



Bayesian characterization of uncertainty in intra-subject non-rigid registration

Citation

Risholm, Petter, Firdaus Janoos, Isaiah Norton, Alex J. Golby, and William M. Wells. 2013. "Bayesian Characterization of Uncertainty in Intra-Subject Non-Rigid Registration." *Medical Image Analysis* 17 (5) (July): 538–555. doi:10.1016/j.media.2013.03.002.

Published Version

doi:10.1016/j.media.2013.03.002

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:34341843>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Published in final edited form as:

Med Image Anal. 2013 July ; 17(5): 538–555. doi:10.1016/j.media.2013.03.002.

Bayesian Characterization of Uncertainty in Intra-Subject Non-Rigid Registration

Petter Risholm^a, Firdaus Janoos^a, Isaiah Norton^b, Alex J. Golby^b, and William M. Wells III^a

Petter Risholm: pettri@bwh.harvard.edu

^aDepartment of Radiology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, MA 02115

^bDepartment of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, MA 02115

Abstract

In settings where high-level inferences are made based on registered image data, the registration uncertainty can contain important information. In this article, we propose a Bayesian non-rigid registration framework where conventional dissimilarity and regularization energies can be included in the likelihood and the prior distribution on deformations respectively through the use of Boltzmann's distribution. The posterior distribution is characterized using Markov Chain Monte Carlo (MCMC) methods with the effect of the Boltzmann temperature hyper-parameters marginalized under broad uninformative hyper-prior distributions. The MCMC chain permits estimation of the most likely deformation as well as the associated uncertainty. On synthetic examples, we demonstrate the ability of the method to identify the *maximum a posteriori* estimate and the associated posterior uncertainty, and demonstrate that the posterior distribution can be non-Gaussian. Additionally, results from registering clinical data acquired during neurosurgery for resection of brain tumor are provided; we compare the method to single transformation results from a deterministic optimizer and introduce methods that summarize the high-dimensional uncertainty. At the site of resection, the registration uncertainty increases and the marginal distribution on deformations is shown to be multi-modal.

Keywords

Bayesian; uncertainty; elastic; non-rigid; registration; MCMC

1. Introduction

The vast majority of non-rigid registration methods only report a single transformation (i.e. a *point-estimate*) result without any estimate of the registration uncertainty, which could provide valuable information whenever important clinical decisions, or high-level analysis, are based on registered data. Various factors such as the ill-posed nature of non-rigid registration, the stochastic nature of the images, the large variability of anatomy, the presence of homogeneous intensity regions, and imaging artifacts like distortion and bias-field, may negatively affect the registration uncertainty. Despite this, state-of-the-art

© 2013 Elsevier B.V. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

registration methods are typically evaluated on a limited set of representative images through estimation of the alignment error of sparse homologous landmarks, or regions, based on a single registration result. Suppose a new dataset, with new pathology or image artifacts that are not represented in the evaluation dataset, was to be registered – then it cannot be assumed that the resulting deformation point-estimate will carry the same accuracy as the previous evaluation results would indicate. If the registration result was associated with some uncertainty measure, or confidence limits, it could be possible to detect that the method had difficulty in finding a proper alignment of the images in certain locations, and the surgical risk involved with making decisions based on the registered data could be evaluated. It should be noted that a highly uncertain (low precision) registration result does not necessarily mean that the result is also inaccurate and vice versa, so systematic errors (bias) of the registration model should be evaluated before placing trust in precise registration results.

In image-guided neurosurgery, where surgeons assess the operative risk based on registered image data, misplaced confidence in registration results may have severe consequences. Important functional areas of the brain are commonly defined from pre-operative studies. If these eloquent areas are located adjacent to a tumor, it is critical that their locations are mapped as precisely as possible to the intra-operative scans. Factors such as resection of tissue, degraded intra-operative image quality, or blood and edema around a tumor, may all contribute to an increase in registration uncertainty and/or decrease of registration accuracy, especially in the vicinity of the surgical field, precisely where accuracy is most essential. For a neurosurgeon, whose goal is to maximize the resection of tumor tissue while preserving eloquent areas, information of the uncertainty in the location of the functional areas in relation to the tumor is just as important as the optimal estimate.

Current image guided navigation systems use rigid registration to map between the pre- and intra-operative space (Shamir et al., 2009; Gumprecht et al., 1999), and researchers have focused on estimating the distribution of errors in rigid registration, especially in fiducial-based registration where a set of homologous points are used to drive the rigid transform that aligns the two spaces (Maurer et al., 1997). It is generally known that the Fiducial Registration Error (FRE) is a poor predictor of real registration error (Danilchenko and Fitzpatrick, 2011). Hence, recent works have focused on characterizing the distribution of the Target Registration Error (TRE), i.e. the distance between homologous points other than the fiducials used in the FRE, after registration (Danilchenko and Fitzpatrick, 2011; Seginer, 2011). An extension to the standard rigid fiducial registration algorithm was presented by Pennec and Thirion (1997) where, given a set of matched geometric features in the form of points or frames, they estimate a rigid motion as well as the covariance matrix which is compatible with the matched points. Another well known method, the Iterative Closest Point (ICP) algorithm, which does not assume point homology, has been extended to include point uncertainties (Maier-Hein et al., 2012; Fietsen et al., 2010).

In rigid intensity-based registration, Kybic (2008) assumed normally distributed registration (translation) parameters and used the Hessian of the similarity criterion to compute confidence intervals of the registration parameters. Robinson and Milanfar (2004) used the Cramèr-Rao bounds to find a lower bound on the covariance of the parameters in a 2D-translational registration. Bansal et al. (2009) introduced a framework for quantifying errors in registration by computing the confidence intervals of the estimated parameters (translation, rotation and scale) for the similarity transform. They show that the parameters are multi-variate normally distributed and use the covariance matrix to compute the confidence interval of the transformation parameters.

Rigid registration cannot account for the general motion of brain tissue that occurs during neurosurgery which has spurred interest in non-rigid registration methods. However, few authors have attempted to quantify the full uncertainty of the estimates produced by non-rigid registration methods. Hub et al. (2009) developed a heuristic method for quantifying uncertainty in B-spline image registration by perturbing the B-spline control points and analyzing the effect on the similarity criterion. A statistical approach was presented by Kybic (2010) who considered images to be random processes and proposed a bootstrapping method to compute statistics on registration parameters. However, their method was only demonstrated on 2D datasets with low-dimensional transformation models. Gee et al. (1994, 1995a) described a Bayesian approach using a prior based on linear elasticity to regularize the problem. By using local quadratic approximations of the likelihood term, they constructed a Gibbs sampler to estimate the posterior mean (Gee et al., 1995a), while a deterministic method was used to estimate the posterior mode (Gee et al., 1994). They reported results of estimating transformation parameter variance for the case of 2D registrations. Another fast Bayesian approach for approximating the posterior uncertainty on registration parameters using variational Bayes was introduced by Simpson et al. (2012) where an approximate posterior distribution on the registration and hyper-parameters is estimated based on a set of simple parametric distributions and has shown to be useful in classification (Simpson et al., 2011a) and segmentation propagation tasks (Simpson et al., 2011b).

In Bayesian inference, typical approaches to finding the *maximum a posteriori* (MAP) estimate of non-parametric distributions involve some form of local gradient based maximization of the log posterior. Unfortunately, there are usually no guarantees that the optimization method finds the global maximum, nor are “error-bars” or credible intervals of the MAP results usually provided. Because the posterior distribution is non-parametric, and assumed to have heavy tails and possibly be multi-modal, Markov Chain Monte Carlo (MCMC) is the most principled approach to infer posterior estimates, including the uncertainty (Gelman et al., 2003). MCMC numerically approximates the posterior distribution by a set of posterior samples. While it asymptotically characterizes the entire posterior distribution, a large number of samples may be needed to characterize the posterior to the needed accuracy, and MCMC methods are consequently compromised by a high computational burden. As a result, few applications of MCMC have been reported in medical image analysis. Recently, Smal et al. (2012) applied MCMC for motion analysis of tagged magnetic resonance images. For segmentation purposes, Fan et al. (2007) used an MCMC method for sampling curves from the exponentiated negative segmentation energy and showed how to summarize posterior marginal distributions in terms of marginal probability maps. An extension of that work was presented in Chang and Fisher (2011) where they sample level sets instead of curves, which facilitates topology changes. MCMC has been widely used in functional Magnetic Resonance Imaging (fMRI) analysis (Woolrich et al., 2006; de Pasquale et al., 2008) and Diffusion Tensor Imaging (DTI) (Behrens et al., 2003; McNab et al., 2009).

The posterior distribution over deformations is a highly informative high-dimensional spatially varying object with local correlations. In this article, we introduce a Bayesian non-rigid registration framework where the posterior distribution on deformation parameters is characterized by MCMC. Although posterior approximations through, for instance, variational Bayes may provide fast estimates of the posterior, it is not known how well they approximate the unknown shape of the posterior distribution. Even though MCMC comes at a hefty computational price, and consequently may not be directly applicable in many clinical applications today, the insights gained into the complexities of the posterior distribution can be used to guide construction of new and fast posterior approximations and help point the way for designing future clinical systems that puts clinically relevant error-

bars on registration results to reduce surgical risk. The framework allows for any traditional dissimilarity and regularization energy to be included in the likelihood and prior distribution respectively using Boltzmann's distribution. We note that other authors have derived prior probability distributions on deformations from regularization energies using Gibbs distribution (e.g. Gee and Bajcsy (1999) and Papademetris et al. (2001)) which is closely related to Boltzmann's distribution.

We demonstrate the method with a Sum of Squared Differences (SSD) based likelihood, where the Boltzmann temperature models the image noise variance. The prior model over deformations is based on a linear elastic bio-mechanical model which, together with Boltzmann's distribution, implements a phenomenological approach to regularization where the Boltzmann temperature modulates the inverse stiffness of the tissue material. Higher temperatures (lower stiffness) facilitate more radical changes to the tissue configuration, while lower temperatures (higher stiffness) prevent large non-smooth deformations. Consequently, one major concern is the effect of model hyper-parameters, in our case the mechanical properties of the tissue and the variance of the image noise, on the registration results. The hyper-parameters are typically specified manually on an ad-hoc basis because it is difficult to find appropriate point-estimates for them. In the context of prostate registration (Risholm et al., 2012), we have shown that the posterior distribution is highly sensitive to the particular value of these hyper-parameters. If training data is available, e.g. in the form of homologous structures or landmarks, the hyper-parameters can be estimated. Examples of this in registration are e.g. in a classic machine learning framework for model selection through optimization of the cross-validation error (Yeo et al., 2010), or by maximum-likelihood estimation (Risholm et al., 2012). However, if no such data is available, the standard way of treating the uncertainty in hyper-parameter values in a Bayesian framework is by equipping them with uninformative hyper-prior distributions and marginalizing them out from the final registration estimates. Another advantage of marginalization of broad hyper-parameters is that the likelihood and prior can, when appropriate, become heavy tailed. Assuming they were modeled with Gaussian distributions, then the marginalized distribution can be viewed as a mixture of Gaussian distributions with varying variance, and such distributions are well equipped for handling outliers. Working with the posterior distribution under marginalized hyper-parameters is complex and computationally expensive and may necessitate fast approximations such as variational Bayes (Simpson et al., 2012). By using a system of sequential local Laplace approximations, we show how to marginalize model hyper-parameters representing image noise and tissue stiffness over broad uninformative log-normal hyper-priors (Janoos et al., 2012).

While a lot of research has gone into finding ways of reducing high-dimensional uncertainty information into simple local low-dimensional objects that can be easily visualized (Spiegelhalter et al., 2011), limited work has been reported on posterior uncertainty visualization in medical applications; including visualization of uncertainty in image guided needle insertions (Simpson et al., 2006), uncertainty in radiation dose control in radiation oncology (McCormick et al., 2004), determining uncertainty in white matter fiber orientation of diffusion tensors (Jones, 2003) and visualization of the local registration variance characterized by variational Bayes (Simpson et al., 2011a) or Gibbs sampling (Gee et al., 1995a). Gee and Bajcsy (1999) also showed how to derive Bayesian credible intervals about deformed point estimates from the posterior variance. While the main focus of this paper is non-parametric estimation of the posterior distribution over deformations, we also show techniques to extract simple and robust low-dimensional posterior summaries, such as the joint mode and the registration uncertainty, and show the importance of such summaries in the realm of neurosurgery.

2. Methods

Consider two images $m: \Omega_m \rightarrow \mathbb{R}$ and $f: \Omega_f \rightarrow \mathbb{R}$, where we want to align the moving image $m(\mathbf{x})$, $\mathbf{x} \in \Omega_m$, with a fixed image $f(\mathbf{x})$, $\mathbf{x} \in \Omega_f$ where $\Omega_f, \Omega_m \subset \mathbb{R}^d$ are the domains of the d -dimensional fixed and moving images respectively. The usual aim of non-rigid registration (Crum et al., 2004) is to find a high-dimensional displacement field $\mathbf{u}(\mathbf{x}): \mathbf{x} \in \Omega_f \mapsto \mathbf{u}(\mathbf{x}) + \mathbf{x} \in \Omega_m$ that minimizes an energy function composed of an image dissimilarity measure $E_s(\mathbf{u}; f, m)$ and a regularization term $E_r(\mathbf{u})$ on the displacements. Although this framework does not depend on the specific form of the dissimilarity energy, regularization energy or transformation model, for clarity and computational tractability, we restrict discussion in this paper to the Sum of Squared Differences (SSD) dissimilarity measure, a linear elastic regularization model and a Finite Element (FE) based transformation model.

FE-modeling (Zienkiewicz and Taylor, 2000; Bro-nielsen, 1998) is a popular method for solving bio-mechanical problems and has shown great potential in modeling tissue deformations (Miller et al., 2010). Here, the non-rigid transformation is effected through deformations of the FE mesh, which implements the bio-mechanical model based on tissue stiffness. Furthermore, it provides machinery to rapidly compute the energy functionals E_s and E_r as well as interpolating the displacement $\mathbf{u}(\mathbf{x})$ at inter-vertex positions. The FE method is described further in Appendix A.

We start by describing the Bayesian formulation of the registration problem in Section 2.1. This is followed by a description of the estimation of the posterior distribution by way of MCMC in Section 2.2, and the hyper-parameter marginalization using local Laplace approximations is described in Section 2.2.3. Finally, in Section 2.3, an algorithm for efficiently finding the mode of the posterior deformation samples is presented.

2.1. Bayesian Non-Rigid Registration Framework

In our Bayesian setting, depicted by the hierarchical graphical model in Fig. 1, the fixed and moving images, f and m respectively, are treated as random variables and are assumed to be generated by a model $m = f \circ \mathbf{u} + \epsilon$ that introduces both noise ϵ and displacement \mathbf{u} in the image generation process, where displacement is also treated as a random variable with an associated probability measure. The likelihood, $p(m | \mathbf{u}, \tau_s, f)$, expresses the probability of observing the moving image, given a deformation \mathbf{u} , under the noise model parameterized by τ_s . The prior $p(\mathbf{u} | \tau_r)$ contributes additional information to regularize the ill-posed maximum likelihood problem where τ_r controls the spatial regularization. With accurate point-estimates for the hyper-parameters $\mathcal{T} \triangleq \{\tau_r, \tau_s\}$, they can be modeled with delta distributions, or alternatively with maximum entropy distributions. In this paper we model them with log-normal distributions $p(\tau_r | \lambda_r, \mu_r)$ and $p(\tau_s | \lambda_s, \mu_s)$, where $\Gamma \triangleq \{\mu_r, \lambda_r, \mu_s, \lambda_s\}$ are the log-normal hyper-prior parameters.

As can be seen from the graphical illustration of the model in Fig. 1, the joint distribution can be factorized as follows (omitting the hyper-prior parameters Γ):

$$p(m, f, \mathbf{u}, \tau_s, \tau_r) = p(m | f, \mathbf{u}, \tau_s) p(\mathbf{u} | \tau_r) p(f) p(\tau_s) p(\tau_r). \quad (1)$$

By Bayes' theorem, the posterior distribution on deformation and hyperparameters $\mathcal{T} \triangleq \{\tau_r, \tau_s\}$ can be written as:

$$p(\mathbf{u}, \mathcal{T} | m, f) \propto p(m | \mathbf{u}, \tau_s, f) p(\mathbf{u} | \tau_r) p(\mathcal{T}). \quad (2)$$

Assuming the dissimilarity and regularization energies (E_s and E_r respectively) as the sufficient statistics of the posterior $p(\mathbf{u} | \mathcal{T}, m, f)$, then the Boltzmann's distribution obtained from these energies has maximum entropy that satisfies this sufficiency property. Modeled with Boltzmann's distribution, the likelihood with a SSD energy functional is now equivalent to a voxel-based i.i.d. Gaussian noise model with zero mean and variance τ_s . As a precursor to registration, images are often smoothed, which introduces local correlation in the noise and the i.i.d. assumption is theoretically violated. One approach to model such noise correlation is to decimate the data, or through the inclusion of a virtual decimation factor in the likelihood as proposed by Groves et al. (2011). This approach is especially important if the main goal is to infer correct values for hyper-parameters, e.g. noise and regularization/stiffness parameters (Simpson et al., 2012). However, as we will discuss in Section 4, the (in)validity of the i.i.d. assumption is not a problem in practice.

The linear elastic energy discretized by the FE-method in Appendix A takes the form of Eqn. (A.3) which is quadratic in the displacement, and the prior distribution takes a form similar to a multi-variate normal with the stiffness matrix serving the role of a sparsely banded (and singular) precision matrix. It should be noted that regularizers typically used in non-rigid registration are designed so as to not penalize rigid transformations such as rotations and/or translations. This implies invariance of the regularization energy under all deformations equivalent up to a rigid transform. Therefore, as the regularization term has a non-negligible kernel, the prior obtained in this manner is improper, i.e. does not have finite measure on the domain \mathbb{R}^D , where D is the dimensionality of \mathbf{u} . In the elastic energy model of the FE-method used here, which is only translation, and not rotation invariant, the stiffness matrix \mathbf{K} (see Eqn. (A.3)) is positive semi-definite and yields an improper prior on \mathbb{R}^D . However, in this case, when combined with the likelihood term and considering boundary constraints on the feasible values of the deformation (yielding a compact domain), this technicality does not pose any problems.

With fixed values of the temperature hyper-parameters (HPs) \mathcal{T} , the posterior distribution on the displacement field takes the following form:

$$p(\mathbf{u} | \mathcal{T}, m, f) \propto \frac{1}{Z(\mathcal{T})} \exp \left\{ -\frac{E_s(\mathbf{u}; m, f)}{2\tau_s} - \frac{E_r(\mathbf{u})}{2\tau_r} \right\}, \quad (3)$$

where $Z(\mathcal{T}) = \int_{\text{Dom}(\mathbf{u})} \exp \left\{ -\frac{E_s(\mathbf{u}; m, f)}{2\tau_s} - \frac{E_r(\mathbf{u})}{2\tau_r} \right\} d\mathbf{u}$,

and $\text{Dom}(\mathbf{u})$ denotes the domain of the transformation \mathbf{u} . The computation of the posterior for a given pair of images is complicated not only by the need to estimate the partition function $Z(\mathcal{T})$, but also by the sensitivity of the shape of the posterior distribution on the specific value of the temperatures, viz. τ_s which controls the image noise model and τ_r which determines the tissue stiffness. While the ratio of these HPs affects the location of the mode, which is important in optimization-based (i.e. *maximum a posteriori* (MAP)) registration, the absolute value of each HP affects the spread and skew of the posterior distribution and also, in case of multiple modes, the relative weighting of the modes. The Bayesian approach encourages marginalizing out these HPs as nuisance variables under an appropriate hyper-prior. Assuming a finite range in the orders of magnitude for τ_r and τ_s , i.e. $\text{Var}[\ln\tau_r] = \lambda_r^{-1}$ and $\text{Var}[\ln\tau_s] = \lambda_s^{-1}$, the temperatures are equipped with maximum entropy prior distributions: $\ln\tau_r \sim \mathcal{N}(\mu_r, \lambda_r^{-1})$ and $\ln\tau_s \sim \mathcal{N}(\mu_s, \lambda_s^{-1})$. Then the complete probability model is:

$$\ln p(\mathbf{u}, \mathcal{T} | m, f) = -\frac{1}{2} [E_r/\tau_r + E_s/\tau_s] - \ln Z(\mathcal{T}) - \frac{1}{2} [\ln(2\pi) - \ln\lambda_r + \lambda_r (\ln\tau_r - \mu_r)^2] - \frac{1}{2} [\ln(2\pi) - \ln\lambda_s + \lambda_s (\ln\tau_s - \mu_s)^2]. \quad (4)$$

The posterior distribution with hyper-parameters marginalized out is:

$$p(\mathbf{u}|m, f, \Gamma) \propto \int_{\tau_s} \int_{\tau_r} p(m|f, \mathbf{u}, \tau_s) p(\mathbf{u}|\tau_r) p(\mathcal{T}|\Gamma) d\tau_r d\tau_s. \quad (5)$$

The marginalization in combination with the likelihood defined as a function of arbitrary and possibly non-smooth images, obviates any meaningful parametric approximation of this distribution and necessitates non-parametric analysis or approximations.

2.2. Posterior Estimation through MCMC Simulation

2.2.1. Metropolis-Hastings—One method to characterize the posterior distribution $p(\mathbf{u} | m, f, \Gamma)$ with HPs marginalized (see Eqn. (5)) is to perform Metropolis Hastings (MH) sampling (Gelman et al., 2003) from the joint distribution on displacement and temperatures $p(\mathbf{u}, \mathcal{T} | m, f, \Gamma)$ (see Eqn. (4)) and discard the temperature values, effectively marginalizing them out. With $\omega \triangleq \{\mathbf{u}, \mathcal{T}\}$, the MH algorithm starts from an initial sample ω^0 and generates a proposal sample, ω' at iteration $n + 1$ from a simple proposal distribution which only depends on the previous sample $\omega' \sim \pi(\omega | \omega^n)$. The proposal sample ω' is accepted (i.e. $\omega^{n+1} = \omega'$) or rejected (i.e. $\omega^{n+1} = \omega^n$) with acceptance probability:

$$\mathcal{A}(\omega', \omega^n) = \min \left(1, \frac{p(\omega' | m, f, \Gamma) \pi(\omega^n | \omega')}{p(\omega^n | m, f, \Gamma) \pi(\omega' | \omega^n)} \right). \quad (6)$$

Because MH-MCMC has no parametric assumptions about the posterior distribution, and does not require the computation of gradients, it is general in terms of the types of likelihood functions and prior distributions that can be used. If the posterior distribution includes non-linear constraints, e.g. prevention of folding of the FE-mesh or movement of the mesh into image regions such as the skull, the proposal sample should be checked before computing the MH-criterion. If the constraints are violated, a new sample should be drawn.

With constant values of the temperature HPs, the partition function cancels in the MH criterion in Eqn. (6). However, when also sampling over temperatures \mathcal{T} , it becomes necessary to estimate the partition function ratio $Z(\mathcal{T}^n)/Z(\mathcal{T}')$ in the MH criterion. Partition ratios are often estimated with an importance sampler (Bishop, 2006), and in our case that would entail generating representative importance samples of \mathbf{u} that cover the parameter space by running the MH sampler at high temperature values. The drawback of this method is that it requires a very long MCMC chain to generate importance samples to ensure low bias and variance in the estimates of the partition function. In Section 2.2.3, we describe an alternative approach where deformations are sampled directly from the posterior distribution with hyper-parameters marginalized out using Laplace approximations.

2.2.2. Proposal Distribution—The acceptance and convergence rate of the MH sampler is dependent on the size and shape of the proposal distribution (Gelman et al., 2003). It is clear that the most effective proposal distribution would be the posterior distribution itself. However, with limited information available regarding the posterior, the typical choice of proposal distribution is a normal distribution because it can be efficiently sampled from, and because of its symmetry, the proposal distributions cancel out in the MH-criterion in Eqn. (6).

Displacement proposals \mathbf{u}' are sampled from a multi-variate normal centered on the previous displacement \mathbf{u}^n :

$$\mathbf{u}' \sim \pi(\mathbf{u}' | \mathbf{u}^n) = \mathcal{N}(\mathbf{u}' ; \mathbf{u}^n, \sigma_u^2 \Sigma). \quad (7)$$

If little was known about the posterior, a diagonal proposal covariance would be most suitable, but convergence would be slow. However, since the transformation of neighboring FE-vertices is correlated, enforcing a neighborhood structure on the correlation matrix Σ of the proposal distribution improves performance. The proposal correlation is set to $\Sigma_{ij} = \exp(-\mathcal{D}(\mathbf{v}_i, \mathbf{v}_j)/\rho)$, where \mathcal{D} is the distance between FE-vertices i and j and ρ is a constant. If a proposal sample causes folding of the deformation field, the sample is discarded and a new sample is drawn. Performance of MH sampling is characterized by the proportion of jumps that are accepted. Suppose that the proposal kernel $\sigma_u^2 \Sigma$ has the same shape as the posterior distribution, then it is known that the optimal jumping rule has acceptance rate around 23% in high dimensions (Gelman et al., 2003). In our simulations, the acceptance rate is tuned automatically to be approximately 23% by adjusting the proposal variance σ_u^2 .

2.2.3. Hyper-Parameter Marginalization—Although the hyper-prior is log-normal, the shape of the posterior distribution on temperatures is unknown. Consequently, if we were to sample temperatures, a standard choice for proposal distribution would be a univariate normal, or possibly a log-normal, centered on the previously accepted temperature. However, instead of sampling from the joint posterior on displacement and temperatures, which requires complicated tuning of three proposal scale parameters (σ_u^2 and two variances for the temperature proposal distributions) as well as partition ratio estimates via MCMC integration, we use the approach presented in Janoos et al. (2012) where hyper-parameters (HP) are marginalized using local Laplace approximations of the hyper-parameter posteriors. The problem of sampling from the posterior distribution of the registration parameters with HPs marginalized out is formulated in the variational free-energy framework, where the variational density turns out to be equivalent to the Laplace approximation of the HP posterior. Estimating the mode of the HP posterior requires derivatives of the partition function of the deformation parameter posterior, which are in turn quickly estimated using a local Laplace approximation. By applying Laplace's approximation locally during the intermediate steps of marginalizing out the HPs, the effect of this inaccuracy on the posterior density of the deformation parameter is restricted (Friston et al., 2007). A detailed description of the marginalization procedure is included in Appendix B. The MCMC scheme for sampling deformation parameters \mathbf{u} from the posterior distribution in Eqn. (4) with Laplace marginalization of hyper-parameters is summarized in Algorithm 1.

2.3. Posterior Estimates

With a collection of samples $\{\mathbf{u}^s\}_{s=1, \dots, S}$ characterizing the posterior distribution in Eqn. (5), statistics such as the MAP estimate of the deformation, its uncertainty and the uncertainty in the deformations at individual FE-vertices can be extracted, as explained next.

2.3.1. Posterior Modes—If the deformations are sampled from a distribution with fixed hyper-parameters, the joint mode on deformations can easily be found by selecting the sample with the highest un-normalized log posterior probability according to Eqn. (3). However, when marginalizing hyper-parameters the value of the partition function is required to calculate individual posterior log probabilities. A more tractable approach in this case is to use the density defined by the samples along with a non-parametric mode finding method such as *mean-shift* (Cheng, 1995).

Because the high-dimensional probability space is only sparsely sampled, the kernel density estimation for mean-shift, when directly applied on the joint distribution of deformations, is poorly conditioned. However, the Markov structure of the FE-mesh allows a factorization of the joint distribution on deformations into a product of low-dimensional conditional distributions on which the mean-shift procedure is well posed. Consider the 2-D example in Fig. 2 for illustrative purposes. Furthermore, assume a set of samples $\{\mathbf{u}^s\}_{s=1,\dots,S}$ are given, where $\mathbf{u}^s = [\mathbf{u}_a^s, \mathbf{u}_b^s, \mathbf{u}_c^s, \mathbf{u}_d^s]$ are the deformations of a triangular FE-mesh with four 2-D vertices and two triangular elements $\varepsilon_1 = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ and $\varepsilon_2 = \{\mathbf{b}, \mathbf{d}, \mathbf{c}\}$. The joint posterior distribution on deformations can then be factorized as:

$$\begin{aligned} p(\mathbf{u}_a, \mathbf{u}_b, \mathbf{u}_c, \mathbf{u}_d) &= p(\mathbf{u}_a | \mathbf{u}_b, \mathbf{u}_c) p(\mathbf{u}_b | \mathbf{u}_c, \mathbf{u}_d) p(\mathbf{u}_c | \mathbf{u}_d) p(\mathbf{u}_d) \\ &= p(\mathbf{u}_a, \mathbf{u}_b, \mathbf{u}_c) p(\mathbf{u}_b, \mathbf{u}_c, \mathbf{u}_d) / p(\mathbf{u}_b, \mathbf{u}_c). \end{aligned} \quad (8)$$

For high-dimensional joint distributions, the dimensionality of the factorized distributions is dependent on the mesh topology and in the worst case equals the number of vertices in the Markov blanket of the vertex with the highest valency. In general, the vertices are ordered according to valency and the vertex with the lowest valency is factorized first. Defining the vector $\mathbf{u}_{bc} \triangleq [\mathbf{0}, \mathbf{u}_b, \mathbf{u}_c, \mathbf{0}]$ and similarly for \mathbf{u}_{abc} and \mathbf{u}_{bcd} , then the mean-shift procedure on the joint log-posterior in Eqn. (8) factorizes according to $\delta(\mathbf{u}) = \delta(\mathbf{u}_{abc}) + \delta(\mathbf{u}_{bcd}) - \delta(\mathbf{u}_{bc})$

where $\delta(\mathbf{u}) = \frac{\sum_{s=1}^S K(\mathbf{u}^s - \mathbf{u}) \mathbf{u}^s}{\sum_{s=1}^S K(\mathbf{u}^s - \mathbf{u})}$ is the mean-shift operator and K is a radial basis function (typically a Gaussian kernel). The mean-shift estimation sets $\mathbf{u} \leftarrow \delta(\mathbf{u})$ and continues until convergence.

2.3.2. Marginal Posterior Distributions—The marginal posterior distribution of a deformation component \mathbf{u}_i is defined by (where $Dom(\mathbf{u}_i)$ denotes the domain of all components except the i -th):

$$p(\mathbf{u}_i | m, f) = \int_{Dom(\mathbf{u}_i)} p(\mathbf{u} | m, f) d\mathbf{u}_i, \quad (9)$$

i.e. the probability distribution of the movement of one FE vertex along one axis in the deformation field. In a sampling framework, the posterior samples of this particular deformation component characterizes the marginal distribution. Posterior marginal modes can easily be found with the mean shift method on the samples from the specific deformation component. The deformation estimates should fall within a certain confidence interval which can be robustly conveyed by e.g. the Inter-Quartile Range (IQR). As the IQR and displacements are only defined at FE vertices, we use linear FE-based interpolation to determine them at inter-vertex locations.

3. Results

The proposed method for characterizing the posterior distribution over deformations was evaluated on both synthetic and clinical image data. In Section 3.1 the method is evaluated on a synthetic dataset with a non-convex likelihood and known ground-truth deformation, while Section 3.2 demonstrates the feasibility of registering clinical Magnetic Resonance (MR) images acquired during neurosurgery for tumor resection. All clinical data was used with informed consent under an institutional review board approved protocol.

To evaluate the MAP mode estimated with the mean-shift method described in Section 2.3.1, the posterior mode of Eqn. (2) was also estimated with a global non-linear optimization procedure implemented in C++ and NLOpt (Johnson, 2012). The global stochastic optimizer, a Multi-Level Single-Linkage (MLSL) algorithm (Rinnooy Kan and

Timmer, 1987), was coupled with a local gradient-based Method of Moving Asymptotes (MMA) (Svanberg, 2002), and non-linear constraints to prevent flipping of tetrahedra (positive deformation Jacobian over the element) were included through an augmented Lagrangian method which adds a penalty to the objective function for any violated constraints. After characterizing the posterior distribution with Algorithm 1 on a given dataset, the deterministic optimizer was run with the temperatures fixed to the mean of the temperatures marginalized through Laplace approximations.

All algorithms were implemented in C++ with extensive use of OpenMP® for parallelization whenever feasible, and the experiments were run on Linux machines with 2×6 -core Intel Xeon™ 2.67GHz CPUs, 12GB DDR-3 RAM.

For all experiments, the material parameters of the linear elastic regularizer were initialized with a Young's modulus of $E = 1.0$ and Poisson's ratio of $\nu = 0.49$. Because the elastic model is used as a regularizer, as opposed to a pure mechanical description of the deformation determined from boundary conditions and external forces, the Young's modulus is unitless and only relative stiffness can be prescribed. Every drawn proposal sample that caused folding of the FE-mesh was rejected.

3.1. Synthetic Experiment

This section evaluates and provides insights into the complex posterior distribution using the synthetic dataset of Fig. 3 which simulates a typical deformation seen during neurosurgical procedures. A synthetic deformation field was created with a thin-plate spline deformation model (Bookstein, 1989) simulating brain shift with an approximately 12mm collapse of the brain surface towards the mid-brain. To make the problem even more challenging, a sinusoidal component was added to the deformation in the x - and y - directions with magnitude 6mm and 4mm and wavelength $w/2$ and $w/4$ respectively, where w is the image width. A clinical T1 MR-image of size $256 \times 256 \times 88$ and $1.0 \times 1.0 \times 2.0 \text{mm}^3$ voxel size acquired prior to neurosurgery was deformed by the simulated deformation field to obtain a synthetic proxy for an intra-operative image. Image intensities were normalized to the range $[0, 1]$ and Gaussian noise with mean $\mu = 0$ and variance $\sigma^2 = 0.06$ was added to the intra-operative image.

The reference image was skull stripped using Brain Extraction Tool (BET) (Smith, 2002) to create an approximate label-map of the brain tissue, which was then used to build the FE-mesh with $V = 317$ vertices and $E = 1177$ tetrahedra using the Computational Geometry Algorithms Library (CGAL) (CGAL, 2008). The minimum and maximum element volumes were 340mm^3 and 2430mm^3 respectively, while the total mesh volume was 1530cm^3 . From the simulated deformation field, the ground truth deformation of the vertices was available, with maximum absolute deformation of $[11.5, 10.9, 3.4] \text{mm}$ and mean absolute deformation of $[4.0, 2.6, 2.0] \text{mm}$.

Broad log-normal hyper-priors model the temperature parameters ($\lambda_r = 1/100$, $\mu_r = \ln(100)$, $\lambda_s = 1/200$, $\mu_s = \ln(0.01)$), and the proposal covariance matrix was constructed with $\rho = 100$. The scale of the proposal distribution $\sigma_u^2 = 0.05$ was tuned to achieve an acceptance rate of approximately 23% as described in Section 2.2.2.

3.1.1. Convergence—The MCMC sampler was evaluated with the Geweke time-series criterion (Geweke, 1991) to test the intra-chain convergence using long MCMC chains from instances of the sampler initialized at dispersed starting points. Inter-chain convergence was also evaluated with the potential scale reduction criterion (Gelman et al., 2003). The Geweke time-series approach (Geweke, 1991) applies a two-sided z test for convergence by comparing the mean of the first half of the chain with the mean of a certain number of sub-

chains from the second half of the chain. The scale reduction criterion uses a one-sided variance ratio test to check whether a set of parallel chains have converged to the same target distribution, where a large potential scale reduction indicates a lack of convergence.

We used three parallel samplers initialized within approximately 4mm in all dimensions around the ground truth and $5 \cdot 10^6$ samples were generated in each chain. After discarding the initial 20% of the samples for chain burn-in followed by thinning with a factor of 20, $2 \cdot 10^5$ samples were retained. At a sampling rate of approximately 30 samples per second, the total running time for a single chain was approximately 46 hours. Figure 4 shows intra-chain convergence evaluated with the Geweke criterion for each chain. The chains are seen to have converged with the majority of the z -scores within two standard errors of the mean. However, with low Markov chain mixing time (Gelman et al., 2003), the chains may be stuck in different regions of the probability space. The inter-chain convergence through the potential scale reduction measures the amount to which different instances of the sampler have converged to the same equilibrium distribution, which is confirmed by the results in Fig. 5. Following convergence evaluation, the samples were then pooled together for a total of $6 \cdot 10^5$ samples, which were used in the next set of results.

3.1.2. Posterior Estimates—Figures 6(a)-6(b) plots the distribution of the marginalized temperatures under the log-normal hyper-priors.

The mode of the joint distribution was estimated with the mean-shift method described in Section 2.3.1, and the results compared with the ground truth and the results from the MAP estimate of the deterministic optimizer (with temperatures fixed to the mean of the marginalized temperatures, i.e. $\tau_r = 1174$ and $\tau_s = 0.19$). From Fig. 7, it can be seen that the estimated posterior mode is highly concordant with that found by the optimizer, and has a maximum and mean absolute error in each of the three dimensions with respect to the ground truth deformation of [2.4, 1.9, 1.6]mm and [0.3, 0.2, 0.3]mm respectively. Table 1 summarizes the registration results. The marginal distribution of the component with the maximal error (2.4mm) is shown in Fig. 8(a) with the joint mode and true displacement overlaid. The reason for the misalignment may be due to the regularizer, which introduces a bias in the registration changing the location of the MAP estimate away from the ground truth. We also found that at the mode, 80 out of 951 deformation components fell outside the \pm IQR/2 range of the ground truth values as shown in Fig. 7(c), while 2/951 fell outside the 95% interval. Hence, in this case the ground truth deformation falls inside the distribution prescribed by the IQR with $(951 - 80)/951 \approx 92\%$ certainty. The marginal distribution for the component with the highest IQR (5.0mm) is shown in Fig. 8(b) while the distribution of IQRs is shown in Fig. 8(c).

3.1.3. Hyper-parameter Sensitivity—To explore the sensitivity of the posterior distribution to the hyper-parameters, we ran the MCMC sampler with a set of fixed HPs and compared the resulting posterior modes with the ground truth and the posterior summaries from the posterior with marginalized HPs. We selected four sets of fixed HPs based on over-dispersed values from the distribution of marginalized temperatures shown in Fig. 6. The specific temperature values and the resulting deviation from the ground truth is summarized in Table 2. It is clear that modeling HPs with point-estimates when having little information regarding the correct value of these HPs can negatively affect the posterior distribution. In our case, the posterior mode has a maximum error of 8.56mm when using dispersed point-estimate HPs compared to 2.41mm when marginalizing HPs (see Table 1).

3.1.4. Normality of Posterior Distribution over Deformations—To investigate the deviation from normality of the marginal distributions, Fig. 9 plots the marginal distribution of the components with the highest/lowest kurtosis and skewness estimates. A sufficient

condition to show non-normality of the posterior distribution is to show that one of the posterior marginal distributions deviates from normality. The plots indicate that the marginal distributions have a bias for leptokurtic distributions, i.e. a distribution with a heavy tail. Mardia's test (Mardia, 1970), which uses skewness and kurtosis statistics to test for multi-normality, further confirms that the posterior distribution does not pass the test for multivariate skewness nor kurtosis under a 5% significance level.

3.2. Clinical Experiments

The clinical experiments were carried out on a neurosurgical MR-dataset acquired on a Siemens Magnetom Verio 3T MR machine in the Advanced Multi-modal Imaging and Guidance Operating (AMIGO) suite at Brigham and Women's hospital. A pre-operative T2 Fast Spin Echo MR image of size $512 \times 512 \times 100$ and resolution $0.5 \times 0.5 \times 1.5 \text{mm}^3$ was acquired prior to surgery. After resection, an intra-operative T2 BLADE MR image of size $320 \times 320 \times 33$ with resolution $0.6 \times 0.6 \times 5.0 \text{mm}^3$ was acquired. Because of brain-shift and the missing tissue due to the resection, large tissue movements are observed around the resection site. Corresponding slices of the dataset can be seen in Fig. 10.

BET was used to skull-strip the pre-operative image, and the corresponding label-map defining the brain-tissue was used to construct a FE-mesh ($V = 858$ vertices and $E = 3791$ tetrahedra) which discretizes the image domain covering the brain. The minimum and maximum element volumes were 140mm^3 and 660mm^3 respectively while the total element volume was 1554cm^3 .

Both the pre- and intra-operative images went through a number of preprocessing stages: 1) the image intensities were normalized through histogram matching, 2) normalization of the intensities to the interval $[0, 1]$, 3) subsampling the pre-operative image by a factor of 2 in all dimensions to reduce the data size, and 4) Gaussian smoothing with variance 2mm^2 .

To assess MCMC chain convergence, three MCMC parallel chains were generated from dispersed starting points ($\pm 4 \text{mm}$ in each dimension) around $\mathbf{u} = \mathbf{0}$. Each chain used identical parameters with the log-normal hyper-prior parameters set to $\lambda_r = 1/20$, $\mu_r = \ln(100)$, $\lambda_s = 1/200$, $\mu_s = \ln(0.01)$. This represents nearly identical hyper-prior parameters as used in the synthetic experiments, with the exception of an increased precision (λ_p) of the regularization hyper-prior to incorporate the a-priori knowledge that deformations observed in the clinical case require more regularization due to missing correspondences due to resection and manipulation of tissue during the procedure. The proposal covariance was constructed with $\rho = 100$, and the proposal scale was tuned to $\sigma_u^2 = 0.12$ to ensure an acceptance rate of approximately 23%. A total number of $5 \cdot 10^6$ samples were generated for each chain, the first $1 \cdot 10^6$ samples were thrown away as burn-in samples and the rest of the samples were thinned out with a factor of 20. With a sampling rate of 27 samples per second, the total simulation time was approximately 50 hours. The chains were found to have converged because 99.5% of the Geweke z-scores were within 2σ and the maximum scale reduction was 1.1. The remaining samples from the three MCMC chains were pooled together for a total of $6 \cdot 10^5$ samples that were used in the following set of results.

Figure 6(c)–6(d) shows histogram plots of the marginalized temperatures under the log-normal hyper-priors using the method in Section 2.2.3. The MCMC posterior mode was compared to the MAP estimate computed with the deterministic optimizer (with temperatures fixed to the mean of the marginalized temperatures, i.e. $\tau_s = 1.59$ and $\tau_r = 7528$).

3.2.1. Posterior Summaries—Figure 9(f) shows a scatter plot of skewness vs. kurtosis for the marginal distributions which are biased toward leptokurtic distributions. The

minimum and maximum skewness was found to be -0.75 and 1.95 , while the kurtosis was in the range $[-0.50, 6.64]$ with a median of 0.11 . It is evident that the posterior distribution is not normally distributed which is confirmed by Mardia's test for multivariate normality (Mardia, 1970) which rejects the null-hypothesis that the posterior distribution is multivariate normal under a 5% significance level, both with regards to kurtosis as well as skewness. The non-normality of the posterior distribution is further emphasized by Fig. 11 where we show marginal distributions with multiple modes.

Table 3 summarizes the results in terms of the distance between the MAP estimate of the deterministic optimizer, the joint mode of posterior deformation samples found with mean shift, as well as the marginal posterior deformation modes found with mean shift. The results indicate that the marginal modes have a maximum error of 1.85mm from the joint mode, while the joint and optimizer modes match well with a maximum absolute distance of 0.45mm . The discrepancy between the marginal and joint mode is natural for high-dimensional non-Gaussian distributions.

3.2.2. Clinically Relevant Posterior Summaries—In Risholm et al. (2010) we introduced posterior predictive summaries for portraying the location of deformed DTI and fMRI activated areas. Given S posterior deformation samples, a marginal posterior distribution on whether a voxel in the intra-operative domain is inside or outside a functional activated area is constructed by deforming the fMRI activated area with each of the S deformation samples and counting how many times a voxel fell inside the deformed area. Similarly, fiber tracts extracted from DTI are deformed with each of the S deformation samples and a voxel visitation count volume is created by incrementing each voxel a deformed fiber crosses. To highlight the clinical importance of characterizing deformation uncertainty in neurosurgery, we compare a posterior predictive distribution of a motion activated center as found by fMRI with residual tumor as delineated by a neurosurgeon in Fig. 13. Notice that the posterior predictive distribution may provide important information which can make it easier for the surgeon to minimize the surgical risk involved with resecting more of the tumor.

In Fig. 11(a) we visualize the expected deformed pre-operative image with regards to the posterior distribution. With a Gaussian posterior distribution, this posterior expected deformed image could be constructed by first deforming the pre-operative image with the MAP estimate followed by Gaussian smoothing with spatial variance according to the registration uncertainty. However, since the posterior distribution is clearly not Gaussian, a more accurate posterior expected image can be estimated according to

$\mathbb{E}_{p(\mathbf{u}|f,m)}(f \circ \mathbf{u}) = \frac{1}{S} \sum_{s=1}^S f \circ \mathbf{u}^s$ which takes into account the skewness and leptokurticity of the posterior distribution. In such a posterior expected image, areas of high uncertainty will be relatively blurry compared to areas of high certainty. The posterior expected image shows the resection area as more blurry compared to areas away from the resection. In the same figure (Fig. 11) we show the corresponding spatial uncertainty in terms of the IQR, and it can be seen that the uncertainty around the resection is considerably higher than in other areas. To investigate this further we plot the marginal distributions for the x -, y - and z -components of the FE vertex that is closest to the resection cavity and find that the marginal distributions have a high spread (IQR) and are multi-modal.

In Fig. 12 we visualize the directional uncertainty in terms of the IQRs along the x -, y - and z -directions. The IQRs indicate that the uncertainty is highest at the resection site, but also high along the boundary of the mesh/brain which is caused by the reduced valence of the boundary vertices. One approach to reduce the registration uncertainty along the boundary of the mesh is to include suitable boundary conditions in the registration, e.g. to add an

additional prior distribution which prevents boundary vertices from moving into the skull extracted from the intra-operative image. In our framework, this can be implemented by discarding proposal samples which move a boundary vertex into the skull.

4. Discussion

It has been our goal in this investigation to characterize the posterior distribution in some intra-subject non-rigid registration problems. Our longer range goal is to provide uncertainty information to clinical end-users, because we feel that it should be factored into decision making or treatment optimization. For example, communicating the distributional information in Fig. 13(c) gives the hypothetical end-user the option to discount the utility of the registration information if the uncertainty in the result is too large; this opportunity is not available if the MAP result, i.e. Fig. 13(b), is the only information about the registration result that is provided.

A Bayesian non-rigid registration framework, that characterizes the posterior distribution on deformations by MCMC, was presented. The likelihood and prior distribution on deformations is constructed by converting dissimilarity and regularization energies into probability density functions by way of Boltzmann's distribution. Proper treatment of hyper-parameters in registration, e.g. noise precision and the tissue parameters/regularization weight, is an important but often neglected problem. Non-rigid registration is a highly under-determined problem and therefore generally needs a prior (regularizer) to reduce the solution space. The prior introduces a bias, but with a large reduction in uncertainty. Decreasing the importance of the prior, e.g. by setting the prior temperature to a large value, effectively makes the regularization negligible which increases the posterior uncertainty. Even though registration results are sensitive to these parameters, they are in current practice often set on an ad-hoc basis. One approach to determine hyper-parameters that are highly informative for a specific clinical application is to learn them from a collection of representative clinical datasets where homologous landmarks or structures are identified (Risholm et al., 2012). However, in cases where such data is not available, and little information is available regarding the optimal value of these parameters, the recommended Bayesian approach to handle them is to equip them with uninformative hyper-priors and marginalize them out. We formulated the posterior distribution over deformations, with the log-normal hyper-parameters marginalized over, in a variational free-energy framework where the variational density is the Laplace approximation of the hyper-parameter posterior. A log-normal hyper-distribution was chosen because it provides very broad hyper-priors (with variation over orders of magnitude), reflecting the uncertainty in the exact values of the hyper-parameters. The hyper-prior ensures that while tissue- stiffness and image-noise variance parameters are allowed to vary over a very wide range of values, they are kept physically sensible by disallowing fantastically huge or small values. However, ideally, they should be centered around some sensible values derived for instance through hyper-parameter optimization using homologous landmarks/structures. The posterior estimation of the prior temperature can be seen as an estimate of the homogeneous inverse stiffness of the underlying tissue. In the future we plan to extend the current methodology to marginalize/estimate compartment- or element-wise tissue stiffness.

Our experiments have shown that marginal posterior distributions on deformations may be multi-modal – which does not imply that the complete posterior distribution is necessarily multimodal. However, because we have shown that the marginal distributions are non-Gaussian, and possibly multi-modal, we can claim that the posterior distribution is non-Gaussian. While it may be surprising to some that the marginal distributions can have multiple modes, our experience in image registration has been that registration systems can become trapped in local extrema in the face of truncation or resection. Likelihood functions

are complicated and highly non-convex functions of the transformation parameters; when there is structural disagreement among images, the likelihood functions frequently exhibit multiple modes that reflect spurious matches of image structures. Because of this, multiple modes in posterior marginals, or even the complete posterior distribution, are a possibility. Furthermore, increasing the degrees of freedom of the warp, which is similar to reducing the smoothing in the objective function, seems unlikely to reduce the occurrence of multiple modes in the posterior marginal; frequently it happens that increasing the degrees of freedom leads to more extrema in optimization problems.

While the proposed MCMC method is a principled approach to Bayesian inference, and provides a numerical approximation to the exact posterior distribution, it does come with high computational cost which is not unreasonable in an exploratory investigation. Although the computational cost may currently be too high for use in neurosurgery, it is straightforward to apply the method to new applications; e.g. to monitor delivered dose in adaptive radiotherapy where each fraction is delivered with approximately 24 hour intervals. For other applications, the proposed methodology can provide valuable insights into complex posterior distributions to help guide the design of faster methods that find simple robust and accurate approximations to the complex posterior distribution. One example is the variational Bayes approach taken by Simpson et al. (2012) which provides a fast exact analytical solution to a parametric approximation to the posterior distribution that facilitates high-dimensional registration. However, the effect of the approximation on the estimated posterior statistics is unknown and may be significant because of the possibly multi-modal and non-Gaussian posterior distribution. Parametric approximations of the posterior should therefore be compared against the asymptotically correct posterior distribution that can be generated by MCMC.

Another MCMC approach to registration, the Gibbs sampler proposed by (Gee et al., 1995b,a), requires approximations of the likelihood function. By using MH, such likelihood approximations are not required and the proposed method is consequently more general in terms of the dissimilarity energies that can readily be used. On the other hand, Gibbs sampling should in theory converge quicker on the posterior distribution because every sample is accepted. The proposed marginalization over hyper-parameters using local Laplace approximations can readily be incorporated with the Gibbs sampler proposed by Gee et al. (1995b,a).

The sampling approach we have used here is computationally intensive. Consequently, for the experiments in this article, we have used meshes with approximately 1000–2500 degrees of freedom, which corresponds to a vertex spacing of approximately 15mm in the clinical case; this is comparable to control point spacing typically used in intra-subject registration (Rueckert et al., 1999). Our vertex spacing is larger than the 5mm B-spline spacing that is commonly used in inter-subject registration (Klein et al., 2009), and it is possible that with higher resolution meshes and data, the estimated posterior uncertainty would decrease. With higher degrees of freedom comes not only higher computational cost, but also higher storage requirements because of the need to store enough samples to characterize the posterior. It is common practice to thin the MH sequence by a factor of at least ten to reduce intra-chain correlation and to discard the first half of the sample chain to remove burn-in effects. Consequently, if 10^7 samples are required to characterize a posterior distribution with 100k degrees of freedom, the storage requirements will be approximately 200Gb after thinning and burn-in. This does not pose a significant problem in practice, especially if advanced storage formats such as HDF5 (The HDF Group, 2000–2010), which can sequentially compress large files, are used.

Because of the high computational complexity of MCMC, some simplifying model assumptions were made to achieve tolerable running times. First, the likelihood assumes that the intensity difference between registered voxels differ by i.i.d. Gaussian noise. Theoretically, this is not a valid assumption, especially when Gaussian blurring is applied, but has been shown to work well in practice. In our work, where light blurring was used, we have expanded on the Gaussian noise model by marginalizing over the noise precision under a log-normal hyper-prior, which effectively can convert the Gaussian noise distribution into a distribution with heavy tails which is better suited for handling outliers, e.g. areas of resection. A well founded approach for handling the noise correlation caused by smoothing of the images is to include a constant Virtual Decimation (VD) factor in the likelihood term (Simpson et al., 2012; Groves et al., 2011), which is similar to decimating the data. This is especially important if the main goal is to infer correct noise and regularization (stiffness) parameters. However, in our case, where the main goal is to characterize the posterior distribution on deformations, the constant VD factor will be absorbed into the marginalized noise temperature and will not have an affect in practice, aside from changing the scale of the temperature parameter. Although the SSD-based likelihood does not currently accommodate for functional or statistical intensity differences, such complex relationships can easily be included in the model, though likely with some computational cost.

Secondly, the FE-based transformation model was a natural choice given the use of the phenomenologically inspired linear elastic prior, but other popular transformation models such as B-splines (Rueckert et al., 1999) can also be incorporated.

Thirdly, linear elasticity has been widely used as a prior in non-rigid registration, but more complex prior models such as visco- and poro-elasticity (Kyriacou et al., 2002) may be more appropriate. Choosing a suitable prior that is a good model for the underlying tissue under deformation may improve the registration results and reduce registration uncertainty. If a segmentation of the underlying tissue is available, and relative stiffness between tissue types is known, a mechanical model that reflects this can easily be constructed and spatially varying (pseudo-)stiffness parameters can easily be encoded in the model without any modifications to the proposed posterior sampling scheme. It is also possible to use a random variable for each tissue type, and equip it with a hyper-prior and marginalize over it, but this requires changing the Laplace approximation accordingly, or using MH of the per tissue type hyper-parameters.

We have used common modeling assumptions, and some approximations, and our goal is to understand the implications of those assumptions on the uncertainty in the posterior distribution (or in deterministic terms, the landscape of the objective function). In the future, it is likely that developments in likelihood modeling, prior modeling, FEM technology, inference algorithms, and application specific development and engineering will lead to faster methods that could have tighter posterior distributions and associated marginals than those that are illustrated in Fig. 13(c). Our experiments indicate that, depending on the application, it may be appropriate to consider the implications of non-Gaussian, and potentially multi-modal, posterior distributions in the design of the system. Approaches that should be investigated to speed up chain convergence, reduce inter-sample correlations and the random walk behavior of MH include Hybrid Monte Carlo (Duane et al., 1987), which uses gradients of the log posterior distribution to take larger proposal steps while keeping the rejection rate low, and adaptive MCMC where the proposal distribution is dynamically fitted to match the posterior based on the accepted samples (Andrieu and Thoms, 2008). One approach for learning the best compromise between likelihood model, prior model, mesh size, accuracy and uncertainty is through model selection using e.g. the Deviance Information Criterion (DIC) which is a suitable model selection criterion when the posterior is obtained by MCMC (Spiegelhalter et al., 2002).

In future work we will adapt the framework to registration of MRI and ultrasound where we suspect the posterior distribution will be multi-modal because of the high noise level and lack of contrast in ultrasound images. We also suspect that exploiting the high-dimensional information available in the posterior distribution on deformations when doing high-level inferences based on the registered data will open up new and exciting research opportunities. Simpson et al. (2011a) showed that utilizing the registration uncertainty in classification of Alzheimer patients improved their classification results. In adaptive radiotherapy, monitoring the radiation dose delivered to critical tissue is important due to changes in anatomy over the treatment period. In Risholm et al. (2011) we showed that the uncertainty in registration can induce significant uncertainty in the estimated cumulative dose delivered to critical tissue during head and neck radiotherapy. The estimated delivered dose and the associated uncertainty was visualized in terms of dose volume histograms. Finding low-dimensional representations of the posterior distribution that are suitable for visualization is a difficult task and an active research topic (Spiegelhalter et al., 2011). In this paper we used simple posterior summaries to convey uncertainty. However, one can envision that more complex approaches to interacting and extracting clinical relevant summaries from the posterior distribution can be designed in the future – for instance by exploiting the local correlations in the posterior.

5. Conclusion

We presented a Bayesian non-rigid registration framework where the posterior distribution is characterized through MCMC sampling. In contrast to conventional registration approaches that only report posterior modes, the posterior distribution, represented by a set of MCMC samples, contains other valuable information such as the registration uncertainty; this is important information when higher level inferences are drawn based on the data being registered.

Acknowledgments

This work was funded in parts by NIH grants R01CA138419, P41RR013218, P41EB015902, P41EB015898 and P41RR019703.

References

- Andrieu C, Thoms J. A tutorial on adaptive MCMC. *Statistics and Computing*. 2008; 18:343–373.
- Bansal R, Staib LH, Laine AF, Xu D, Liu J, Posecion LF, Peterson BS. Calculation of the confidence intervals for transformation parameters in the registration of medical images. *Med Image Anal*. 2009; 13:215–223. [PubMed: 19138877]
- Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn Reson Med*. 2003; 50:1077–88. [PubMed: 14587019]
- Bishop, CM. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc; Secaucus, NJ, USA: 2006.
- Bookstein F. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Mach Intell*. 1989; 11:567–585.
- Bro-nielsen, M. *Finite Element Modeling in Surgery Simulation*. Proc IEEE; 1998. p. 490-503.
- CGAL. *Computational Geometry Algorithms Library*; 2008. www.cgal.org
- Chang, J.; Fisher, JW. Efficient MCMC sampling with implicit shape representations. *Proc IEEE Comp Vis Pattern Recognit*; 2011. p. 2081-88.
- Cheng Y. Mean shift, mode seeking, and clustering. *IEEE Trans Pattern Anal Mach Intell*. 1995; 17:790–799.
- Crum WR, Hartkens T, Hill DLG. Non-rigid image registration: theory and practice. *Br J Radiol*. 2004; 77:S140–153. [PubMed: 15677356]

- Danilchenko A, Fitzpatrick JM. General approach to first-order error prediction in rigid point registration. *IEEE Trans Med Imaging*. 2011; 30:679–693. [PubMed: 21075718]
- Davis, PJ.; Rabinowitz, P. *Methods of Numerical Integration*. 2. Dover Publications; 2007.
- Duane S, Kennedy A, Pendleton BJ, Roweth D. Hybrid Monte Carlo. *Physics Letters B*. 1987; 195:216–222.
- Fan, AC.; Fisher, JW.; Wells, WM.; Levitt, JJ.; Willsky, AS. MCMC curve sampling for image segmentation. In: Ayache, N.; Ourselin, S.; Maeder, AJ., editors. *Med Image Comput Comput Assist Interv*. Springer; 2007. p. 477-485.
- Fieten LJ, Radermacher K, Kernbach MA, Heger S. Integration of model-based weighting into an ICP variant to account for measurement errors in intra-operative A-mode ultrasound-based registration. *Conf Proc IEEE Eng Med Biol Soc*. 2010; 1:6264–7. [PubMed: 21097352]
- Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the Laplace approximation. *Neuroimage*. 2007; 34:220–234. [PubMed: 17055746]
- Gee, JC.; Bajcsy, RK. *Brain Warping*. Academic Press; San Diego: 1999. Chapter 11 - elastic matching: Continuum mechanical and probabilistic analysis; p. 183-197.
- Gee, JC.; Haynor, DR.; Reivich, M.; Bajcsy, R. Finite element approach to warping of brain images. *Proc SPIE Conf Med Imaging: Im Process*; 1994. p. 18-27.
- Gee, JC.; Le Briquer, L.; Barillot, C.; Haynor, DR. Probabilistic matching of brain images. In: Bizais, Y.; Barillot, CRD., editors. *Inf Process Med Imaging*. Kluwer Academic Publishers; 1995a. p. 113-126.
- Gee, JC.; Le Briquer, L.; Barillot, C.; Haynor, DR.; Bajcsy, R. Bayesian approach to the brain image matching problem. *Proc SPIE Conf Med Imaging: Im Process*; 1995b. p. 154-156.
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian Data Analysis, Second Edition (Texts in Statistical Science)*. 2. Chapman & Hall/CRC; 2003.
- Geweke, J. Staff Report. Vol. 148. Federal Reserve Bank of Minneapolis; 1991. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.
- Groves AR, Beckmann CF, Smith SM, Woolrich MW. Linked independent component analysis for multimodal data fusion. *NeuroImage*. 2011; 54:2198–2217. [PubMed: 20932919]
- Gumprecht HK, Widenka DC, Lumenta CB. Brainlab vectorvision neuronavigation system: Technology and clinical experiences in 131 cases. *Neurosurgery*. 1999; 44:97–104. [PubMed: 9894969]
- Hub M, Kessler M, Karger C. A stochastic approach to estimate the uncertainty involved in B-spline image registration. *IEEE Trans Med Imaging*. 2009; 28:1708–1716. [PubMed: 19447703]
- Janoos, F.; Risholm, P.; Wells, WM. Robust non-rigid registration and characterization of uncertainty. In: Zhou, K.; Duncan, JS.; Ourselin, S., editors. *IEEE Workshop Math Methods Biomed Image Analysis*; 2012. p. 4321-4326.
- Johnson, SG. The NLOpt nonlinear optimization package. 2012. <http://abinitio.mit.edu/nlopt>
- Jones DK. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magn Reson Med*. 2003; 49:7–12. [PubMed: 12509814]
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*. 2009; 46:786–802. [PubMed: 19195496]
- Kybic, J. Fast no ground truth image registration accuracy evaluation: Comparison of bootstrap and Hessian approaches. *Proc IEEE Int Symp Biomed Imaging, IEEE*; 2008. p. 792-795.
- Kybic J. Bootstrap resampling for image registration uncertainty estimation. *IEEE Trans Image Process*. 2010; 19:64–73. [PubMed: 19709978]
- Kyriacou SK, Mohamed A, Miller K, Neff S. Brain mechanics for neurosurgery: modeling issues. *Biomech Model Mechan*. 2002; 1:151–164.
- Maier-Hein L, Franz A, dos Santos T, Schmidt M, Fangerau M, Meinzer H, Fitzpatrick J. Convergent iterative closest-point algorithm to accommodate anisotropic and inhomogeneous localization error. *IEEE Trans Pattern Anal Mach Intell*. 2012; 34:1520–1532. [PubMed: 22184256]

- Mardia KV. Measures of multivariate skewness and kurtosis with applications. *Biometrika*. 1970; 57:519–530.
- Maurer CR, Fitzpatrick JM, Wang MY, Galloway RL, Maciunas RJ, Allen GS. Registration of head volume images using implantable fiducial markers. *IEEE Trans Med Imaging*. 1997; 16:447–462. [PubMed: 9263002]
- McCormick T, Dink D, Orcun S, Pekny J, Rardin R, Baxter L, Thai V, Langer M. Target volume uncertainty and a method to visualize its effect on the target dose prescription. *Int J Radiat Oncol Biol Phys*. 2004; 60:1580–1588. [PubMed: 15590190]
- McNab JA, Jbabdi S, Deoni SC, Douaud G, Behrens TE, Miller KL. High resolution diffusion-weighted imaging in fixed human brain using diffusion-weighted steady state free precession. *NeuroImage*. 2009; 46:775–785. [PubMed: 19344686]
- Meurant G. Estimates of the trace of the inverse of a symmetric matrix using the modified Chebyshev algorithm. *Numerical Algorithms*. 2009; 51:309–318.
- Miller K, Wittek A, Joldes G. Biomechanics of the brain for computer-integrated surgery. *Acta Bioeng Biomech*. 2010; 12:25–37. [PubMed: 20882939]
- Papademetris, X.; Onat, ET.; Sinusas, AJ.; Dione, DP.; Constable, RT.; Duncan, JS. *Inf Process Med Imaging*. Springer-Verlag; London, UK: 2001. The active elastic model; p. 36-49.
- de Pasquale F, Gratta CD, Romani G. Empirical Markov chain Monte Carlo bayesian analysis of fMRI data. *NeuroImage*. 2008; 42:99–111. [PubMed: 18538586]
- Pennec X, Thirion JP. A framework for uncertainty and validation of 3-D registration methods based on points and frames. *Int J Comput Vis*. 1997; 25:203–229.
- Rinnooy Kan A, Timmer G. Stochastic global optimization methods part II: Multi level methods. *Mathematical Programming*. 1987; 39:57–78.
- Risholm, P.; Balter, J.; Wells, WM. Estimating delivered dose in radiation therapy: The influence of registration uncertainty. In: Fichtinger, G.; Martel, A.; Peters, T., editors. *Med Image Comput Comput Assist Interv*. Vol. 6891. Springer; 2011. p. 548-555.
- Risholm, P.; Janoos, F.; Fedorov, A.; Tempany, C.; Cormack, RA.; Wells, WM. Selection of optimal hyper-parameters for estimation of uncertainty in MRI-TRUS registration of the prostate. In: Ayache, N.; Delingette, H.; Golland, P.; Mori, K., editors. *Med Image Comput Comput Assist Interv*. Springer; 2012.
- Risholm, P.; Pieper, S.; Samset, E.; Wells, WM. Summarizing and visualizing uncertainty in non-rigid registration. In: Jiang, T.; Navab, N.; Pluim, J.; Viergever, M., editors. *Med Image Comput Comput Assist Interv*. Vol. 6362. Springer; 2010. p. 554-561.
- Robinson MD, Milanfar P. Fundamental performance limits in image registration. *IEEE Trans Image Process*. 2004; 13:1185–1199. [PubMed: 15449581]
- Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Trans Med Imaging*. 1999; 18:712–721. [PubMed: 10534053]
- Seginer A. Rigid-body point-based registration: The distribution of the target registration error when the fiducial registration errors are given. *Med Image Anal*. 2011; 15:397–413. [PubMed: 21515088]
- Shamir R, Joskowicz L, Spektor S, Shoshan Y. Localization and registration accuracy in image guided neurosurgery: a clinical study. *Int J Comput Assist Radiol Surg*. 2009; 4:45–52. [PubMed: 20033601]
- Simpson, AL.; Ma, B.; Chen, ECS.; Ellis, RE.; Stewart, AJ. Using registration uncertainty visualization in a user study of a simple surgical task. In: Larsen, R.; Nielsen, M.; Sporring, J., editors. *Med Image Comput Comput Assist Interv*. Springer; 2006. p. 397-404.
- Simpson II, Schnabel JA, Groves AR, Andersson JL, Woolrich MW. Probabilistic inference of regularisation in non-rigid registration. *NeuroImage*. 2012; 59:2438–51. [PubMed: 21939772]
- Simpson, II.; Woolrich, MW.; Groves, AR.; Schnabel, JA. Longitudinal brain MRI analysis with uncertain registration. In: Fichtinger, G.; Martel, AL.; Peters, TM., editors. *Med Image Comput Comput Assist Interv*. Springer; 2011a. p. 647-654.
- Simpson, II.; Woolrich, MW.; Schnabel, JA. Probabilistic segmentation propagation from uncertainty in registration. *Proc Medical Image Understanding and Analysis*; 2011b. p. 331-335.

- Smal I, Carranza-Herrezuelo N, Klein S, Wielopolski P, Moelker A, Springeling T, Bernsen M, Niessen W, Meijering E. Reversible jump MCMC methods for fully automatic motion analysis in tagged MRI. *Med Image Anal.* 2012; 16:301–324. [PubMed: 21963294]
- Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp.* 2002; 17:143–155. [PubMed: 12391568]
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian Measures of Model Complexity and Fit (with Discussion). *J R Stat Soc Series B Stat Methodol.* 2002; 64:583–616.
- Spiegelhalter DJ, Pearson M, Short I. Visualizing uncertainty about the future. *Science.* 2011; 333:1393–1400. [PubMed: 21903802]
- Svanberg K. A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM J Optimization.* 2002; 12:555–573.
- The HDF Group. Hierarchical data format version 5. 2000–2010. <http://www.hdfgroup.org/HDF5>
- Woolrich MW, Chiarelli P, Gallichan D, Perthen J, Liu TT. Bayesian inference of hemodynamic changes in functional arterial spin labeling data. *Magn Reson Med.* 2006; 56:891–906. [PubMed: 16964610]
- Yeo BTT, Sabuncu MR, Vercauteren T, Holt DJ, Amunts K, Zilles K, Golland P, Fischl B. Learning task-optimal registration cost functions for localizing cytoarchitecture and function in the cerebral cortex. *IEEE Trans Med Imaging.* 2010; 29:1424–1441. [PubMed: 20529736]
- Zienkiewicz, O.; Taylor, R. *The Finite Element Method.* 5. Butterworth-Heinemann; 2000.

Appendix A. Linear Finite Elements

In the Finite Element (FE) setting (Zienkiewicz and Taylor, 2000; Bronielsen, 1998) with tetrahedral FEs, Ω_f is discretized into disjoint tetrahedral elements connected at the set of mesh vertices $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_V\}$ where \mathbf{v}_j has coordinates $(x_j, y_j, z_j)^\top$ in the fixed image. Associated with each vertex \mathbf{v}_j is a displacement vector $\mathbf{u}_j = (\delta x_j, \delta y_j, \delta z_j)^\top$. The displacement vector at any voxel $\mathbf{x} \in \Omega_f$ can be determined using linear interpolation of the displacements at the vertices of its enclosing tetrahedral element $\varepsilon = \{\mathbf{v}_{e_1} \dots \mathbf{v}_{e_4}\}$:

$\mathbf{u}(\mathbf{x}) = \sum_{j=1}^4 b_x(\mathbf{v}_{e_j}) \mathbf{u}_{e_j}$, $\forall \mathbf{x} \in \text{ConvexHull}\{\varepsilon\}$, where $b_x(\mathbf{v}_{e_j})$ is the barycentric weight of vertex \mathbf{v}_{e_j} at voxel \mathbf{x} . In matrix notation, this is represented by $\mathbf{u}(\mathbf{x}) = \mathbf{B}_x \mathbf{u}$, where \mathbf{B}_x is the $3 \times V$ interpolation coefficient matrix for voxel \mathbf{x} (with 4 non-zero 3×3 diagonal blocks) and $\mathbf{u} = (\mathbf{u}_1^\top, \mathbf{u}_2^\top \dots \mathbf{u}_V^\top)^\top$ is the $3V \times 1$ vector of displacements. Therefore, for each $\mathbf{x} \in \Omega_f$ under a given deformation of the FE mesh \mathbf{u} , the corresponding voxel in the moving image has location $\mathbf{x} + \mathbf{u}(\mathbf{x}) = \mathbf{x} + \mathbf{B}_x \mathbf{u}$.

Although any alternative measure can be used in this framework, for ease of presentation the sum-of-square differences (SSD) is used in the following:

$$E_{ssd}(\mathbf{u}; f, m) = \sum_{\mathbf{x} \in \Omega_f} (f(\mathbf{x}) - m(\mathbf{x} + \mathbf{B}_x \mathbf{u}))^2 \quad (\text{A.1})$$

$$= \sum_{e \in \mathcal{E}} \sum_{\mathbf{x} \in \Omega_e} (f(\mathbf{x}) - m(\mathbf{x} + \mathbf{B}_x \mathbf{u}))^2. \quad (\text{A.2})$$

By pre-computing $\mathbf{x} \in \Omega_e$, this similarity measure is highly parallelizable.

We assume that the underlying tissue which is discretized by FEs can be represented by an isotropic elastic material parameterized by stiffness (Young's modulus) and compressibility (Poisson's ratio). The elastic energy stored in a deformed object is equal to the sum of the stress times strain over the FEs. By assuming a linearized strain tensor, the linear elastic

energy of a deformation, which we use as a regularizing term to penalize physically implausible solutions, takes the form $E_e(\mathbf{u}) = \int_{\mathbf{x} \in \Omega_f} \mathbf{u}(\mathbf{x})^\top \mathbf{K}_x \mathbf{u}(\mathbf{x}) d\mathbf{x}$, where \mathbf{K}_x is the stiffness matrix that is directly proportional to the Young's modulus. Under the linear element model, this is equivalent to

$$E_{el}(\mathbf{u}) = \sum_{i=1}^V \sum_{j=1}^V \mathbf{u}_i^\top \mathbf{K}_{i,j} \mathbf{u}_j = \mathbf{u}^\top \mathbf{K} \mathbf{u}, \quad (\text{A.3})$$

where $\mathbf{K}_{i,j}$ is non-zero only if vertices \mathbf{v}_i and \mathbf{v}_j share a tetrahedral element.

Consequently, both the SSD and elastic energy contributions of one vertex \mathbf{v}_i under displacement \mathbf{u}_i depends only on the displacements of vertices \mathbf{v}_j that share tetrahedral elements with \mathbf{v}_i . Therefore, the energy term exhibits Markov structure and the Markov blanket for vertex \mathbf{v}_i is $\mathcal{L}_i = \cup_e \mathcal{E}_e \setminus \mathbf{v}_i$ such that $\mathbf{v}_j \in \mathcal{E}_e$.

Even though linear elasticity puts a penalty on non-smooth deformations, it does not prevent unwanted folding of tissue/elements. Folding of elements can be prevented by restricting the determinant of the Jacobian over the deformed element to be positive $|\mathcal{J}(\mathbf{u}_e(\Omega_e))| > 0$.

Appendix B. Marginalization through Laplace Approximations

To simplify the following derivations, we define $\theta_r \triangleq \ln \tau_r$ and $\theta_s \triangleq \ln \tau_s$, so that the log-normal priors over \mathcal{T} imply normal priors over $\Theta \triangleq \{\theta_r, \theta_s\}$. Introducing a variational density $q(\Theta)$ over the HPs Θ , the marginal distribution of the deformations \mathbf{u} in Eqn. (4) can be decomposed into a free-energy and KL-divergence term as:

$$\ln p(\mathbf{u}|m, f) = \ln \int p(\mathbf{u}, \Theta|m, f) d\Theta = F + \mathbb{KL}(q \| p(\Theta|\mathbf{u}, m, f)), \quad (\text{B.1})$$

where

$$F = \int q(\Theta) \ln p(\mathbf{u}, \Theta|m, f) d\Theta - \int q(\Theta) \ln q(\Theta) d\Theta. \quad (\text{B.2})$$

As the KL divergence term is positive, F acts as a lower bound on the estimate of $\ln p(\mathbf{u} | m, f)$ and is exact for $q = p(\Theta | \mathbf{u}, m, f)$.

Restricting the variational density $q(\Theta)$ to the family of normal distributions, the minimizer of $\mathbb{KL}(q \| p(\Theta | \mathbf{u}, m, f))$ is $\mathcal{N}(\mathbb{E}[p(\Theta | \mathbf{u}, m, f)], \text{Var}[p(\Theta | \mathbf{u}, m, f)])$. However, as the mean and variance of the posterior do not have closed-form solutions, and to avoid expensive Monte Carlo integration, we use a Laplace approximation of $p(\Theta | \mathbf{u}, m, f)$ by replacing the mean with the mode and the variance with the local curvature of the mode. Consequently, $q(\Theta) \sim \mathcal{N}(\Theta^*, \Lambda^{*-1})$, where $\Theta^* = \text{argmax}_{\Theta} \ln p(\Theta | \mathbf{u}, m, f)$ is the conditional mode and $\Lambda^* = -2\nabla\nabla^\top \ln p(\Theta | \mathbf{u}, m, f)$ is the negative Hessian of $\ln p(\Theta | \mathbf{u}, m, f)$ respectively, which are equivalent to the mode and negative Hessian of $\ln p(\mathbf{u}, \Theta | m, f)$. The maximum of $\ln p(\mathbf{u} = \mathbf{u}', \Theta | m, f)$ with respect to Θ (see Eqn. (4)) can be computed with a Newton-like method using its analytical gradient:

$$\begin{aligned} \frac{\partial \ln p}{\partial \theta_r} &= -\frac{1}{2} \left[-e^{-\theta_r} E_r + 2\lambda_r \theta_r - 2\lambda_r \mu_r \right] - \frac{\partial \ln Z}{\partial \theta_r}, \\ \frac{\partial \ln p}{\partial \theta_s} &= -\frac{1}{2} \left[-e^{-\theta_s} E_s + 2\lambda_s \theta_s - 2\lambda_s \mu_s \right] - \frac{\partial \ln Z}{\partial \theta_s}, \end{aligned} \quad (\text{B.3})$$

and Hessian:

$$\begin{aligned}\frac{\partial^2 \ln p}{\partial \theta_r^2} &= -\frac{1}{2} \left[e^{-\theta_r} E_r + 2\lambda_r \right] - \frac{\partial^2 \ln Z}{\partial \theta_r^2}, \\ \frac{\partial^2 \ln p}{\partial \theta_s^2} &= -\frac{1}{2} \left[e^{-\theta_s} E_s + 2\lambda_s \right] - \frac{\partial^2 \ln Z}{\partial \theta_s^2}, \\ \frac{\partial^2 \ln p}{\partial \theta_r \partial \theta_s} &= -\frac{\partial^2 \ln Z}{\partial \theta_r \partial \theta_s}.\end{aligned}\quad (\text{B.4})$$

This maximization requires evaluating the log-partition function $\ln Z(\Theta)$ and its derivatives, which is described further in Appendix C.

Therefore, under the Laplace approximation $q \approx \mathcal{N}(\Theta^*, \Lambda^*)$ the two components of the free energy in Eqn. (B.2) take the form

$$\int_{\Theta} q(\Theta) \ln q(\Theta) d\Theta = \frac{1}{2} [-\ln |\Lambda^*| + 2 \ln |2\pi| + 2]$$

and

$$\begin{aligned}\int_{\Theta} q(\Theta) \ln p(\mathbf{u}, \Theta | m, f) d\Theta \\ &= -\frac{1}{2} \left[\mathbb{E}_q[e^{-\theta_r}] E_{cl} + \mathbb{E}_q[e^{-\theta_s}] E_{ssd} \right] \\ &\quad + \lambda_r \mathbb{E}_q[\theta_r^2] \\ &\quad - 2\lambda_r \mathbb{E}_q[\theta_r] \mu_r \\ &\quad + \lambda_s \mathbb{E}_q[\theta_s^2] \\ &\quad - 2\lambda_s \mathbb{E}_q[\theta_s] \mu_s \\ &\quad - \mathbb{E}_q[\ln Z(\Theta)] + c,\end{aligned}\quad (\text{B.5})$$

where $\mathbb{E}_q[e^{-\theta_r}] = \exp\left\{-\theta_r^* + \frac{\Lambda_{2,2}^*}{2|\Lambda^*|}\right\}$ and $\mathbb{E}_q[e^{-\theta_s}] = \exp\left\{-\theta_s^* + \frac{\Lambda_{1,1}^*}{2|\Lambda^*|}\right\}$

Appendix C. Estimating the Partition Function

Evaluating the partition function

$$Z(\Theta) = \int \exp\left\{-\frac{e^{-\theta_r}}{2} E_{cl}(\mathbf{u}) - \frac{e^{-\theta_s}}{2} E_{ssd}(\mathbf{u})\right\} d\mathbf{u} \quad (\text{C.1})$$

requires integration of the moving image as a function of the displacement vector over the space of all possible displacements and therefore does not have a closed form solution. Moreover, as it needs to be computed at every step during the optimization of $\arg\max_{\Theta} \ln p(\Theta | \mathbf{u}, m, f)$, numerical or Monte Carlo integration is not computationally feasible.

In order to make the estimation of $Z(\Theta)$ tractable, Laplace's method is again applied to replace $E(\mathbf{u}) = e^{-\theta_r} E_{cl}(\mathbf{u}) + e^{-\theta_s} E_{ssd}(\mathbf{u})$ with a quadratic expansion around the minimum $\mathbf{u}^* = \arg\min_{\mathbf{u}} E(\mathbf{u})$. Specifically, the Taylor series expansion is

$E(\mathbf{u}) = E_{cl}(\mathbf{u}) + E_{ssd}(\mathbf{u}^*) + (\mathbf{u} - \mathbf{u}^*)^T \nabla E_{ssd}(\mathbf{u}^*) + \frac{1}{2} (\mathbf{u} - \mathbf{u}^*)^T \nabla \nabla^T E_{ssd}(\mathbf{u}^*) (\mathbf{u} - \mathbf{u}^*) + \mathcal{O}(\Theta^3)$.
Substituting in Eqn. (C.1) gives:

$$\begin{aligned} Z(\Theta) &\cong \exp\left\{-\frac{e^{-\theta_s}}{2}a(\mathbf{u}^*)\right\} \int \exp\left\{-\frac{1}{2}[(\mathbf{u}-\mu_{\mathbf{u}|\Theta})^\top \Lambda_{\mathbf{u}|\Theta}(\mathbf{u}-\mu_{\mathbf{u}|\Theta})]\right\} d\mathbf{u}. \\ &= \exp\left\{-\frac{e^{-\theta_s}}{2}a(\mathbf{u}^*)\right\} \frac{\Lambda_{\mathbf{u}|\Theta}^{-\frac{1}{2}}}{2\pi}, \end{aligned} \quad (\text{C.2})$$

where $\Lambda_{\mathbf{u}|\Theta} \triangleq e^{-\theta_r} \mathbf{K} + \frac{e^{-\theta_s}}{2} \mathbf{D}$ and $\mu_{\mathbf{u}|\Theta} \triangleq \Lambda_{\mathbf{u}|\Theta}^{-1} \left[\frac{1}{2} e^{-\theta_s} (\nabla E_{ssd}(\mathbf{u}^*) + \mathbf{D} \mathbf{u}^*) \right]$. Here, the notation $\mathbf{D} \triangleq \nabla \nabla^\top E_{ssd}(\mathbf{u}^*)$ and $a(\mathbf{u}^*) \triangleq [E_{ssd}(\mathbf{u}^*) + \mathbf{u}^{*\top} \nabla E_{ssd}(\mathbf{u}^*) + \frac{1}{2} \mathbf{u}^* \mathbf{D} \mathbf{u}^*]$ is being used for conciseness¹.

Therefore,

$$\ln Z(\Theta) = -\frac{1}{2} \ln |\Lambda_{\mathbf{u}|\Theta}| - \frac{e^{-\theta_s}}{2} a(\mathbf{u}^*) + c = -\frac{1}{2} [-N\theta_s + \ln |D| + \ln |I + e^{\theta_s - \theta_r} b D^{-1} \mathbf{K}| + e^{-\theta_s} a(\mathbf{u}^*)] + c.$$

Typically in the applications of this method the value of $e^{\theta_s - \theta_r} \ll 1$, therefore using the fact that for small Θ it is the case that $|I + \Theta \mathbf{A}| = 1 + \tau_r(\mathbf{A})\Theta + \mathcal{O}(\Theta^2)$, the approximation:

$$\ln Z(\Theta) \cong -\frac{1}{2} \left[-N\theta_s + \ln |D| + \ln |1 + e^{\theta_s - \theta_r} \tau_r(\mathbf{D}^{-1} \mathbf{K})| + e^{-\theta_s} a(\mathbf{u}^*) \right] + c \quad (\text{C.3})$$

is highly accurate.

Introducing the term $\xi \triangleq \tau_r(\mathbf{D}^{-1} \mathbf{K}) e^{\theta_s - \theta_r} / (1 + e^{\theta_s - \theta_r} \tau_r(\mathbf{D}^{-1} \mathbf{K}))$, the gradient of the approximate log-partition function becomes:

$$\frac{\partial \ln Z(\Theta)}{\partial \theta_r} = \frac{\xi}{2}, \quad \frac{\partial \ln Z(\Theta)}{\partial \theta_s} = \frac{1}{2} \left[N - \xi + e^{-\theta_s} a(\mathbf{u}^*) \right], \quad (\text{C.4})$$

and its Hessian becomes:

$$\begin{aligned} \frac{\partial^2 \ln Z(\Theta)}{\partial \theta_r^2} &= -\frac{1}{2} [\xi - \xi^2], & \frac{\partial^2 \ln Z(\Theta)}{\partial \theta_r \partial \theta_s} &= -\frac{1}{2} [-\xi + \xi^2], \\ \frac{\partial^2 \ln Z(\Theta)}{\partial \theta_s^2} &= -\frac{1}{2} [\xi - \xi^2 + e^{-\theta_s} a(\mathbf{u}^*)]. \end{aligned} \quad (\text{C.5})$$

From eqns. (B.4) and (C.5), it can be seen that $\ln p(\Theta|\mathbf{u}, m, f)$ is concave in Θ , implying the existence of a global and unique maximum.

To avoid the expensive inversion of the $3N \times 3N$ matrix \mathbf{D} at each new value of deformation \mathbf{u}^* , as required in the evaluation of ξ , the term $\tau_r(\mathbf{D}^{-1} \mathbf{K})$ is directly estimated from $\mathbf{K}^{-1} \mathbf{D}$ using Gauss quadrature and the modified Chebyshev algorithm along with stochastic estimates of the modified moments (Meurant, 2009). In this calculation, pseudo-inverse \mathbf{K}^{-} (which can be pre-computed) is used since the translation invariance property of \mathbf{K} renders $\det \mathbf{K} = 0$. The expected value of $\ln Z(\Theta)$ under the variational distribution $q(\Theta) = \mathcal{N}(\Theta^*, \Lambda^{*-1})$ is $\mathbb{E}_q[\ln Z(\Theta)] = \frac{1}{2} [-N\Theta_s^* \ln |D| + \mathbb{E}_q[\ln |1 + e^{\theta_s - \theta_r} \tau_r(\mathbf{D}^{-1} \mathbf{K})|] + \mathbb{E}_q[e^{-\theta_s} a(\mathbf{u}^*)]]$, where $\mathbb{E}_q[\ln |1 + e^{\theta_s - \theta_r} \tau_r(\mathbf{D}^{-1} \mathbf{K})|]$ is evaluated numerically using Gauss-Hermite quadrature (Davis and Rabinowitz, 2007).

¹The gradient and Hessian of the similarity energy $E_{ssd}(\mathbf{u}) = \frac{1}{2} \sum_{\mathbf{x} \in \Omega_f} |m(\mathbf{x} + \mathbf{B}_x \mathbf{u}) - f(\mathbf{x})|^2$ with respect to \mathbf{u} can be expressed as transformations of the moving image gradient and Hessian as follows:

$$\begin{aligned} \nabla_{\mathbf{u}} E_{ssd}(\mathbf{u}) &= \sum_{\mathbf{x}} 2 [m(\mathbf{x} + \mathbf{B}_x \mathbf{u}) - f(\mathbf{x})] [\nabla_{\mathbf{x}} m(\mathbf{x} + \mathbf{B}_x \mathbf{u}) \mathbf{B}_x], \\ \nabla \nabla_{\mathbf{u}}^\top E_{ssd}(\mathbf{u}) &= \sum_{\mathbf{x}} \left\{ 2 [\nabla_{\mathbf{x}} m(\mathbf{x} + \mathbf{B}_x \mathbf{u}) \mathbf{B}_x] [\nabla_{\mathbf{x}} m(\mathbf{x} + \mathbf{B}_x \mathbf{u}) \mathbf{B}_x]^\top + 2 [m(\mathbf{x} + \mathbf{B}_x \mathbf{u}) - f(\mathbf{x})] [\mathbf{B}_x^\top \nabla \nabla_{\mathbf{x}}^\top m(\mathbf{x} + \mathbf{B}_x \mathbf{u}) \mathbf{B}_x] \right\}. \end{aligned}$$

Highlights

- Characterize non-parametric posterior distribution on deformations with MCMC.
- Marginalize over hyper-parameters.
- The registration uncertainty is estimated.
- The posterior distribution on deformations can be non-Gaussian and multimodal.

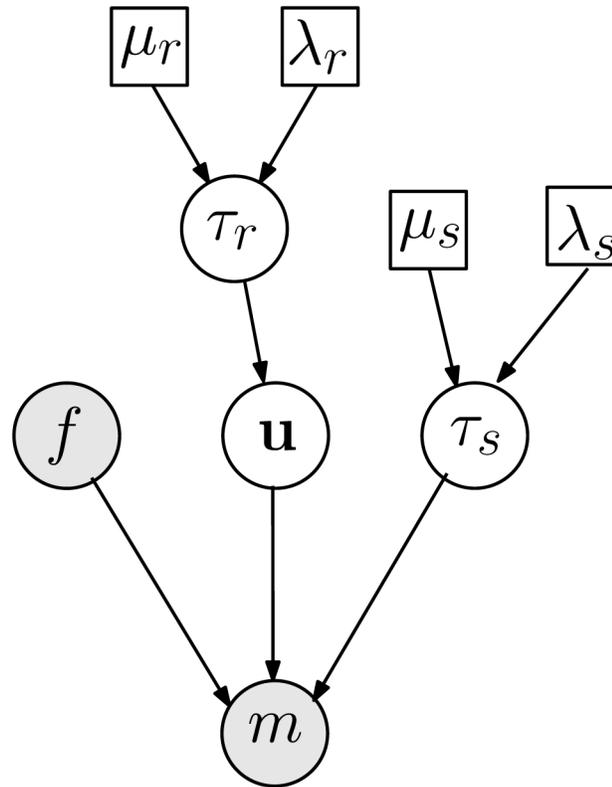


Figure 1.

Graphical model of the Bayesian image registration framework. Using a generative form, $m = f \circ \mathbf{u} + \varepsilon$, for the registration problem where the moving image m is a realization of a deformed fixed image f plus some noise ε . The hyper-parameter τ_r controls the spatial precision of the deformation field \mathbf{u} under a log-normal hyper-prior parametrized by the mean μ_r and precision λ_r , and is inversely proportional to tissue stiffness in the elastic regularization model used in this article. A log-normal hyper-prior, with mean μ_s and precision λ_s , models the uncertainty in the parameter τ_s which is proportional to the variance of noise introduced during formation of image m .

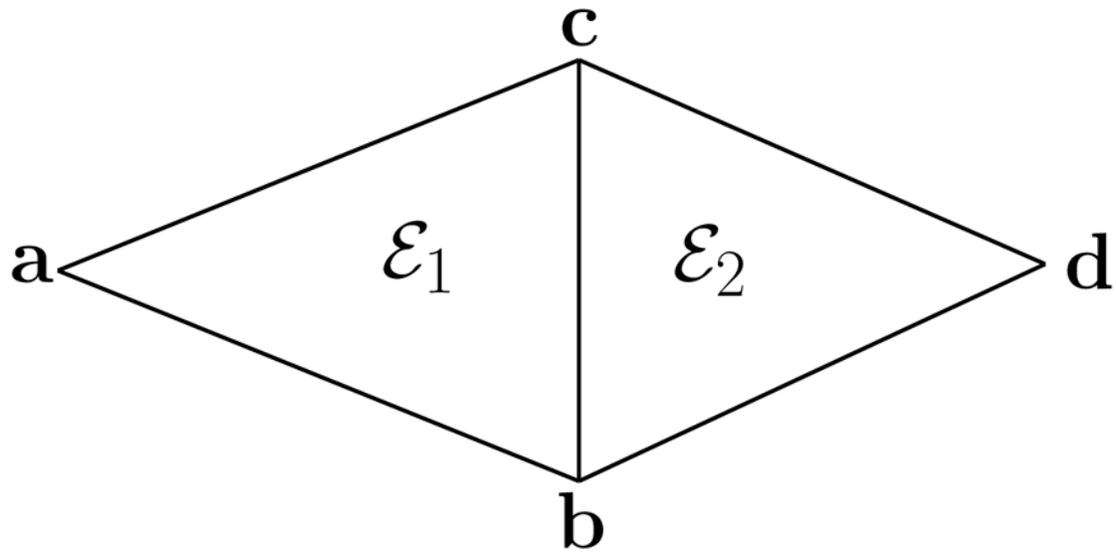


Figure 2.

A simple FE-mesh which consists of vertices **a**, **b**, **c** and **d** and two triangular elements $\varepsilon_1 = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ and $\varepsilon_2 = \{\mathbf{b}, \mathbf{d}, \mathbf{c}\}$. Assume \mathbf{u}_a , \mathbf{u}_b , \mathbf{u}_c and \mathbf{u}_d are displacements associated with the FE-vertices, then, due to the Markov structure of the FE-mesh, the high-dimensional joint distribution on deformations can be factorized into lower-dimensional marginal distributions $p(\mathbf{u}_a, \mathbf{u}_b, \mathbf{u}_c, \mathbf{u}_d) = p(\mathbf{u}_a, \mathbf{u}_b, \mathbf{u}_c) p(\mathbf{u}_b, \mathbf{u}_c, \mathbf{u}_d) / p(\mathbf{u}_b, \mathbf{u}_c)$ according to Eqn. (8).

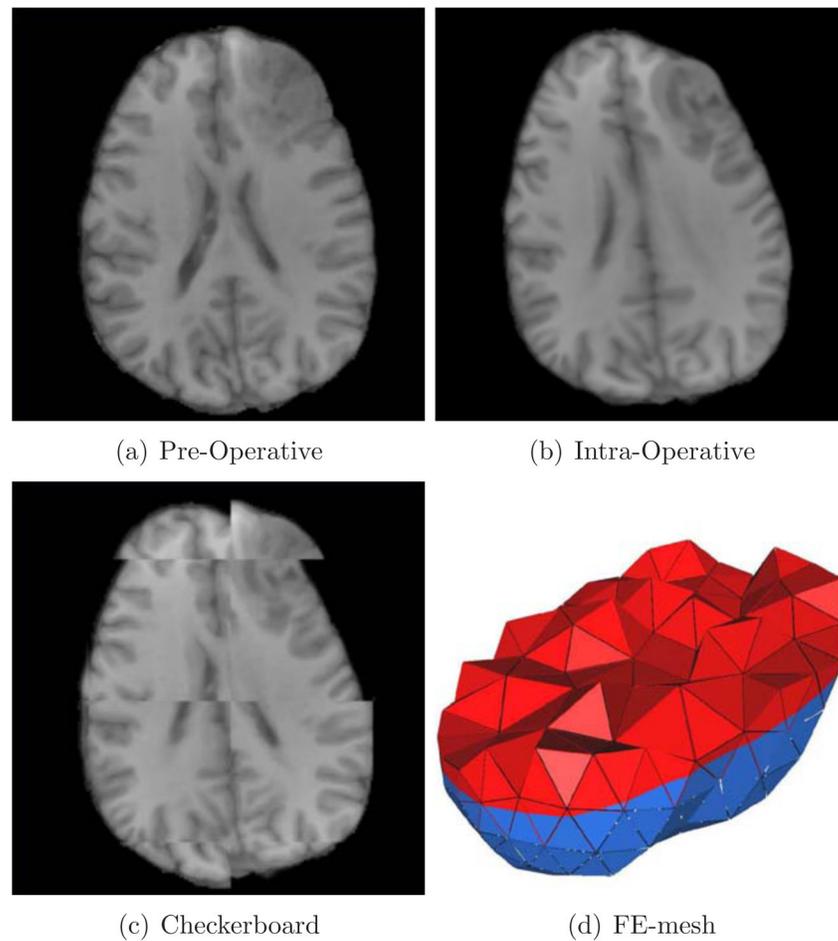


Figure 3.

Corresponding axial slices of the synthetic data used for the simulations. **(a)** A T1 MR-image ($256 \times 256 \times 88$) with $1.0 \times 1.0 \times 2.0 \text{mm}^3$ spacing acquired prior to neuro-surgery with intensities in the range $[0, 1]$. **(b)** Using a B-spline deformation model, a brain-shift of about 12mm was simulated together with global sinusoidal components with magnitude of 6mm and 4mm in the x - and y -components as explained in the text. The resulting ground-truth deformation field was applied to (a) to create the synthetic “intra-operative” image in (b). **(c)** A 4×4 checkerboard pattern of the images in (a) and (b). **(d)** A 3D view of the corresponding FE-mesh consisting of $V = 317$ vertices and $E = 1177$ tetrahedra. The blue color indicates that the element surface is on the boundary of the mesh, while red indicates that the surface is shared between two elements.

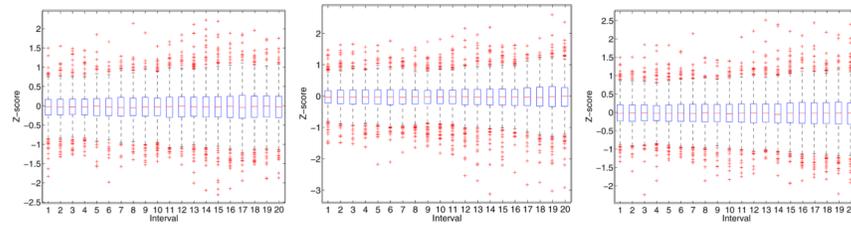
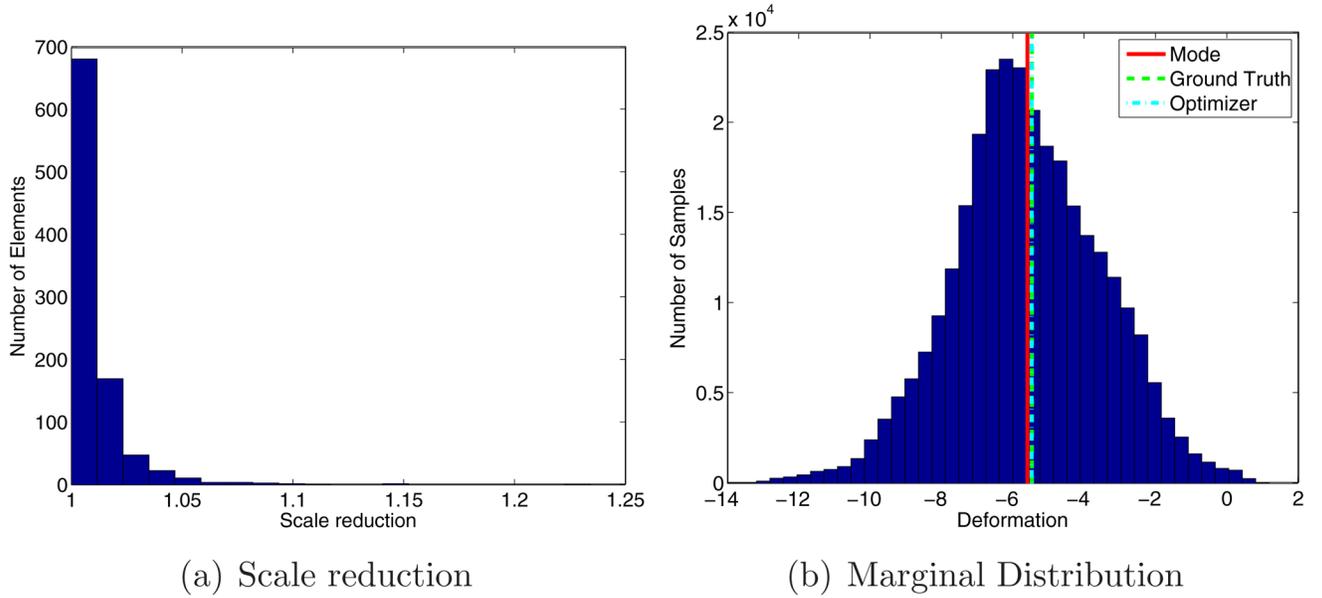


Figure 4.

Geweke time-series results for the 3 MCMC chains for the synthetic data in Fig. 3. The latter half of each chain after thinning ($1 \cdot 10^5$ samples) was divided into 20 subchains, and z -scores comparing the first half of a chain with each of the 20 subintervals were computed for each of the 951 deformation components. Each figure shows box-plots of the 951 z -scores for the 20 sub-intervals, and we can conclude that the chains have converged because the majority of the z -scores are within two standard errors of the mean.

**Figure 5.**

Scale reduction results from the experiment on the synthetic MRI data in Fig. 3. **(a)** Histogram of the potential scale reduction for the deformation components. The general guideline is that the scale reduction should be below ≈ 1.1 for each component before terminating the chains. In this case, four components exceed a scale reduction of 1.1, with the max scale reduction at 1.23. **(b)** Marginal distribution of the component with the largest scale reduction after pooling together the samples from the three individual MCMC chains. Even though the scale reduction criterion is violated, the marginal distribution is relatively well-characterized. Notice that the red line denotes the joint mode, which for high-dimensional non-Gaussian distributions do not necessarily align with the marginal mode. In this case it is highly concordant with the ground truth, but not with the marginal mode.

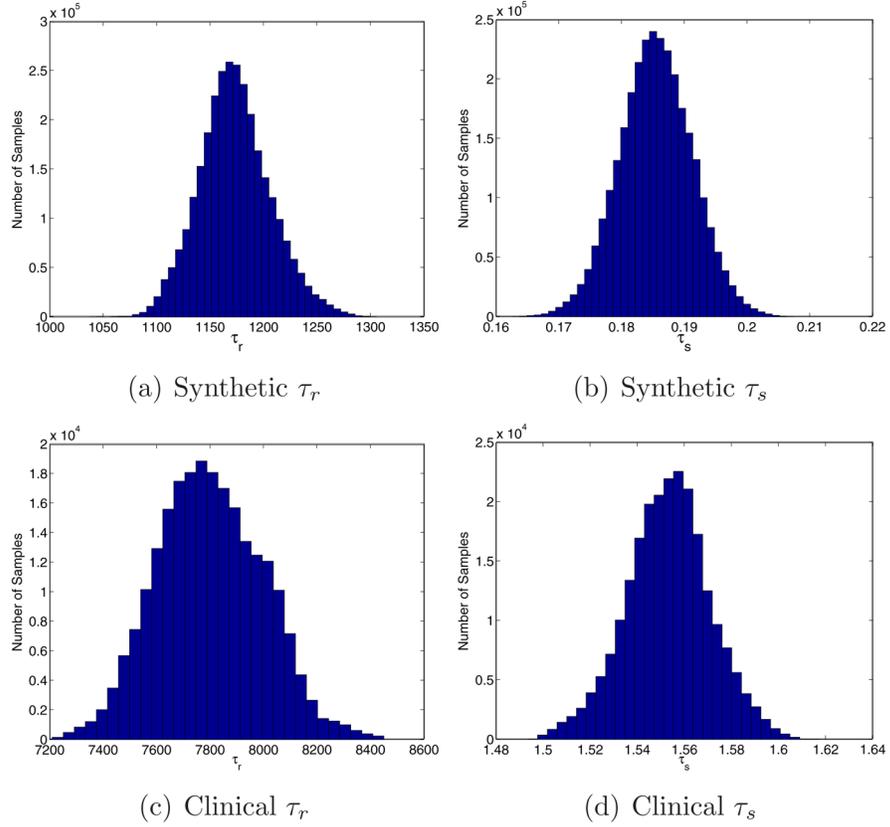
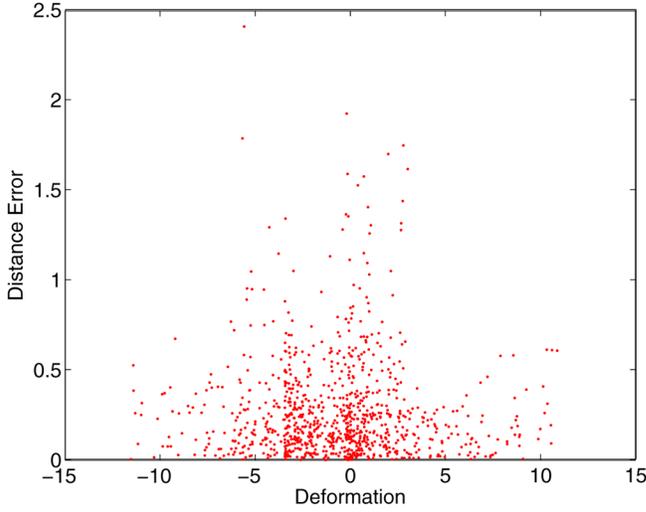
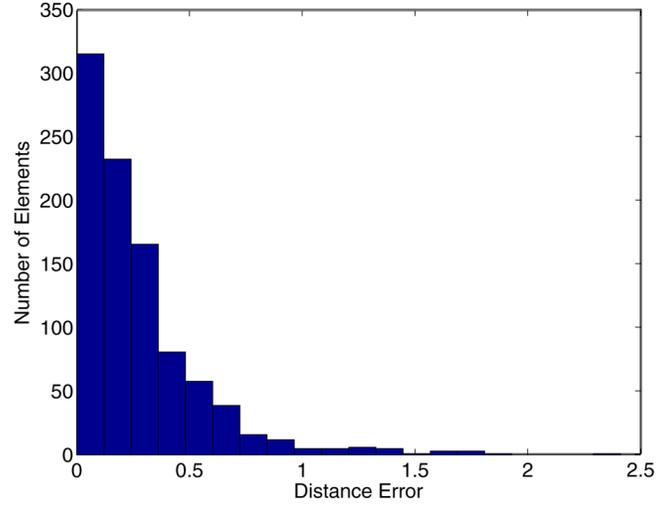


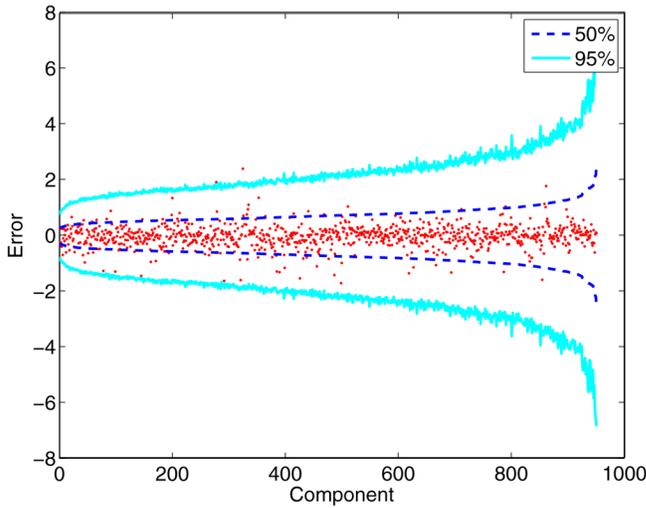
Figure 6. Histogram plots of the marginalized temperatures using Laplace approximations under the hyper-prior distributions with the method described in Section 2.2.3. **(a,b)** The distribution of marginalized temperatures from registering the synthetic dataset in Fig. 3. The mean and variances were $\hat{\tau}_r = 1174$, $\hat{\tau}_s = 0.19$ and $\hat{\sigma}_{\tau_r}^2 = 1117$, $\hat{\sigma}_{\tau_s}^2 = 3.4 \cdot 10^{-5}$. **(c,d)** The distribution of marginalized temperatures from registering the clinical dataset in Fig. 10. The mean and variances were $\hat{\tau}_r = 7528$, $\hat{\tau}_s = 1.59$ and $\hat{\sigma}_{\tau_r}^2 = 7.1e^5$, $\hat{\sigma}_{\tau_s}^2 = 0.04$.



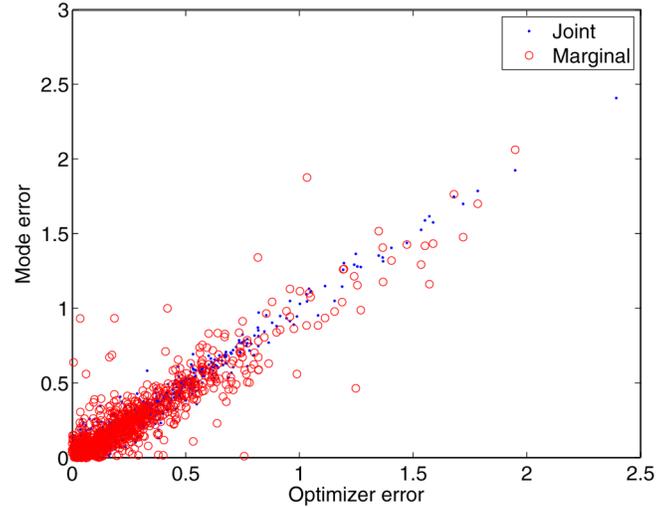
(a) Error Plot



(b) Error Histogram



(c) IQR vs Error

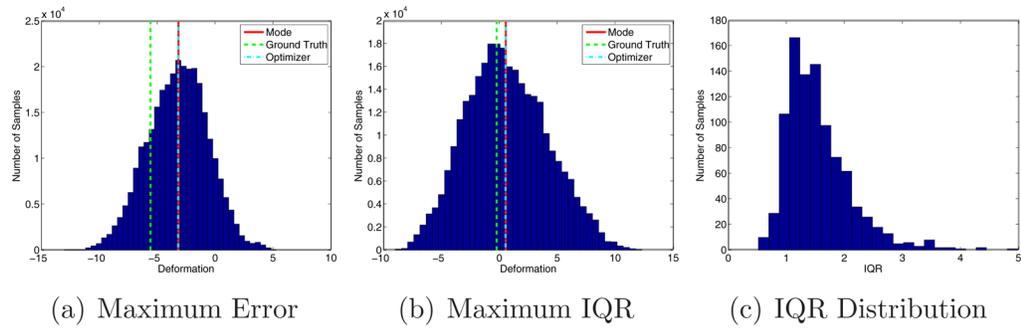


(d) Marginal vs Joint Error

Figure 7.

Registration error plots from the experiment on the synthetic MRI data in Fig. 3. **(a)** Error between the joint mode, estimated non-parametrically from the MCMC samples, and the ground-truth plotted against the ground-truth deformation value. A maximum absolute error of 2.4mm was found in the x -direction, while the mean error over all dimensions was 0.27mm. Notice that the highest errors are found for components with small deformations which may be explained by the regularizer which puts more restrictions on large deformations. **(b)** Histogram of the absolute distances between ground truth and the mode. **(c)** The distance between ground truth and the mode in relation to the spread of the 50% middle samples (IQR) and the spread of the 95% middle samples of the corresponding deformation component. The components are sorted in ascending order of IQR. **(d)** The marginal and joint error in the MCMC non-parametric mode (as compared to ground truth) plotted against corresponding error in optimizer estimates. It can be observed that there is a

high correlation between the non-parametric joint error and optimizer error, while the errors of the marginal estimates are less correlated with regards to the optimizer and joint estimates. In Section 3.1.4, the posterior is shown to not be normally distributed which explains the fact that the marginal modes do not correspond well with the joint mode.

**Figure 8.**

Histogram plots of the deformation components with (a) the highest error and (b) highest IQR from registering the synthetic MRI data in Fig. 3. The component in (a) has an absolute error of 2.4mm while the component in (b) has an IQR of 5.0mm. (c) The distribution of the 951 IQRs (one for each deformation component). The smallest and largest IQR were 0.52mm and 5.00mm respectively.

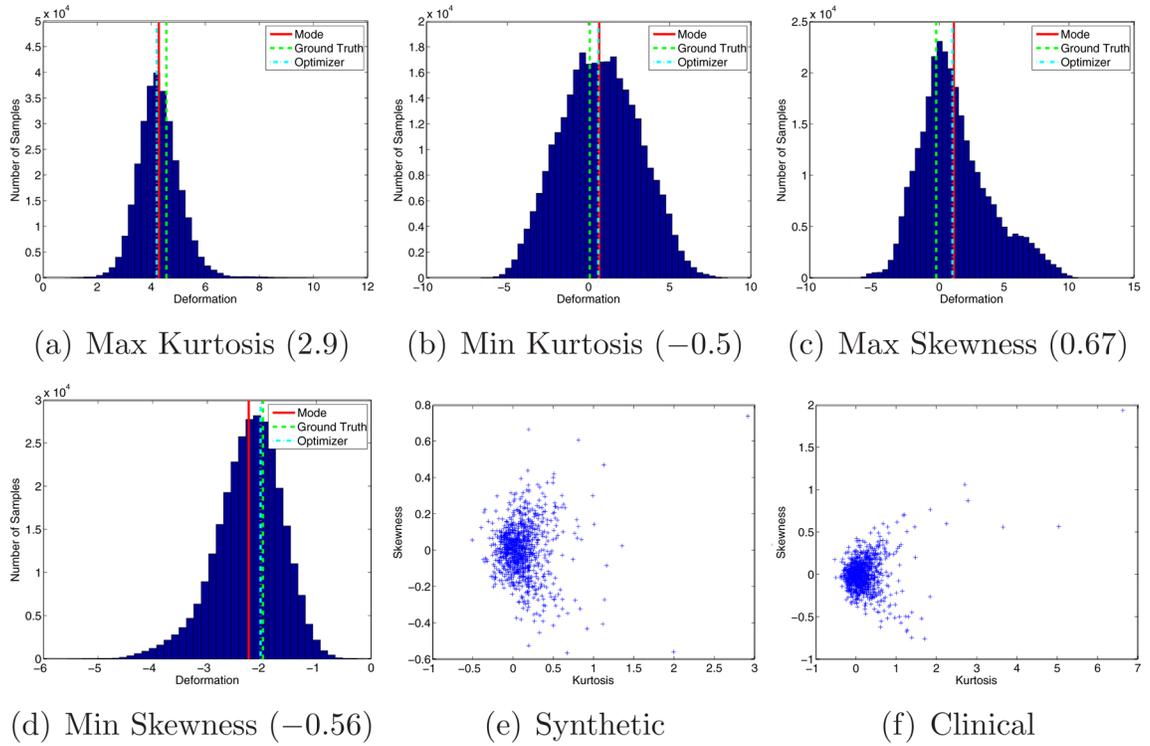


Figure 9.

Marginal distributions of the components with min/max skewness and kurtosis from the posterior distribution characterizing the displacement field which aligns the synthetic MRI dataset in Fig. 3. A normal distribution is assumed to have zero skewness and kurtosis. **(a)** A plot of the marginal distribution with the highest kurtosis (2.9). **(b)** The marginal distribution with the lowest kurtosis (-0.5). **(c)** Marginal distribution with the highest skewness (0.67). **(d)** Marginal distribution with the lowest skewness (-0.56). **(e)** A scatter plot of kurtosis vs skewness for the marginal distributions on displacements for the synthetic data in Fig. 3. **(f)** A scatter plot of the kurtosis vs skewness for the marginal distributions on displacements for the clinical data in Fig. 10. The marginal distributions have a bias towards being leptokurtic (positive kurtosis) which is shown in the distributions by an acute peak and heavier tails than a normal distribution.

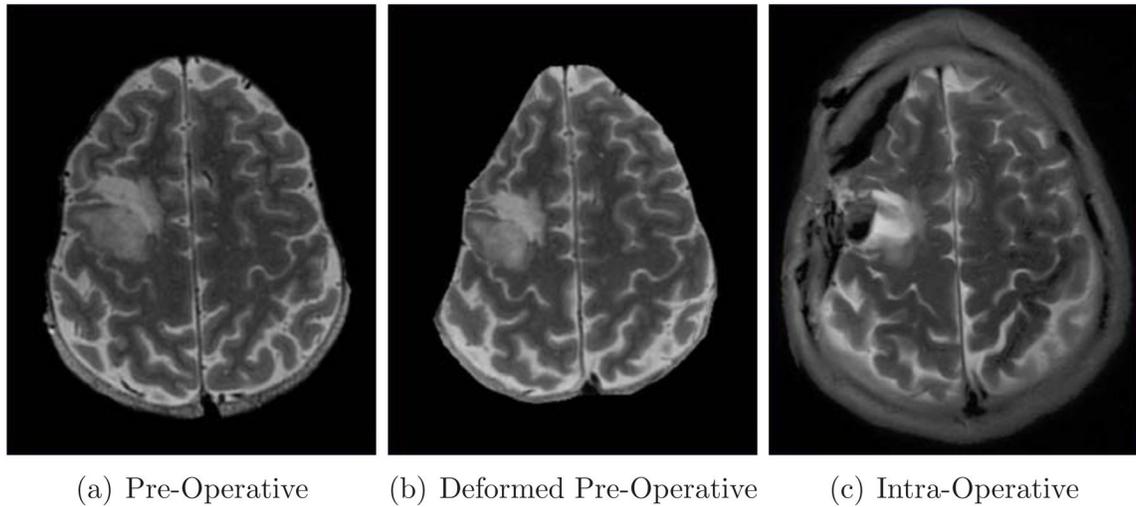
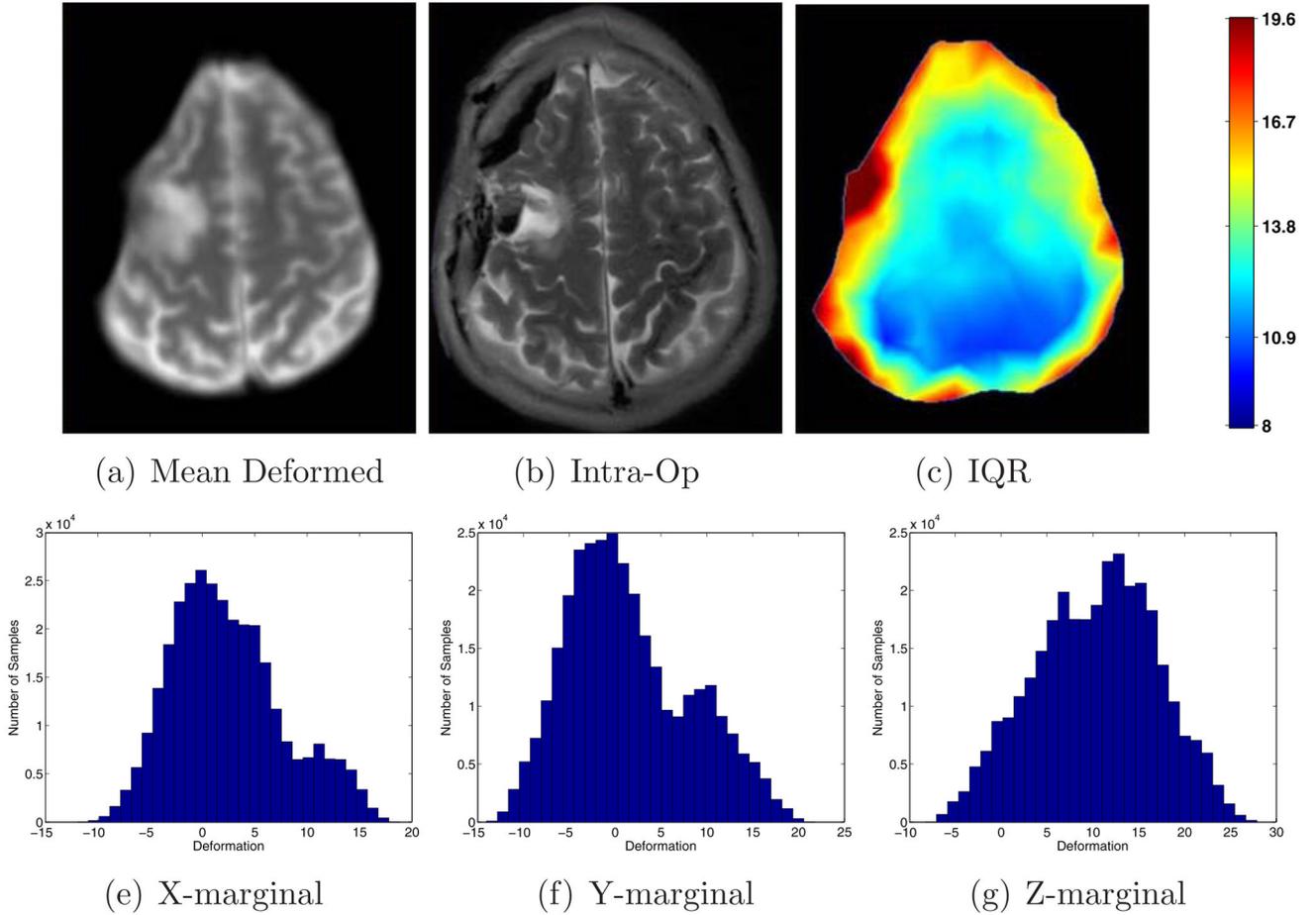


Figure 10. Corresponding axial slices of the clinical MRI data used for the simulations. **(a)** A skull stripped pre-operative T2 FSE MR-image acquired prior to neurosurgery. The FE-mesh discretizes the image domain represented by the remaining brain tissue after the skull is stripped. **(b)** Deformed pre-operative image to match the intra-operative image. **(c)** An intra-operative T2 BLADE MR image acquired after resection. Note that skull stripping the intra-operative image can be challenging due to the resection and is not required in the proposed registration procedure.

**Figure 11.**

Qualitative registration results for the clinical dataset in Fig. 10. All images show corresponding axial slices. **(a)** Mean marginal deformed pre-operative image. Given the posterior deformation samples $\{\mathbf{u}^s\}_{s=1,\dots,S}$, we sum over all deformed pre-operative images

and normalize with the number of samples $\widehat{f}_{\mathbf{u}} = \frac{1}{S} \sum_{s=1}^S f \circ \mathbf{u}^s$. This is equivalent to the expected deformed pre-operative image with respect to the posterior distribution over deformations $E_{\mathbf{u}|f,m}(f \circ \mathbf{u})$. Areas of high uncertainty are relatively blurred compared to areas of high certainty. Notice that the resection area is relatively blurry compared to other areas. **(b)** The axial slice of the intra-operative image corresponding to the slice shown in (a). **(c)** The uncertainty of the estimated deformation in terms of the IQRs (summed over the IQR for the x-, y- and z- components of the deformation field) for the same slice as visualized in (a) and (b). The units of the colorbar is in millimeters. Notice the high uncertainty around the resection where the (summed) IQR is 19.6mm. **(d-f)** The marginal distributions (x-, y- and z-components) for the FE-vertex closest to the resection cavity. We notice that the distributions are multi-modal, skewed and consequently non-Gaussian.

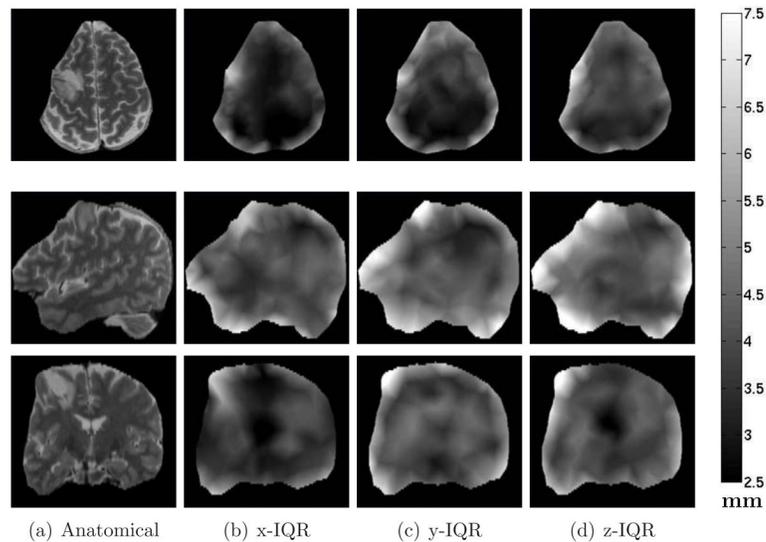


Figure 12.

This figure visualizes the directional uncertainty for the estimated deformation that aligns the clinical dataset in Fig. 10 for an axial (row 1), sagittal (row 2) and coronal (row 3) slice. **(a)** Deformed (with MAP estimate) pre-operative image. **(b)** The IQR along the x-direction (left-right according to patient). **(c)** The IQR along the y-direction (anterior-posterior). **(d)** The IQR along the z-direction (superior-inferior). Uncertainty is clearly increased around the resection. It is worth noting that the uncertainty is in general also somewhat higher along the cortical surface, i.e. the outer boundary of the FE-mesh because there are fewer constraints in the FE-mesh due to the reduction in connected elements/vertices. However, it is straightforward in our Bayesian registration framework to reduce the boundary uncertainty by including prior information, e.g. that boundary vertices should not deform into the skull defined in the intra-operative domain.

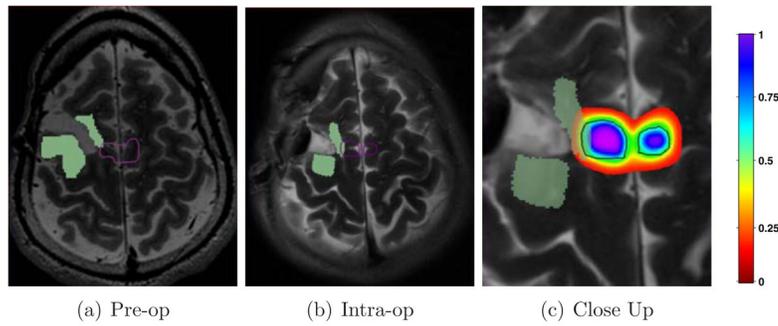


Figure 13.

Predictive posterior distribution of fMRI activated area from hand tapping in relation to intra-operative residual tumor. **(a)** fMRI is acquired pre-operatively and associated with a pre-operative anatomical image. Here we show the pre-operative image with the tumor, as delineated by a neurosurgeon, overlaid in green and the outline of the fMRI activated area overlaid in purple. Notice that the fMRI activated area is dangerously close to the tumor. **(b)** Because fMRI cannot be acquired intra-operatively, it needs to be registered to an intra-operative image to locate it in relation to the tumor during surgery. Here we show the intra-operative image overlaid with the residual tumor in green, and the fMRI activated area, deformed by the most probable deformation estimate, in purple. **(c)** Close-up of the location in (b), but with the predictive posterior distribution on the location of the deformed motion activated area overlaid. The most probable deformation of the fMRI activated area is shown with a black outline. In this surgical case, where the posterior predictive distribution was computed retrospectively, it was deemed too risky to resect the tumor close to the fMRI activated area because the true intra-operative location of the fMRI area could not be assessed properly. However, if the full posterior predictive distribution on the location of the functional area in the intra-operative space was available, e.g. as in Fig. (c), the surgeon may have been able to better assess the surgical risk and continued the resection.

Table 1

Summary of the results from registration of the synthetic dataset. The ground truth deformation \mathbf{u}_g was compared with the result from the deterministic optimizer \mathbf{u}_o , the posterior joint mode \mathbf{u}_j and the posterior marginal mode \mathbf{u}_m . The deformation vectors are of length $3V$ where $V=317$ is the number of vertices in the FE-mesh. All values are in millimeters. Note that the optimizer and the non-parametric mode using mean-shift locate the same MAP estimate (with a maximum error of 0.25mm), while the marginal estimates disagree by a maximum of 2.80mm.

Distance Type	Maximum	95 percentile	Median
$ \mathbf{u}_g - \mathbf{u}_j $	2.41	0.79	0.19
$ \mathbf{u}_g - \mathbf{u}_m $	2.80	0.78	0.18
$ \mathbf{u}_g - \mathbf{u}_o $	2.39	0.78	0.20
$ \mathbf{u}_o - \mathbf{u}_j $	0.25	0.12	0.03
$ \mathbf{u}_j - \mathbf{u}_m $	0.90	0.21	0.03

Table 2

The sensitivity of the posterior distribution to the HPs was explored by estimating the posterior mode for four separate sets of dispersed HPs selected from the tails of the marginalized HPs shown in Fig. 6. This table summarizes the specific values of the HPs in the first column, and the maximum error, 95th percentile error and median error, all reported in millimeters, are summarized in the following columns. The posterior mode was found to be sensitive to using point-estimates of the HPs, with a maximum error compared to the ground truth in the range of 6.43mm-8.56mm compared to a maximum error of 2.41mm when marginalizing HPs under uninformative hyper-priors.

(τ_r, τ_s)	Maximum	95th percentile	Median
(1050, 0.16)	6.43	1.71	0.34
(1350, 0.22)	8.56	2.07	0.35
(1050, 0.22)	6.72	1.98	0.32
(1350, 0.16)	8.29	2.12	0.32
Marginalized	2.41	0.79	0.19

Table 3

Summary of the results (in millimeters) from the registration with the neurosurgical dataset in Fig. 10. The MAP estimate \mathbf{u}_o from the deterministic optimizer described in Section 3.1.2 is compared with the posterior joint mode \mathbf{u}_j and the posterior marginal modes \mathbf{u}_m . The deformation vectors are of length $3V$ where $V = 858$ is the number of vertices in the FE-mesh. Notice that the optimizer and mean shift on the posterior samples are finding the approximately same MAP estimate (with a maximum error of 0.46mm), while the marginal estimates are off by a maximum of 1.85mm.

Distance Type	Maximum	95th percentile	Median
$ \mathbf{u}_o - \mathbf{u}_j $	0.46	0.19	0.05
$ \mathbf{u}_m - \mathbf{u}_o $	1.82	0.77	0.17
$ \mathbf{u}_m - \mathbf{u}_j $	1.85	0.76	0.16

Algorithm 1

Outline of Monte Carlo sampling algorithm with marginalization of hyper-parameters through Laplace approximations. The initial transformation \mathbf{u}^0 can either be set to a zero deformation, or alternatively to the result of a fast deterministic registration method. The log temperatures τ^0 are typically set to the mean of the log normal prior. Note that in Appendix B, where the marginalization through Laplace approximations is derived, $\theta_r \triangleq \ln \tau_r$ and $\theta_s \triangleq \ln \tau_s$ and $\Theta = \{\theta_r, \theta_s\}$ to simplify the mathematical derivation.

```

begin
  Initialize  $\mathbf{u}^0$  and  $\mathcal{T}^0$ .
for  $s := 1$  to  $S$  do
  // MCMC proposal sampling
  Generate a sample  $\mathbf{u}'$  from the proposal distribution in Eqn. (7).

  // Laplace approximation
  Set  $\mathcal{T} := \mathcal{T}^{s-1}$ 
  repeat
    // Maximize  $\ln p(\mathcal{T}, \mathbf{u}'|f, m)$  wrt  $\mathcal{T}$ 
    For current value of  $\mathcal{T}$  compute
     $\mathbf{u}^* = \operatorname{argmin}_{\mathbf{u}} [E_{el}(\mathbf{u})/\tau_r + E_{ssd}(\mathbf{u})/\tau_s]$ 
    Compute  $\ln Z(\mathcal{T})$ ,  $\nabla_{\mathcal{T}} \ln Z(\mathcal{T})$  and  $\nabla \nabla_{\mathcal{T}}^{\top} \ln Z(\mathcal{T})$  as per
    eqns. (C.3)– (C.5) around  $\mathbf{u}^*$ 
    Do a Newton update for  $\mathcal{T}$  as per eqns. (B.3)– (B.4)
  until Convergence of updates to  $\ln p(\mathcal{T}, \mathbf{u}'|m, f)$ 
  Set  $\mathcal{T}^s := \mathcal{T}^*$ , the maximum of  $\ln p(\mathcal{T}, \mathbf{u}'|m, f)$ 

  // Evaluate MH criterion
  Compute MH criterion in Eqn. (6) with  $\ln p(\mathbf{u}'|f, m)$  as given in
  Eqn. (B.5) and  $q$  approximated at  $\mathcal{T}^s$ 
  If accepted,  $\mathbf{u}^s = \mathbf{u}'$ , otherwise  $\mathbf{u}^s = \mathbf{u}^{s-1}$ 

```
