior. The minus-log-posterior is minimized as a function of any parameters that are to be estimated. See Figure 2 for the graphical program that was used to analyze the data discussed in this article.

3.1. The object model

The object model in this case is the parametric model for the 3D radiotracer intensity distribution. In the BIE, we always convert parametric models to non-parametric ones (uniformly sampled grids) so that complex models can easily be built through combination of parametric models after conversion to a non-parametric form. The parameteric model for the bounding surface is a a set of 482 vertices with (x,y,z) components, and a connectivity network that creates 960 triangles by **x**. The lower left red box (UniformGrid3D) contains the voxellated model, **v**. The surface is transformed into a voxellated connecting vertices together. The 1446-point vector that lists the (x,y,z) components for each of the 482 vertices is denoted **x**. This parametric model is converted to a non-parametric uniformly-voxellated grid by setting the value of each voxel in **m** to the fraction of that voxel that is contained within the volume described by the triangulated surface[4]. The mask **m** is

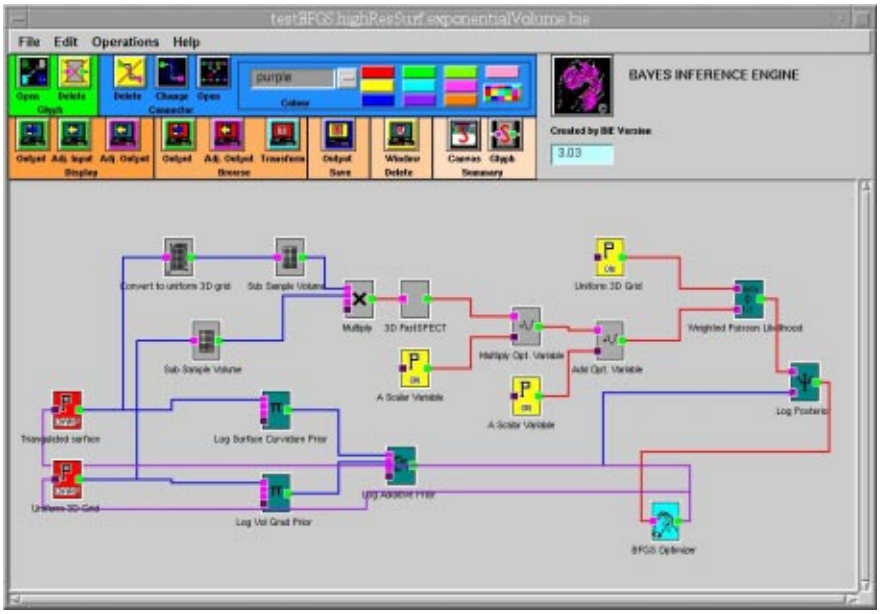


Figure 2. BIE canvas used to analyze FASTSPECT data. The upper left red box (TriangulatedSurface) contains the surface, grid (upper path) to produce a (mostly binary) mask, **m**, that is multiplied by the embedded volume model, Embed(**v**), to produce, **f**, the final voxellated grid model for the tracer distribution. **f** is multiplied by the matrix, **H**, to produce **Hf(x,v).** Finally the normalizer, I, is applied and the additive background constant, s, is added, to produce the predicted detector pixel rates **g**=I**Hf**(**x,v**)+s.

multiplied by the embedded voxellated grid, Embed(**v**), so that the final tracer distribution model, **f**=**m**(**x**)*Embed(**v**), allows for smooth voxel-to-voxel variations within the closed surface and rapid transitions in intensity across the surface boundary. **f** must have the proper dimensions in order to be operated on by the large matrix, **H**, which is the reason for the Embed(*) operation (**v** has the minimal number of parameters needed for a rectangular volume to cover the region of detectable non-zero intensity). The reader is referred to past work[2] for a brief discussion of how our approach contrasts with other deformable model approaches.

3.2. The measurement model

The measurement model uses as input the nonparametric version of the object model, **f,** and produces a set of predicted data elements, **g**, in this case a Poisson rate for each detector pixel. For the FASTSPECT machine, the measurement model is merely the very large matrix, **H**, along with a single additive constant that models the background (the same background constant is used for all 24 detectors), so that **g=**I**Hf+s.** The constant, I, is used to normalize the solution so that the voxel values, **v**, are nominally 1.0 if no intensity variation is required. The background, s**,** must be jointly estimated from the data along with the object model parameters. The nature of our object model allows us to speed up the calculation **Hf** dramatically since only a few percent of the voxels in **f** are nonzero. Simply skipping over the calculation **Hf** for values of **f** that are zero allows us to calculate **Hf** in about 300 msec on a DEC Alpha 500/500. The same speedup applies in the adjoint direction, wherein derivatives are propagated according to the chain rule in the direction opposite to the path that transforms object parameters into predicted data[5].

3.3. The probability model

The prior probability model for the surface, **x**, penalizes a discrete approximation to the local curvature at every edge shared by two triangles on the surface in order to enforce smoothness of the estimated surface[2-4]. Let $\mathbf{n_i}$ be the normal to the ith triangle. We define θ_{ij} to be the angle between $\mathbf{n_i}$ and $\mathbf{n_j}$. Then, if A_i is the area of the ith triangle, and l_{ij} is 1/3 the height of the ith triangle relative to the edge shared by triangles i and j, the curvature prior is defined as

$$\pi_{\text{surf}}(x) = \sum_i A_i \sum_j \left(\frac{\tan(\theta_{ij}/2)}{l_{ij}} \right)^2, \tag{1}$$

where **x** is the list of (x,y,z) components for each of the 482 vertices, i indexes over all triangles on the surface and j indexes over all triangles that share an edge with the ith triangle.

The prior probability model for the volume, **v**, penalizes the voxel-to-voxel variations in intensity in order to make inhomogeneities in the tracer distribution within the closed surface smooth:

$$\pi_{\text{volume}}(v) = \sum_{ijk} \left\| \nabla v_{ijk} \right\|^2 \tag{2}$$

where

$$\nabla v_{ijk} = \begin{pmatrix} (v_{i+1,j,k} - v_{i,j,k}) \text{ if } 1 \le i \le N_x; 0 \text{ else} \\ (v_{i,j+1,k} - v_{i,j,k}) \text{ if } 1 \le j \le N_y; 0 \text{ else} \\ (v_{i,j,k+1} - v_{i,j,k}) \text{ if } 1 \le k \le N_z; 0 \text{ else} \end{pmatrix}$$

where N_x=35, N_y=34, and N_z=16. Our first definition of the volume prior used the symmetric definition of numerical gradient, but this form admitted significant Nyquist-frequency oscillations in the MAP solns and so we discarded it in favor of Equation (2).

The probability model for the likelihood is the Poisson distribution with mean value equal to the predicted data (predicted detector pixel Poisson rates, **g=**I**Hf(x,v)+**s) and count values equal to the actual detector counts, **k**:

$$\begin{aligned} \phi(\mathbf{x},\mathbf{v},I,s) &= -\ln \text{Prob}[\text{data} \mid \text{predicted data}] \\ &= \Sigma_i [-k_i \ln g_i + g_i] + C(k), \end{aligned} \tag{3}$$

where C(**k**) has terms that depend only on **k.** The dependence of the predicted data **g** on the underlying parameters **x, v,** I, and s, is understood.

3.4. The estimation problem

The Bayesian estimation problem is to find the values for the object model parameters, **x, v,** that produce the maximum *a posteriori* (MAP) probability, or the minimum minus-log posterior:

$$(\mathbf{x}^{MAP}(\alpha,\beta), \mathbf{v}^{MAP}(\alpha,\beta)) = \arg\min_{x,v} [\phi(\mathbf{x},\mathbf{v},I,s) + \alpha\pi_{surf}(\mathbf{x}) + \beta\pi_{volume}(\mathbf{v})], \tag{4}$$

for some fixed values of the hyperparameters, α, β. The higher-order problem is to determine the values of α, β, from the data. One Bayesian solution to the higher-order problem is to select the hyperparameters that yield the greatest evidence for the data, where the evidence is the integral w.r.t. parameters over the joint posterior distribution of parameters and data (leaving just the probability of the data given the hyperparameters, called the evidence)[6].

Previously[2], we determined α using the L-curve[7], the continuum of 2D points,

$$(\phi(\mathbf{x}^{MAP}(\alpha)), \pi(\mathbf{x}^{MAP}(\alpha))), \tag{5}$$

parameterized by α. The value of α chosen for the final estimate was the one that yielded the point on the L-curve that was closest to the "corner". The L-curve approach is traditionally used for linear least-squares with quadratic regularization, but, to the degree that the minus-log posterior is quadratic, an L-curve approach should be reasonable. In our previous work, we were satisfied with the performance of the L-curve for selecting a reasonable value for the one hyperparameter in the problem. The ad hoc extension of the L-curve approach to problems with 2 hyperparameters is to generate an L-surface[8]:

$$(\phi(\mathbf{x}^{MAP}(\alpha,\beta),\mathbf{v}^{MAP}(\alpha,\beta),I,s), \pi_{surf}(\mathbf{x}^{MAP}(\alpha,\beta)),\pi_{volume}(\mathbf{v}^{MAP}(\alpha,\beta))), \tag{6}$$

an example of which is plotted in Figure 3.

One can search the L-surface for a point of maximum mean curvature, in analogy with the L-curve approach. The main advantage of the L-surface approach over computing the evidence is computation time, since only the MAP estimators for **x** and **v** need be found in order to implement the approach.3.5 Evaluating the evidence

3.5.1. Motivation

The evidence in the data for the choice of α,β is

$$\text{Evidence}(\alpha,\beta) = \Pr[k \mid \alpha,\beta] = \iint \Pr[k, x, v \mid \alpha,\beta]\, dx\, dv \tag{7}$$

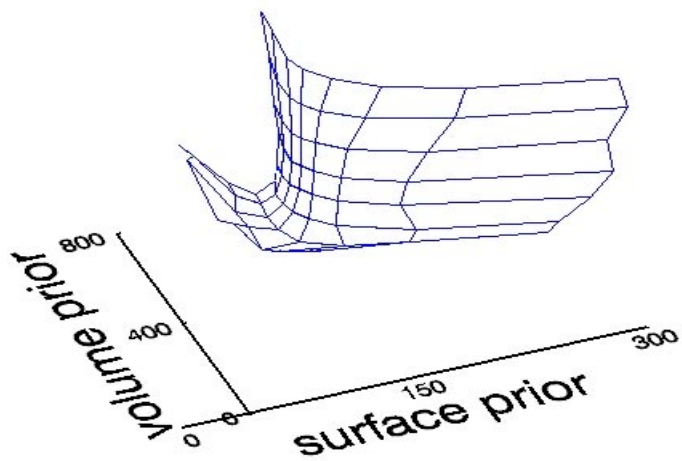


Figure 3. The L-surface for frame 46. The height above the plane is the minus-log-likelihood. The point of maximum curvature appears to be on the far left (very strong surface prior) and midway up the curve (medium-strength volume prior).

If one conceptually thinks of the pair (α,β) as the indices over a continuum of hypotheses for prior probability models of the surface and volume, then maximizing the evidence over (α,β) is equivalent to using a likelihood ratio test to select the hypothesis with the most evidence in its favor (the one that the data best supports). We attempt to perform that maximization to select hyperparameter values[9], as an alternative to the L-surface approach:

$$(\alpha,\beta)_{ML} = \arg\max_{(\alpha,\beta)} \text{Evidence}(\alpha,\beta)$$

3.5.2. The normal approximation

The problem, of course, is that evaluating Equation (7) is nontrivial in general, both analytically and computationally. However, if the joint posterior is normal in **k,x,** and **v,** then the problem is tractable, and the evidence can be evaluated analytically. Thus, we assume that a quadratic expansion of the minus-log-likelihood and minus-log-priors are good approximations in the neighborhood of the MAP solution for each (α,β), and we evaluate the integral of the multi-dimensional normal distribution that results from this approximation to produce an approximate value for the true evidence.

The expansion of the minus-log-likelihood is:

$$\Pr[k \mid x,v] = C(k)e^{-\phi(k,x,v,I,s)} \sim C(k)e^{-\left[\phi_0 + \left(\frac{d\phi}{dz}\right)^T (z-z_{\text{MAP}}) + \frac{1}{2}(z-z_{\text{MAP}})^T J^T \frac{d^2\phi}{dg_i dg_j} J (z-z_{\text{MAP}})\right]} \tag{8}$$

where **z** is the concatenation of **x** and **v**, **J** is the Jacobian of the predicted data, **g**, wrt **z**, and we use the notation **Ω** to indicate the Hessian of the minus log likelihood wrt **g**, so that an approximation to the total Hessian of the minus-log-likelihood wrt **z** is $\mathbf{J}^T \Omega \mathbf{J}$.

The expansion of the minus-log-prior for the surface is

$$\Pr[x \mid \alpha] = Ce^{-\alpha\pi_{\text{surf}}(x)} \sim Ce^{-\alpha\left[\left(\frac{d\pi_{\text{surf}}}{dx}\right)^T (x-x_{\text{MAP}}) + \frac{1}{2}(x-x_{\text{MAP}})^T C_{\text{surf}} (x-x_{\text{MAP}})\right]} \tag{9}$$

where $C \sim [(2\pi)^{N/2}\det(\alpha C_{\text{surf}})^{-1/2}]^{-1}$ in order for the prior to integrate to 1. The approximation for C is good if the mean of the normal approximation to the prior distribution contained in Equation (9) doesn't change significantly as α changes, which we have observed in practice. A similar expansion holds for the volume prior with equality for the normalizing constant since this prior is truly a multi-dimensional normal distribution. The posterior, then, is approximated as

$$\Pr[k,x,v \mid \alpha,\beta] \sim [(2\pi)^{N/2}(\det \alpha C_{\text{surf}})^{-1/2}(2\pi)^{M/2}(\det \beta C_{\text{volume}})^{-1/2}]^{-1} C(k)e^{-\phi_0} e^{-\frac{1}{2}\left[(z-z_{\text{MAP}})^T \left[J^T \Omega J + \begin{pmatrix} \alpha C_{\text{surf}} & 0 \\ 0 & \beta C_{\text{volume}} \end{pmatrix}\right](z-z_{\text{MAP}})\right]}$$

(10)

where N=1446 is the dimension of **x** and M=19040 is the dimension of **v**. The minus-log-evidence is easily evaluated to be

where we have ignored terms in Equation (11) that are independent of α,β. Note that C_{surf} can depend on α, but that C_{volume} does not depend on β.

$$-\ln \text{Evidence}(\alpha,\beta) \sim \phi_o - \frac{1}{2}\ln\det[\beta C_{\text{volume}}] - \frac{1}{2}\ln\det[\alpha C_{\text{surf}}] + \frac{1}{2}\ln\det\left[J^T\Omega J + \begin{pmatrix} \alpha C_{\text{surf}} & 0 \\ 0 & \beta C_{\text{volume}} \end{pmatrix}\right] \quad (11)$$

3.5.3. The Bai *et. al.* algorithm

In order to evaluate the evidence, then, we need to compute the determinant of several large matrices. We use the method of Bai *et. al.*,[10] who have discussed a fast Monte Carlo approach that produces noisy samples of the desired quantity, with the mean of the distribution of these samples being equal to the desired quantity. The algorithm utilizes the simple fact that ln(det(M)), where M is a positive-definite symmetric matrix, is equal to the sum of the natural log of the eigenvalues of M, or, is equal to the trace of "lnM", where lnM can be defined as the limit of a sequence of weighted sums of polynomials in M, which are well-defined.

In order to understand how the algorithm works, consider a simpler problem. If one were interested in evaluating the trace of M itself, a simple Monte-Carlo algorithm to do so is the following: For i=1,N: 1) generate a white, gaussian vector, $\mathbf{w_i}$, 2) evaluate $s_i=\mathbf{w_i}^T M\, \mathbf{w_i}$, 3) evaluate the sum $S_i=(1/N)\sum s_i$. S strongly converges to the trace of M. The extension of this simple algorithm to evaluating the trace of polynomials in M is obvious: evaluate $s_i=\mathbf{w_i}^T\, p(M)\, \mathbf{w_i}$ in step 2), where p(*) is the polynomial function of interest. Note that $\mathbf{w}^T p(M)\, \mathbf{w}$ can be viewed as a Riemann-Stieltjes integral of p(M) over the eigenspace of M, with discrete increments in the measure at the eigenvalues of M that have jump value equal to the norm-squared of **w** projected onto the corresponding eigenvector. Viewed this way, one can construct a sequence of orthonormal polynomials wrt to this measure and evaluate Gaussian quadrature approximations to the integral of lnM over the measure. This is essentially the algorithm. The generation of the orthonormal polynomials equates to generating the Lanczos iterates (which produce a tri-diagonalization of the Krylov matrix for M).

In a technical note[11], Golub and von Matt use this algorithm to find an optimal value for a regularization parameter. The authors determine that the best approach is to use only one Monte Carlo sample (N=1 in the algorithm above) for each possible value of the regularization parameter, since the minimum of the minus-log-evidence as a function of the regularization parameter is robust to the stochastic nature of the samples. We tested this empirical observation on our problem with 2 hyperparameters and report the results in the next section.

3.5.4. Implementation problems

We had several problems in implementing the Bai *et. al.* algorithm for computing the needed determinants. Firstly, the Hessians for the minus-log-priors and minus-log-likelihood are not computed and stored in memory, nor can they be directly accessed to operate on vectors as needed by the Bai *et. al.* algorithm. Instead, multiplying them by a vector **e** can only be done approximately by perturbing the MAP solutions by a small amount in the direction of **e** and re-evaluating the minus-log-probability function and its derivative wrt **x** and **v**. Then, the effect of the Hessian on **e** can be reconstructed as, e.g.

$$C_{surf}\, e \sim \frac{\left[\dfrac{d\pi_{surf}}{dx}\right]_{x+pe} - \left[\dfrac{d\pi_{surf}}{dx}\right]_{x}}{p} \quad (12)$$

A problem that we encountered was that the perturbation size, p, has to be quite small so that nonquadratic behavior can be avoided, but has to be large enough so that numerical noise does not dominate the result. The size of this perturbation depends on the parameters. Non-quadratic behavior exists for many parts of our problem. For example, the surface prior is non-quadratic when α is small and very ragged surfaces result as the MAP solution. The minus-log-likelihood is not quadratic because of the low actual count level in the detector (the maximum count level in most detectors was around 6, and many detector pixels had 0 counts). For small values of β, the MAP estimate for the volume had negative entries when unconstrained, and predicted detector rates also were negative. However, when computing the minus-log-likelihood, a lower threshold was done on the predicted detector rates so that the logarithm could be evaluated. This type of situation results in non-quadratic behavior of the minus-log-likelihood as a function of the underlying volume parameters, **v**, in the

neighborhood of the MAP soln. Our solution was to employ a small, positive lower bound on the values in v, and/or use exponential reparameterization (constrained linear model vs. unconstrained nonlinear model).

Another problem that we encountered early on was the non-empty null space in the Hessians. If the Hessian has a null space, then its determinant is zero, and so the evidence cannot be computed (the integral doesn't exist). A simple example of this is the fact that a constant-valued volume is in the null space of C_{volume} since the volume prior only penalizes functions that have spatial variation (any constant-valued volume has zero volume prior value). This is easy to remedy, though, since one only has to remove the DC part of the volume manually before probing C_{volume}. A more subtle example is that infinitesimal perturbations to vertices along the tangent plane of the surface are in the null space of the Hessians for both the minus-log-likelihood as well as the minus-log-prior since movement of vertices in the tangent plane do not make a difference in the resultant mask, **m**, nor do they significantly change the angles between triangles that share edges on which the vertex lies. However, this is only true if the surface is highly sampled, which is one reason we used a greater number of vertices in this work than in our previous work (we used roughly 3x the number of vertices). Our remedy here was to remove the part of the surface perturbations that were in the tangent plane, which reduces the number of degrees of freedom in the surface model by 2/3!

A final problem worth documenting is that in generating the Lanczos iterates, we found it necessary to perform complete reorthonormalization to make the iteration converge. This topic is not discussed in the Bai *et. al.* paper, but is discussed at length in the excellent overview of matrix computations by Golub[12].

4. RESULTS

We analyzed 4 of the 100 50-msec frames during which the radiotracer is at least part of the way inside the ventricle. Our first goal was to make the optimization completely automated so that we could generate large tables of values for the evidence as a function of the hyperparameters. As a first step, the MAP solutions are needed for large tables of values of the

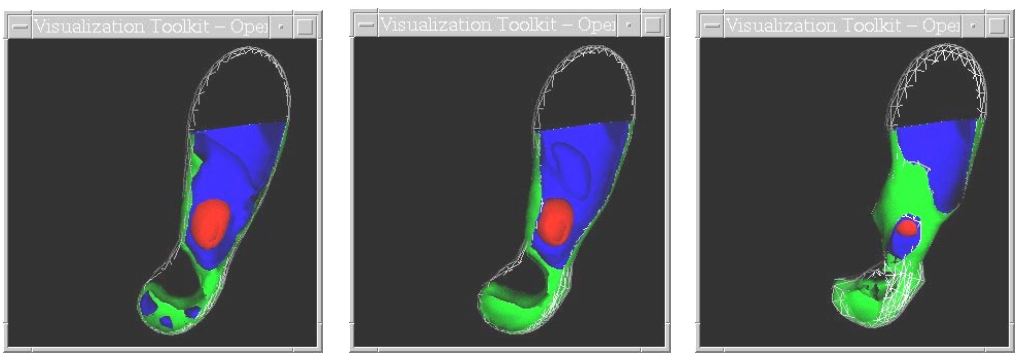


Figure 5. Volume-rendered display of the MAP estimates for the surface and volume for frame 46 using a) α=3.16, β=0.56, b) α=1.0, β=1.0, and c) α=0.2, β=2.2. All 3 MAP estimates have approximately the same likelihood value. The faint outline is a cut through the surface model, while the inner ball is the isosurface of the volume model at half full intensity and the next two isosurfaces are at quarter and eighth full intensity.merely restarted the process at the state immediately after the point in the table at which the code got stuck).

hyperparameters. We employed the BFGS algorithm with backtracking to obtain the MAP solutions over this table of values for the hyperparameters. We set the initial inverse Hessian associated with **x** to the identity matrix times the inverse value of the smallest eigenvalue of C_{surf} (at the MAP solution obtained for a large value of α), and similarly for the inverse Hessian associated with **v**. The scalar multipliers associated with the initial inverse Hessians for **x** and **v** were different by about a factor of 30. We were able to generate 9x9 tables of MAP solutions spanning two orders of magnitude for α and β (0.1 to 10.0 for both) without encountering a software interrupt of the process approximately ½ the time (the other ½ of the time we The range in the minus-log-likelihood for this tabular evaluation for frame 46 is about 500. Several MAP solutions for frame 46 that all have approximately the same value of minus-log-likelihood are shown in Figure 5.

At this early time in the movement of the tracer bolus it is just entering the right ventricle, and there appears to be massive inhomogeneity in intensity value throughout the volume bounded by the surface model, as displayed in Figure 5 by the small hotspot just above what is presumably the valve between the input tube and the ventricle. The tracer appears to have

preferentially squirted through one side of the valve opening (the author was unsure at publication time as to how and where the valve is hinged in the artificial heart).

Our next goal was to automatically generate the approximations to the minus-log-evidence over the same large table of values for α and β. Once the implementation details in the previous subsection were identified and fixed, this part of the code seemed to perform quite robustly, with perhaps 2-3% of the minus-log-evidence calculations appearing to be outliers, perhaps due to poor convergence of our implementation of the Bai *et. al.* algorithm in these cases. The first study we did (in Figure 6) was to investigate whether the minimum in minus-log-evidence corresponded to something reasonable, and whether it was stable over more than one stochastic realization, as earlier proposed.

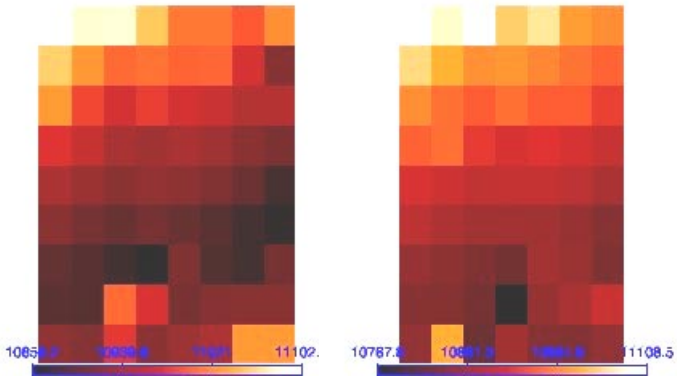


Figure 6. The minus-log-evidence for frame 46 as a function of α and β for a) realization number 1 and b) realization number 2. The range for both α and β is 10.0 (upper left) to 0.1 (lower right) and they are logarithmically sampled over this range. Dark indicates lower minus-log-evidence, and thus higher evidence (more probable). The last column of the image in a) and the last 2 columns in the image in b) have not been computed.

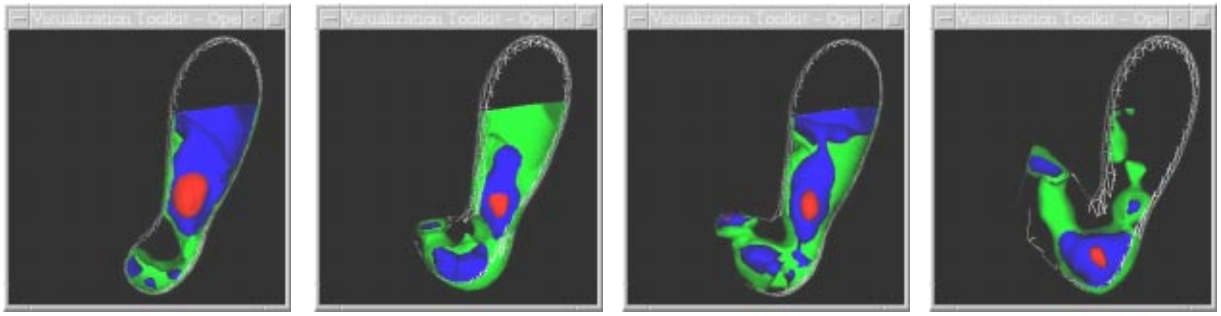


Figure 7. MAP estimates for frames a) 46, b) 51, c) 56 and d) 61 using α=3.16, β=0.56 (the optimal values for frame 46 using the evidence).

There are two interesting conclusions to draw from Figure 6. First, it appears that there is very little evidence in frame 46 data for choosing the model that says weak surface prior/strong volume prior over one that says weak volume prior/strong surface prior (or vice versa), as indicated by the relatively broad valley running from lower left to upper right. This is probably because the volume prior is weak enough throughout this valley that it can properly reconstruct the hot spot on the top side of the valve, and the extremely low intensity on the other side of the valve is so low that there are very few counts in the data to support it, and so one could adequately represent it with either model. The repeatable shape and location of the valley over 2 realizations is somewhat reassuring that making decisions based on single realizations may be applicable here.

When we apply the optimal values for α and β obtained from Figure 6 (α~3.16, β~0.56) to frames 51, 56 and 61, we obtain a sequence of MAP estimates for the tracer intensity that shows persistent inhomogeneity even as the tracer is beginning to be pushed up the output tube (Figure 7). Note that even though the tracer has been almost fully ejected from the input tube, the hot spot has not uniformly mixed in the ventricle. The inability of the surface to admit high-curvature kinks at the tube/ventricle join is apparent in frames 56 and 61 (Figure 7c and 7d). The tracer appears to be concentrated well below

the bottom of the surface in the "saddle" region between the input and output tubes, but the strong curvature prior will apparently not allow the surface to track this interface. This result prompted us to investigate the minus-log-evidence plot for frame 51, to see if the shape and optimal values for α,β change from frame to frame (see Figure 8).

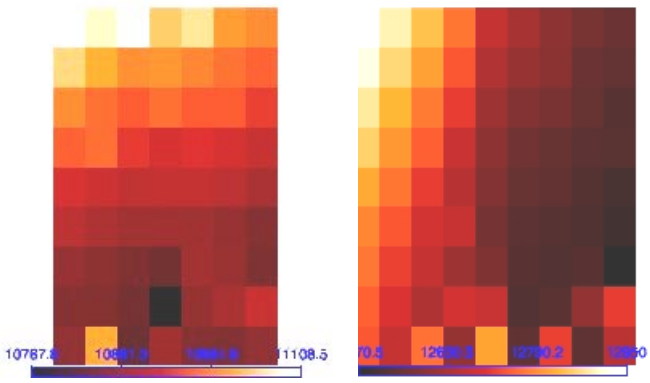


Figure 8. The minus-log-evidence as a function of α and β for a) realization number 2 for frame 46 and b) realization number 1 for frame 51. The range for both α and β is 10.0 (upper left) to 0.1 (lower right) and they are logarithmically sampled over this range. Dark indicates lower minus-log-evidence, and thus higher evidence (more probable). The last 2 columns in the image in a) have not been computed, accounting for the difference the size of a) vs. b)

It appears that the optimal values for α and β do change from frame to frame, as one might expect since the bolus needs to travel a rather torturous (high-curvature) path to get from the input tube to the output tube, and the surface prior would have to be appropriately deweighted if expected to follow this path accurately. The difference between the MAP estimates of the surface for frame 51 using the optimal values of α and β for frame 46 vs. using the optimal values for frame 51 (α~0.1, β~1.0as observed in Figure 7) show significant differences, such as the enhanced ability to follow the top of the ventricle at the input tube/ventricle join, and the enhanced ability to allow small amounts of tracer in the output tube to show up.

6. CONCLUSION

We have formulated the analysis of low-count, first-pass cardiac SPECT data in a Bayesian framework using deformable geometric models that admit smooth intensity variations within a bounding surface, across which rapid transitions in intensity can occur. In particular, the model assumes that the radiotracer distribution within the tubes and ventricle is smoothly distributed inside a volume defined by a closed, triangulated surface with 482 vertices and 960 triangles. We jointly estimate the intensities of the voxels within the bounding surface as well as the positions of the vertices of the surface from the raw data. The optimization process takes, on a DEC Alpha 500/500 (500 MB memory, 500 MHz processor), 5 minutes per frame using a BFGS optimization algorithm with backtracking and a reasonable initial inverse Hessian derived from looking at the eigenvalue structure of the Hessians for the minus-log-priors.

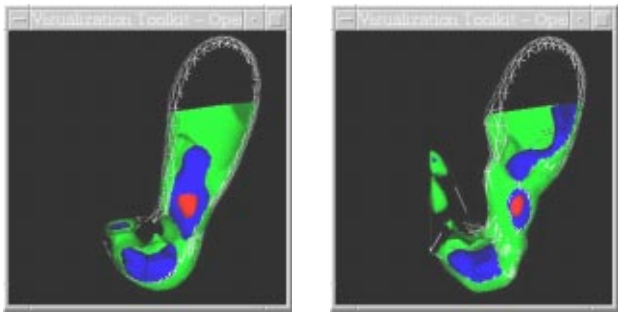


Figure 9. MAP estimates for frame 51 using a) α=3.16, β=0.56 (the optimal values of α and β for frame 46 using the evidence), and b) α=0.1, β=1.0 (the optimal values of α and β for frame 46 using the evidence).

We use the system matrix for FASTSPECT and an unknown constant additive background to model the predicted rates at the detector as a function of volumetric distributions of radiotracer parameterized by the surface and voxels. The system matrix and raw data were provided to us by the University of Arizona. The raw data consist of 24 4mm pinhole views of the radiotracer intensity distribution at each of 100 time frames. The total number of counts, integrated over all 24 detectors, is between 4935 and 5734 for each the four frames we actually analyzed.

We explored the feasibility of computing the minus-log-evidence as a function of the two hyperparameters in the model we used. We first assumed that the posterior was normal by expanding the minus-log-likelihood and minus-log-priors as quadratics. We showed that the integral of this approximate posterior can be calculated if the determinant of several Hessian matrices can be determined, and we explored the utility of a Bai *et. al.*'s fast algorithm for computing the needed determinants. Our conclusions are that the algorithm required very careful attention to details in implementation for our specific problem, but that it appears to produce results that are meaningful. In early time frames, there is no evidence to discriminate between several models that explain the data, while in later times there appears to be evidence that a weak prior on the surface is needed to explain the data properly. These results seem to jibe with the physical situation, wherein the tracer has to navigate a high-curvature spatial path to go from the input tube to the output tube. Details of this path are only observed at later time frames. The results have confirmed suspicions we developed during earlier analyses using a model that assumed the tracer was uniformly distributed within the bounding surface, namely that the tracer appears to exhibit large inhomogeneities in intensity within the bounding surface, particularly at early times, but persisting through the time when the tracer starts to be ejected by the ventricle into the output tube.

One can conceive of a number of extensions to the model that will only improve the performance of the approach, e.g., integrating surface "kinks"[13] into the 3D model that allows for spatial variation of tracer within the bounding surface. Our long-term goal is to explore the utility of incorporating hydrodynamic flow constraints on the time evolution of the 3D tracer distribution.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy under contract 7405-ENG-36. We thank Irene Pang and Harry Barrett of the University of Arizona for supplying us with the FASTSPECT data and associated system matrix, complete with attenuation correction.

REFERENCES

1. W.P. Klein, H.H. Barrett, I. W. Pang, D.D. Patton, M.M. Rogulski, J.J. Sain, and W. Smith "FASTSPECT: Electrical and mechanical design of a high resolution dynamic SPECT imager," *Conference Record of the 1995 IEEE Nucl. Sci. Symp. & Med. Imaging Conf.* (IEEE, Los Alamitos, 1996), Vol. 2, pp. 931-933.
2. G.S. Cunningham, K.M. Hanson, and X.L. Battle, "Three-dimensional reconstructions from low-count SPECT data using deformable models," *Optics Express* **2**, no. 6, pp. 227-236, 1998. Online at http://epubs.osa.org/opticsexpress.
3. X.L. Battle, G.S. Cunningham, and K.M. Hanson, "Tomographic reconstruction using 3D deformable models," *Phys. Med. Biol* **43**, pp. 983-990, 1998.
4. X.L. Battle, G.S. Cunningham, and K.M. Hanson, "3D tomographic reconstruction using geometrical models," *Proc. SPIE* (SPIE, Bellingham, 1996), Vol. 3034, pp. 346-357.
5. K.M. Hanson and G.S. Cunningham, "A computational approach to Bayesian inference," M.M. Meyer and J.L. Rosenberger, eds., *Computing Science and Statistics* (Interface Foundation, Fairfax Station, VA 22039-7460, 1996, Vol. 27, pp. 202-211.
6. D.J.C. MacKay, "Bayesian interpolation," *Neural Comput.* **4**, pp. 415-447, 1992.
7. P.C. Hansen and D.P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.* **14**, pp. 1487-1503, 1993.
8. M. Belge, M. Kilmer, and E.L. Miller, "Simultaneous multiple regularization parameter selection by means of the L-hypersurface with applications to linear inverse problems posed in the wavelet transform domain," Proc. SPIE (SPIE, Bellingham, 1998), Vol. 3459, pp 328-336.
9. R. Molina, A.K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Proc.* **8**, no. 2, pp. 231-246, 1999.
10. Z. Bai, M. Fahey, and G. Golub, "Some large-scale matrix computation problems," *J. Comp. Appl. Math.* **74**, pp. 71-89, 1996.
11. G.H. Golub and U. von Matt, "Generalized cross-validation for large-scale problems," revised version, Stanford University Technical Report.

12. G.H. Golub and C.F. Van Loan, *Matrix Compuations* (3rd edition, chapter 9.2), Johns Hopkins University Press, Baltimore, 1992.
13. K.M. Hanson, R.L. Bilisoly, and G.S. Cunningham, "Kinky tomographic reconstruction," *Proc. SPIE* (SPIE, Bellingham, 1996), Vol. 2710, pp. 156-166.