Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER • OPEN ACCESS

Bayesian non-linear regression with spatial priors for noise reduction and error estimation in quantitative MRI with an application in T1 estimation

To cite this article: Tommy Löfstedt *et al* 2020 *Phys. Med. Biol.* **65** 225036

IOP Publishing Phys. Med. Biol. 65 (2020) 225036 https://doi.org/10.1088/1361-6560/abb9f5

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
17 June 2020

REVISED
24 August 2020

ACCEPTED FOR PUBLICATION
18 September 2020

PUBLISHED
19 November 2020

PAPER

Bayesian non-linear regression with spatial priors for noise reduction and error estimation in quantitative MRI with an application in T1 estimation

Tommy Löfstedt[1,2,3], **Max Hellström**[1,3], **Mikael Bylund**[1] and **Anders Garpebring**[1]

[1] Department of Radiation Sciences, Umeå University, Umeå, Sweden
[2] Department of Computing Science, Umeå University, Umeå, Sweden
[3] Equally contributing authors.

E-mail: max.hellstrom@umu.se

Keywords: Bayesian statistics, quantitative MRI, noise reduction, tissue parameter estimation, WAIC

Abstract

Purpose. To develop a method that can reduce and estimate uncertainty in quantitative MR parameter maps without the need for hand-tuning of any hyperparameters. *Methods.* We present an estimation method where uncertainties are reduced by incorporating information on spatial correlations between neighbouring voxels. The method is based on a Bayesian hierarchical non-linear regression model, where the parameters of interest are sampled, using Markov chain Monte Carlo (MCMC), from a high-dimensional posterior distribution with a spatial prior. The degree to which the prior affects the model is determined by an automatic hyperparameter search using an information criterion and is, therefore, free from manual user-dependent tuning. The samples obtained further provide a convenient means to obtain uncertainties in both voxels and regions. The developed method was evaluated on T_1 estimations based on the variable flip angle method. *Results.* The proposed method delivers noise-reduced T_1 parameter maps with associated error estimates by combining MCMC sampling, the widely applicable information criterion, and total variation-based denoising. The proposed method results in an overall decrease in estimation error when compared to conventional voxel-wise maximum likelihood estimation. However, this comes with an increased bias in some regions, predominately at tissue interfaces, as well as an increase in computational time. *Conclusions.* This study provides a method that generates more precise estimates compared to the conventional method, without incorporating user subjectivity, and with the added benefit of uncertainty estimation.

1. Introduction

Quantitative magnetic resonance imaging (qMRI) can provide measurements of tissue properties that are independent of the exact details of the data acquisition, and simultaneously often provide interpretations of measurements (Tofts 2003). Several applications of qMRI can be found in cancer diagnostics and follow-ups, e.g. prostate cancer staging using apparent diffusion coefficient imaging (Fütterer 2016) and dynamic contrast-enhanced MRI (DCE-MRI) for early response assessment (Pham *et al* 2017). There are also numerous applications outside of oncology, for example T_1 and T_2 relaxometry in assessment of multiple sclerosis (Bitsch *et al* 2001, Tozer *et al* 2005), Parkinson's disease (Nürnberger *et al* 2017), and renal function (Wood 2014).

Quantitative parameters are usually obtained by fitting a non-linear regression model to the data in each voxel. Typically, the model is fit by minimising the mean squared error, corresponding to a maximum likelihood estimator assuming independent Gaussian noise in the data. The voxel-by-voxel approach is intuitive and simple to implement, but may result in noisy parameter maps, in particular when the model contains parameters that are difficult to estimate. The noise in the parameter maps can be reduced by including a regularisation term when fitting the model to the data. In a Bayesian interpretation, this

regularisation term can be seen as the corresponding prior distribution over the model parameters. Structured spatial regularisation terms and priors have been used in qMRI by several authors. They have, for instance, been used when estimating relaxation times and proton density (Wang and Cao 2012, Baselice *et al* 2016, Kumar *et al* 2012, Raj *et al* 2014), in diffusion and intra-voxel incoherent motion (IVIM) estimation (While 2017, Orton *et al* 2014), for B_0-estimation (Baselice *et al* 2010), and for dynamic contrast-enhanced MRI (DCE-MRI) (Schmid *et al* 2006, Kelm *et al* 2009, Sommer and Schmid 2014, Bartos *et al* 2019).

The inclusion of structured spatial regularisation terms or priors usually means there will be regularisation parameters or hyperparameters that must be tuned to the data. Selecting the hyperparameter is difficult, and how it is done varies between studies. Examples of different approaches include visual inspection (Kumar *et al* 2012), finding good parameters for a representative dataset (Bartos *et al* 2019, Wang and Cao 2012, Freiman *et al* 2013), L-curve analysis (Kumar *et al* 2012), constraints on the size of the residuals (Raj *et al* 2014), and non-informative priors have been used in the Bayesian framework (Orton *et al* 2014).

An often overlooked part of qMRI is the uncertainty in the parameter maps. Knowledge of the uncertainty can add much value since it gives a means to determine to what degree the data can be trusted, enables more accurate statistical analysis, and can be used as a tool when optimising the image acquisition. Typically, the uncertainty is estimated in each voxel separately, using for instance linear error propagation (Schabel and Parker 2008, Garpebring *et al* 2013). However, this simple approach is no longer valid if spatial regularisation or priors are used. In that case, one needs to take into account that the noise in the parameter maps is spatially correlated. By modelling the entire distribution for the parameter maps and sampling from this distribution using, for example, Markov chain Monte Carlo (MCMC), it is possible to obtain the uncertainties in the presence of spatial priors, as has been demonstrated in a few studies, e.g. in Orton *et al* (2014)), Schmid *et al* (2006), Sommer and Schmid (2014), Glad and Sebastiani *et al* (1995), De Pasquale *et al* (2000). This approach has the added benefit, relative to linear error propagation, that the non-linear noise propagation is accounted for, which is highly relevant when there is much noise, and for models that are highly non-linear for some parameter combinations. The downside is that MCMC can be computationally expensive.

Ideally, the estimation of parameter maps should incorporate some prior knowledge in order to reduce uncertainties; it should include automatic selection of hyperparameters to make the method less subjective and make it more easy to use; and include error estimation to improve the understanding, interpretability, and reliability of the parameter maps. The purpose of this work was to develop a method that can achieve this, and evaluate it for the case of T_1-estimation based on the variable flip angle (VFA) method.

2. Methods

2.1. Statistical model

In the descriptions that follow, $\boldsymbol{\theta} \in \mathbb{R}^{V\times 2}$ denotes a collection of two tissue parameter maps over V voxels. An individual parameter map is denoted by $\boldsymbol{\theta}_{\cdot,p} \in \mathbb{R}^V$, where $p \in \{1,2\}$ is the index of the particular parameter. Let $v \in \{1,2,\ldots,V\}$ be an index over the voxels in the analysed region. All parameters in the vth voxel are then denoted by $\boldsymbol{\theta}_{v,\cdot} \in \mathbb{R}^2$. The value of a specific tissue parameter, p, at voxel v is denoted $\theta_{v,p}$.

The relationship between the tissue parameters and the measured VFA signal is modelled as

$$y_{v,m} = s_m(\boldsymbol{\theta}_{v,\cdot}) + \varepsilon_{v,m}, \tag{1}$$

where $m \in \{1,2,\ldots,M\}$ is an index for each acquired image, i.e. an index for the flip angles, the analysed magnitude signal, $y_{v,m}$, is an element of $\boldsymbol{y} \in \mathbb{R}^{V\times M}$, the $s_m(\boldsymbol{\theta}_{v,\cdot})$ denotes the spoiled gradient echo (SPGR) signal model, and $\varepsilon_{v,m}$ denotes independent Gaussian noise with mean zero and the same variance, σ^2, for all voxels and flip angles. The assumption of Gaussian distributed noise is valid in our case of high SNR conditions. At lower signal intensities, the noise needs to be modelled as Rician to be properly described (Gudbjartsson and Patz 1995). The SPGR signal model is given by Tofts (2003):

$$s_m(\boldsymbol{\theta}_{v,\cdot}) = \rho_v \sin(\alpha_m) \frac{1-\exp(-T_R/T_{1v})}{1-\cos(\alpha_m)\exp(-T_R/T_{1v})}, \tag{2}$$

where $\boldsymbol{\theta}_{v,\cdot} = [\rho_v, T_{1v}]^{\mathrm{T}}$. At each voxel, indexed by v, the ρ_v is proportional to the proton density and T_{1v} is the spin-lattice relaxation time. The T_R denotes the repetition time and α_m, for $m = 1,\ldots,M$, denotes the different flip angles. Also, note that any dependence on the echo time has been absorbed into ρ_v.

Estimating the parameters of this model has traditionally corresponded to maximising the data likelihood, i.e. $p(\boldsymbol{y}\,|\,\boldsymbol{\theta},\sigma^2)$. Since the model targets, $y_{v,m} \sim \mathcal{N}(s_m(\boldsymbol{\theta}_{v,\cdot}),\sigma^2)$, are assumed independent, the

data likelihood becomes

$$p(\boldsymbol{y}\,|\,\boldsymbol{\theta},\sigma^2) = \prod_{m=1}^{M}\prod_{v=1}^{V} p(y_{v,m}\,|\,\boldsymbol{\theta}_{v,\cdot},\sigma^2)$$
$$= (2\pi\sigma^2)^{-MV/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{m=1}^{M}\sum_{v=1}^{V}\left(y_{v,m}-s_m(\boldsymbol{\theta}_{v,\cdot})\right)^2\right). \tag{3}$$

Equations (1) and (2) can be used for non-linear least squares regression between the measured signal and the VFA signal to estimate the parameters of the model, corresponding to maximising the data likelihood in equation (3). In this work, we present an extension to the maximum likelihood approach where we employ an hierarchical Bayesian model that incorporates a spatial prior over the parameters. The prior probabilities for the tissue parameters are modelled as neighbourhood dependencies expressed using the total variation function (Rudin *et al* 1992), $\mathrm{TV}(\boldsymbol{\theta}_{\cdot,p})$, as the energy function in a Boltzmann distribution (Gibbs distribution).

Let $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]^{\mathrm{T}}$ be positive hyperparameters that relates to the image gradient for the two parameters in each spatial location, then the joint spatial prior distribution over the parameters is given by

$$p_{\mathrm{TV}}(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda}) := C_{\mathrm{TV}}\prod_{p=1}^{2} p_{\mathrm{TV}}(\boldsymbol{\theta}\,|\,\lambda_p)$$
$$\propto \prod_{p=1}^{2}\lambda_p^V e^{-\lambda_p \mathrm{TV}(\boldsymbol{\theta}_{\cdot,p})},$$

in which C_{TV} is the normalising factor (partition function) and $\mathrm{TV}(\boldsymbol{\theta})$ is given as the the common discrete approximation of the total variation function, i.e.

$$\mathrm{TV}(\boldsymbol{x}) \approx \sum_{i,j}\sqrt{\left(\boldsymbol{x}_{v(i+1,j)}-\boldsymbol{x}_{v(i,j)}\right)^2+\left(\boldsymbol{x}_{v(i,j+1)}-\boldsymbol{x}_{v(i,j)}\right)^2}, \tag{4}$$

for the case of a 2D image with voxels indexed by i and j, such that $v(i,j)$ maps the spatial coordinates to the linear indices of the input vector, $\boldsymbol{x}$. That is, the total variation function is the sum of the norms of the spatial gradients at each voxel, here approximated by the forward difference. There is a possibility that a voxel indexed by $(i+1,j)$ or $(i,j+1)$ does not exist (as is the case along the borders and in the corners). In that case, the quadratic term containing this term is set to zero (i.e. essentially using reflection padding). To make the equations easier to read, these special cases have not been included here but are implicitly assumed. Details on how the form of the total variation prior was derived is given in appendix A. The parameters are also constrained by a uniform prior over a compact feasible region. This is needed to make the total variation-based prior proper and to ensure convergence in the sampling-procedure. This is given by

$$p_{\mathrm{unif}}(\boldsymbol{\theta}\,|\,l_\rho,h_\rho,l_{T_1},h_{T_1}) = \prod_{v=1}^{V}\mathcal{U}(\theta_{v,1}\,|\,l_\rho,h_\rho)\mathcal{U}(\theta_{v,2}\,|\,l_{T_1},h_{T_1}), \tag{5}$$

where l and h denote the lower- and upper limits of the uniform prior, $\mathcal{U}$, in terms of ρ and T_1. Hence, the full prior on the parameters is given as

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda},l_\rho,h_\rho,l_{T_1},h_{T_1}) = p_{\mathrm{TV}}(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda})\,p_{\mathrm{unif}}(\boldsymbol{\theta}\,|\,l_\rho,h_\rho,l_{T_1},h_{T_1}). \tag{6}$$

The hyper-prior for σ was defined using the non-informative prior

$$p(\sigma^2) \propto \sigma^{-1}, \tag{7}$$

i.e. the Jefferys' prior for the standard deviation $\sigma > 0$ for Gaussian distributions (Gelman *et al* 2014b). The prior over $\boldsymbol{\lambda}$ was set to

$$p(\boldsymbol{\lambda}) \propto \prod_{p=1}^{2}\lambda_p^{-\kappa V}, \tag{8}$$

which is a non-informative (improper) prior with unknown hyper-parameter κ. Combining equations (3), (6), (7), and (8) gives us the full log posterior that was used in this work, namely,

$$\log p(\boldsymbol{\theta},\boldsymbol{\lambda},\sigma^2\,|\,\boldsymbol{y}) = -(MV+1)\log\sigma \tag{9}$$

$$
\begin{aligned}
&- \frac{1}{2\sigma^2} \sum_{m=1}^{M} \sum_{v=1}^{V} \left(y_{v,m} - s_m(\boldsymbol{\theta}_{v,\cdot})\right)^2 \\
&- \sum_{p=1}^{2} \left[\lambda_p \mathrm{TV}(\boldsymbol{\theta}_{\cdot,p}) - (1-\kappa) V \log \lambda_p\right] \\
&+ C,
\end{aligned}
$$

where $C = \log\left((2\pi)^{-MV/2} p_{\mathrm{unif}}(\boldsymbol{\theta} \,|\, l_\rho, h_\rho, l_{T_1}, h_{T_1})\, C_{\mathrm{TV}}\right)$. The joint prior of the total variation and uniform priors in equation (6) is proper in this case, since its integral over the support of the uniform prior is finite.

2.2. Parameter estimation

Inference was done by drawing samples from the posterior in equation (9) using MCMC. To sample from the posterior density, we utilised a blocked Gibbs sampler, comprised of a Gibbs sampler (Geman and Geman 1984) for the hyperparameters, σ^2 and $\boldsymbol{\lambda}$, and the affine invariant ensemble sampler by Goodman and Weare (2010) for the likelihood parameters, $\boldsymbol{\theta}$.

To sample from $p(\boldsymbol{\theta}, \boldsymbol{\lambda}, \sigma^2 \,|\, \boldsymbol{y})$, we thus need to construct the three conditional densities $p(\sigma^2 \,|\, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{y})$, $p(\boldsymbol{\lambda} \,|\, \boldsymbol{y}, \boldsymbol{\theta}, \sigma^2)$, and $p(\boldsymbol{\theta} \,|\, \boldsymbol{y}, \boldsymbol{\lambda}, \sigma^2)$ and sample from them one at a time.

The conditional densities for the hyperparameters are straight-forward to derive from equation (9), and can easily be found to be

$$
p(\sigma^2 \,|\, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{y}) \propto (\sigma^2)^{-(A_\sigma + 1)} e^{-B_\sigma / \sigma^2},
$$

i.e. an inverse gamma distribution with shape parameter $A_\sigma = (MV - 1)/2$ and scale parameter $B_\sigma = \frac{1}{2} \sum_{m=1}^{M} \sum_{v=1}^{V} (y_{v,m} - s_m(\boldsymbol{\theta}_{v,\cdot}))^2$; and

$$
p(\lambda_i \,|\, \lambda_j, \boldsymbol{y}, \boldsymbol{\theta}, \sigma^2) \propto \lambda_i^{A_{\lambda_i} - 1} e^{-\lambda_i / B_{\lambda_i}},
$$

i.e. a gamma distribution with shape parameter $A_{\lambda_i} = (1-\kappa)V + 1$ and scale parameter $B_{\lambda_i} = 1/\mathrm{TV}(\boldsymbol{\theta}_{\cdot,i})$, with $i \in \{1, 2\}$ and $j = 3 - i$. The conditional density for $\boldsymbol{\theta}$ is a bit more involved. We write the TV term in equation (9), using the approximation in equation (4), as

$$
\sum_{p=1}^{2} \lambda_p \mathrm{TV}(\boldsymbol{\theta}_{\cdot,p}) = \sum_{p=1}^{2} \lambda_p \sum_{i,j} \sqrt{(\boldsymbol{\theta}_{v(i+1,j),p} - \boldsymbol{\theta}_{v(i,j),p})^2 + (\boldsymbol{\theta}_{v(i,j+1),p} - \boldsymbol{\theta}_{v(i,j),p})^2}. \tag{10}
$$

We note that all terms associated with a voxel with index (i, j) can be identified by looking at a 3×3 neighbourhood around that voxel, where voxel (i, j) is in the centre. This is illustrated in figure 1. Hence, a conditional distribution for the centre voxel only depends on the surrounding 3×3 neighbourhood. We exploit this property in order to develop an efficient sampler.

It is clear from figure 1 that only three terms are required in the part associated with the total variation prior for a particular voxel, and hence we can construct a conditional log-posterior density over the parameters in a voxel v as

$$
\begin{aligned}
&\log p\left(\boldsymbol{\theta}_{v(i,j),\cdot} \,|\, \boldsymbol{\theta}_{-v(i,j),\cdot}, \boldsymbol{\lambda}, \sigma^2, \boldsymbol{y}\right) \\
&\quad \propto - \frac{1}{2\sigma^2} \sum_{m=1}^{M} \left(y_{v(i,j),m} - s_m(\boldsymbol{\theta}_{v(i,j),\cdot})\right)^2 \\
&\qquad - \sum_{p=1}^{2} \lambda_p \sqrt{\left(\boldsymbol{\theta}_{v(i+1,j),p} - \boldsymbol{\theta}_{v(i,j),p}\right)^2 + \left(\boldsymbol{\theta}_{v(i,j+1),p} - \boldsymbol{\theta}_{v(i,j),p}\right)^2} \\
&\qquad - \sum_{p=1}^{2} \lambda_p \sqrt{\left(\boldsymbol{\theta}_{v(i,j),p} - \boldsymbol{\theta}_{v(i-1,j),p}\right)^2 + \left(\boldsymbol{\theta}_{v(i-1,j+1),p} - \boldsymbol{\theta}_{v(i-1,j),p}\right)^2} \\
&\qquad - \sum_{p=1}^{2} \lambda_p \sqrt{\left(\boldsymbol{\theta}_{v(i+1,j-1),p} - \boldsymbol{\theta}_{v(i,j-1),p}\right)^2 + \left(\boldsymbol{\theta}_{v(i,j),p} - \boldsymbol{\theta}_{v(i,j-1),p}\right)^2},
\end{aligned} \tag{11}
$$

where $\boldsymbol{\theta}_{-v,\cdot}$ denotes all voxels except the one with index v. Hence, all voxels at least three voxels apart, i.e. those that lie in different 3×3 neighbourhoods, can be sampled independently (and therefore also in parallel).

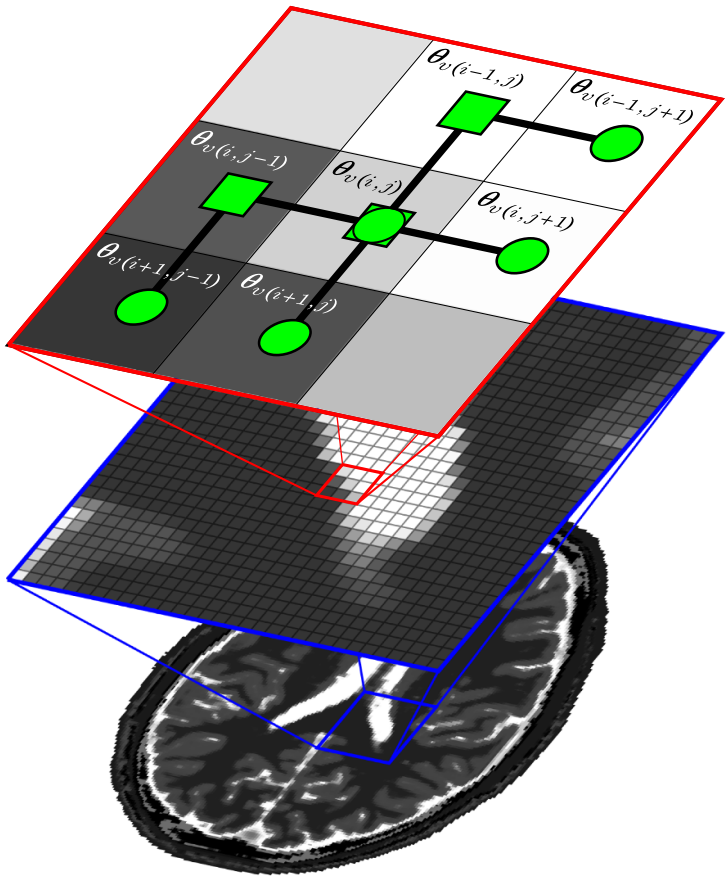

Figure 1. A zoomed-in view of a 3×3 voxel region. Each term in the total variation (equation (10)) has one central voxel (green square) and two neighbouring voxels (green circles). The voxel with index (i, j) will be included in three groups of voxels, illustrated here with black connecting lines holding the voxels in a group together. These three groups correspond to the last three summation terms in equation (11). Since the conditional distribution for the centre voxel only depends on the 3×3 neighbourhood, the analysed region with a total of V voxels can be split into $V/9$ subsets of 3×3 regions for parallel sampling.

2.2.1. Implementation

The proposed sampling method is summarised in algorithm 1 and was implemented in MATLAB R2018b (The MathWorks, Inc., Natick, MA, USA). The algorithm returns N_s samples of $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$, and σ^2. In our implementation, the parameters of the uniform prior, see equation (5), were set to $l_\rho = 0$, $h_\rho = 10^5$, $l_{T_1} = 0$ s, and $h_{T_1} = 10$ s, thus constraining the estimators to exclude physically unreasonable high- and low tissue parameter values.

The algorithm starts by computing an initial guess for the model parameters using InitialGuess, by first spatially averaging all signal in the data and fitting a single curve to this average using an ML estimator. This yields a single pair of T_1 and proton density values which are distributed over all V image voxels. To satisfy the requirement that AffineInvariantSampler needs several (N_w) walkers with nonidentical initialisation, randomness is added to each voxel and for each walker. 10 % uniform noise with zero mean is used for both T_1 and the proton density. A starting guess for σ^2 is obtained from the residual resulting from applying the parameters of the single pair of T_1 and proton density at each voxel.

Before producing the samples, the algorithm generates N_b burn-in samples that are discarded. We adapted the method of Betancourt (2010) to determine when the burn-in phase was over, i.e. computing the number of samples (N_b) required before the chains have converged to the posterior in a sufficient manner for practical use. The proposed method is presented in appendix B.

Thinning was used to reduce the correlation between samples by computing $N_s \times N_t$ samples, from which every N_t sample was kept. To determine the amount of thinning (N_t) to use, we estimated the effective sample size by computing the autocorrelation time, R, of the chains, and then stored every $\lceil 1/R \rceil$ sample (Gelman *et al* 2014b).

The AffineInvariantSampler performs MCMC sampling using the affine invariant sampler of Goodman and Weare (2010). This sampler has been shown to perform significantly faster than standard MCMC algorithms on highly skewed probability distributions. The sampler was tuned for best performance for T_1 estimation (in terms of the number of independent samples per second) by performing a grid search over

Algorithm 1: A blocked Gibbs sampling algorithm that returns $N_s \times N_w$ samples of $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$, and σ^2.

Data: $\boldsymbol{y} \in \mathbb{R}^{V \times M}$, N_s, N_t, N_w
Result: $\boldsymbol{\theta}^{(s,w)} \in \mathbb{R}^{V \times 2}$, $\lambda_{1,w}^{(s)}$, $\lambda_{2,w}^{(s)}$, $\sigma_w^{2,(s)}$ for $s = 1, \ldots, N_s$ and $w = 1, \ldots, N_w$
begin
  $\boldsymbol{\theta}^{(0)}, \sigma^{2,(0)} \leftarrow$ InitialGuess$(\boldsymbol{y})$
  $\boldsymbol{\lambda}^{(0)} = [0, 0]^{\mathrm{T}}$
  for $s \leftarrow 1$ **to** $N_b + N_s$ **do**
    $\widetilde{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}^{(s-1,\cdot)}$
    $\widetilde{\boldsymbol{\lambda}} \leftarrow \boldsymbol{\lambda}^{(s-1)}$
    $\widetilde{\boldsymbol{\sigma}}^2 \leftarrow \boldsymbol{\sigma}^{2,(s-1)}$
    for $t \leftarrow 1$ **to** N_t **do**
      for $v \leftarrow 1$ **to** V **do**
        $\widetilde{\boldsymbol{\theta}}_{v,\cdot} \leftarrow$ AffineInvariantSampler$\left(p(\boldsymbol{\theta}_{v,\cdot} \,|\, \widetilde{\boldsymbol{\theta}}_{-v,\cdot}, \widetilde{\boldsymbol{\lambda}}, \widetilde{\sigma}^2, \boldsymbol{y})\right)$
      end
      for $w \leftarrow 1$ **to** N_w **do**
        $\widetilde{\lambda}_{1,w} \leftarrow$ Sampler$\left(p(\lambda_1 \,|\, \widetilde{\lambda}_{2,w}, \widetilde{\boldsymbol{\theta}}^{(w)}, \widetilde{\sigma}_w^2, \boldsymbol{y})\right)$
        $\widetilde{\lambda}_{2,w} \leftarrow$ Sampler$\left(p(\lambda_2 \,|\, \widetilde{\lambda}_{1,w}, \widetilde{\boldsymbol{\theta}}^{(w)}, \widetilde{\sigma}_w^2, \boldsymbol{y})\right)$
        $\widetilde{\boldsymbol{\lambda}}_w \leftarrow [\widetilde{\lambda}_{1,w}, \widetilde{\lambda}_{2,w}]^{\mathrm{T}}$
        $\widetilde{\sigma}_w^2 \leftarrow$ Sampler$\left(p(\sigma^2 \,|\, \widetilde{\boldsymbol{\theta}}^{(w)}, \widetilde{\boldsymbol{\lambda}}_w, \boldsymbol{y})\right)$
      end
    end
    if $s > N_b$ **then**
      $\boldsymbol{\theta}^{(s-N_b,\cdot)} \leftarrow \widetilde{\boldsymbol{\theta}}$
      $\boldsymbol{\lambda}^{(s-N_b)} \leftarrow \widetilde{\boldsymbol{\lambda}}$
      $\boldsymbol{\sigma}^{2,(s-N_b)} \leftarrow \widetilde{\boldsymbol{\sigma}}^2$
    end
  end
end

relevant sampler settings. A step size of 4.0, thinning of 8, and 10 walkers were found to perform well over a large range of values for κ.

The Sampler denotes any generic function that draws samples from the given distribution; we used the default random number generators that are available in MATLAB for the uniform, gamma, and inverse gamma distributions. The arguments to the samplers are the density functions that samples should be drawn from.

To speed up the algorithm, we make use of the fact that the conditional probability, equation (11), for a particular voxel only contains terms associated with voxels in its immediate surrounding. That is, the loop over V voxels is implemented as a loop over 9 subsets of voxels and the AffineInvariantSampler performs calculations at $V/9$ voxels in parallel, see figure 1 for more details on these voxel subsets. To keep notation simple and not introduce yet another index, this implementation detail is not illustrated in algorithm 1.

To determine the value of the hyperparameter κ, the only unknown hyperparameter in this work, we propose to use the widely applicable information criterion (WAIC, also known as the Watanabe–Akaike information criterion) (Watanabe 2009). We used the criterion denoted WAIC_1 in Watanabe (2009) to evaluate the relative quality of our model as a function of the hyperparameter κ. The WAIC is a fully Bayesian method to estimate the out-of-sample generalisation. It is a generalization of the Akaike information criterion (AIC), an improvement of the deviance information criterion (DIC), and is asymptotically equivalent to Bayesian cross-validation (Watanabe 2009, Watanabe 2010, Gelman *et al* 2014a). The WAIC is described in more detail in appendix C.

A grid search for κ in equation (8) was conducted in the range $[-0.5, 0.99]$ in steps of 0.01. The resulting curve was smoothed using a moving average in order to find a $\kappa_{\min}$ less affected by noise in the estimations, and the κ corresponding to the minimum WAIC value in the grid was selected. Code for running the above implementation, complete with a synthetic dataset, is available at the following GitHub repository: https://github.com/MaxHellstrom/Bayesian-VFA-T1-estimation.

2.3. Data

Two sets of data were used in the evaluation of the proposed method: one synthetic dataset with known ground truth, and one dataset with eight volunteering patients.

2.3.1. Synthetic data

An axial brain slice from the BrainWeb phantom (Kwan *et al* 1996, Cocosco *et al* 1997, Collins *et al* 1998, Kwan *et al* 1999) was generated in MICE Toolkit (NONPI Medical AB, Umeå, Sweden; www.micetoolkit.com) with matrix size 256×256, and voxel size $0.98 \times 0.98 \times 2.00$ mm^3. Ground truth reference of the tissue parameters, $T_{1,\text{ref}}$ and ρ_{ref}, were extracted for all voxels in the acquired image matrix. These tissue parameters were then used to compute spoiled gradient echo (SPGR) data according to Equations (1) and 2, with $T_R = 6.8$ ms, $T_E = 2.1$ ms, flip angles $\alpha = 2^\circ,\ 4^\circ,\ 11^\circ,\ 13^\circ, and\ 15^\circ$. Complex circular Gaussian noise was used to generate synthetic noise, so that the magnitude SPGR signal follows the Rician distribution. The variance of the noise was tuned until it reached the same level as the *in-vivo* dataset by visual comparison. One hundred multi-flip angle images with independent noise were generated this way.

2.3.2. In-vivo *data*

A total of eight patients volunteered for this study and were scanned with a 3D SPGR sequence on a GE Signa 3 T PET/MR scanner. The patient group consisted of 7 men, and 1 woman, in the age span 39 to 65 year old (with a mean age of 52 years). The scans were conducted with identical T_{R}, T_{E}, voxel size, and flip angles as for the synthetic dataset. The acquisition was conducted with a matrix size of $256 \times 256 \times 8$ (8 axial slices) and a pixel bandwidth of 488 Hz/pixel. Bloch–Siegert based B_1 mapping was performed to enable corrections of flip angles. This study was conducted in accordance with the principles embodied in the Declaration of Helsinki and was ethically approved by the regional research ethics committee (dnr: 2019-02666). Informed consent was obtained from all patients.

2.4. Analysis

Three different estimators, referred to as ML, B_{unif}, and B_{TV}, were evaluated on both the synthetic and the *in-vivo* datasets. The first estimator (ML) was a conventional maximum likelihood estimator, where the parameter maps were obtained by minimising the negative log-likelihood independently in each voxel, v, i.e. solving the program

$$\underset{\boldsymbol{\theta}_{v,\cdot} \in \mathcal{R}}{\text{minimise}} \sum_{m=1}^{M} \left(y_{m,v} - s_m(\boldsymbol{\theta}_{v,\cdot})\right)^2,$$

and assigning the result to the parameter estimate $\hat{\rho}_v$ and $\hat{T}_{1v}$, for $v = 1, \ldots, V$, and where $\mathcal{R} \subset \mathbb{R}^2$ is the compact feasible region for the parameter values as defined by the uniform prior. This program was solved using the trust-region-reflective algorithm as implemented in MATLAB R2018b.

The estimators B_{unif} and B_{TV} correspond to Bayesian approaches with different priors. The difference between these two estimators is that B_{unif} only uses the uniform prior on the parameters, while B_{TV} combines the uniform and the total variation priors. Both estimators used the blocked Gibbs sampling algorithm (see algorithm 1) and collected $N_s = 1\,280$ samples (after thinning) from the posterior distribution. Point estimates for B_{unif} and B_{TV} were chosen to be the sample means, i.e.

$$\widehat{\rho}_v = \frac{1}{N_s} \sum_{s=1}^{N_s} \rho_v^{(s)} \tag{12}$$

and

$$\widehat{T}_{1v} = \frac{1}{N_s} \sum_{s=1}^{N_s} T_{1v}^{(s)}. \tag{13}$$

The synthetic dataset with 100 generated multi-flip angle images were used to investigate the performance of the point estimate when using the different estimators. The normalised bias and normalised standard error, i.e. the coefficient of variation, CV, were computed as

$$\text{Bias}_{T_{1v}} = \frac{\bar{\widehat{T}}_{1v} - T_{1v}^{\text{ref}}}{T_{1v}^{\text{ref}}}, \tag{14}$$

$$\text{Bias}_{\rho_v} = \frac{\bar{\widehat{\rho}}_v - \rho_v^{\text{ref}}}{\rho_v^{\text{ref}}}, \tag{15}$$

$$\mathrm{CV}_{T_{1v}} = \frac{1}{T_{1v}^{\mathrm{ref}}}\sqrt{\frac{1}{(N_B-1)}\sum_{b=1}^{N_B}\left(\widehat{T}_{1v,b}-\bar{\widehat{T}}_{1v}\right)^2}, \text{ and} \tag{16}$$

$$\mathrm{CV}_{\rho_v} = \frac{1}{\rho_v^{\mathrm{ref}}}\sqrt{\frac{1}{(N_B-1)}\sum_{b=1}^{N_B}\left(\widehat{\rho}_{v,b}-\bar{\widehat{\rho}}_v\right)^2}. \tag{17}$$

In the equations above, b is an index over the $N_B = 100$ generated multi-flip angle images, $\widehat{T}_{1v,b}$ and $\widehat{\rho}_{v,b}$ are the estimates of T_1 and ρ at voxel v for image b. The average estimates are defined as $\bar{\widehat{T}}_{1v} = N_B^{-1}\sum_{b=1}^{N_B}\widehat{T}_{1v,b}$ and $\bar{\widehat{\rho}}_v = N_B^{-1}\sum_{b=1}^{N_B}\widehat{\rho}_{v,b}$. The references values T_{1v}^{ref} and ρ_v^{ref} correspond to the ground truth images from which the synthetic data was generated. To conduct a quantitative evaluation of the Bayesian estimators $\mathrm{B}_{\mathrm{unif}}$ and B_{TV}, we adapted the method of Sjölund *et al* (2018) and constructed probability–probability (P-P) plots from the posterior samples. To illustrate this concept, consider a set of samples from the posterior distribution with known ground-truth reference in each voxel. A P-P plot can then be created by calculating the frequency, across all voxels, with which the reference value is smaller than the pth percentile of the posterior samples. A one-to-one P-P ratio is an ideal case, i.e. when the reference value is smaller than the pth percentile in p % of the cases. This P-P ratio can act as a sanity-test of our choice in priors. This was conducted exclusively for the synthetic data where ground truth data are available.

3. Results

3.1. Tissue parameter estimation

Parameter maps computed using all three estimators are shown in figure 2. These maps are estimated from the data from one of the patients in the *in-vivo* dataset (patient 1). Previous to the imaging occasion, the patient went through surgical removal of a grade 2 astrocytoma tumour, which is clearly visible in all images. In figure 2, the voxel intensity equals the parameter estimates $\widehat{\rho}_v$ (in 2–c) and $\widehat{T}_{1v}$ (in 2d–f). The same parameter estimates are presented as normalised histograms, $\widehat{\rho}_v$ (in 2g–i) and $\widehat{T}_{1v}$ (in 2j–l). These histograms present the same results as in (2a–f), but sorted in order of increasing parameter values, instead of spatial location. The difference between the three estimators is shown more clearly in figure 3, which shows a zoomed-in version of the T_1 estimate in figure 2(d)–(f). The degree to which fine-grained details are affected by the priors is shown in the top row of figure 3, which shows a region with several fine structures. By comparing 3e and 3f, i.e. the introduction of the TV-term, we see that fine structures are still clearly visible after the smoothing has been applied. The $\mathrm{B}_{\mathrm{unif}}$ estimator delivers estimates in a similar way as the conventional estimate (ML). Further, it is clear that B_{TV} delivers a distinguishable amount of noise reduction, as well as an overall $\widehat{T}_1$ reduction in homogeneous regions of high T_1 value. When comparing the T_1 histograms (2j–l), we see that the distribution produced with B_{TV}(2l) differs significantly from the ones produced with ML and $\mathrm{B}_{\mathrm{unif}}$ (2j-k) in the sense that the distribution is more tightly concentrated at three clearly visible peaks. These peaks are located at approximately 0.6 s, 1.2 s and 4.7 s. Parameter maps computed for the other volunteers in the *in-vivo* dataset were very similar and are presented in appendix D.

Parameter estimates with ML, $\mathrm{B}_{\mathrm{unif}}$, and B_{TV} were also conducted for the synthetic dataset of 100 generated multi-flip angle images. For the *in-vivo* dataset, the mean computation times, including burn-in, were 6.0 (ML), 6.1 ($\mathrm{B}_{\mathrm{unif}}$), and 31 ($\mathrm{B}_{\mathrm{TV}}$) minutes per patient (one axial slice). The corresponding numbers for the synthetic dataset was about 6.1, 5.6, and 28 minutes. For both datasets, the burn-in phase took about 45 % of the total computation time. The difference in computation time between the two datasets is likely affected by the number of pixels in the analysed region. Please note that hyperparameter search for κ is not included in the above stated estimation time. This is presented in detail in section 3.2. All computations were performed on a 3.7 GHz Intel Core i7-8700K processor.

Figure 4 illustrates the utility of obtaining distributions of likely values rather than only a single point-estimate. A ground truth synthetic T_1 map is shown together with histograms of the samples acquired with $\mathrm{B}_{\mathrm{unif}}$ and B_{TV} using one of the 100 synthetically generated multi-flip angle images. The total variation prior in B_{TV} produces more narrow histograms in all four locations. This is most evident in regions with large T_1 values. Further, the magnitude of the estimation bias varies between the different locations and tends to be larger in regions with large T_1 gradient. By observing e.g. figure 4(d), we see that $\mathrm{B}_{\mathrm{unif}}$ produces a positive estimation bias, while B_{TV} produces a negative bias. Please note that the histograms in figure 2 present the distribution of $\widehat{\rho}_v$ and $\widehat{T}_{1v}$ estimates in all voxels, while the histograms in figure 4 present the distribution of T_1 samples in a voxel at four different locations.

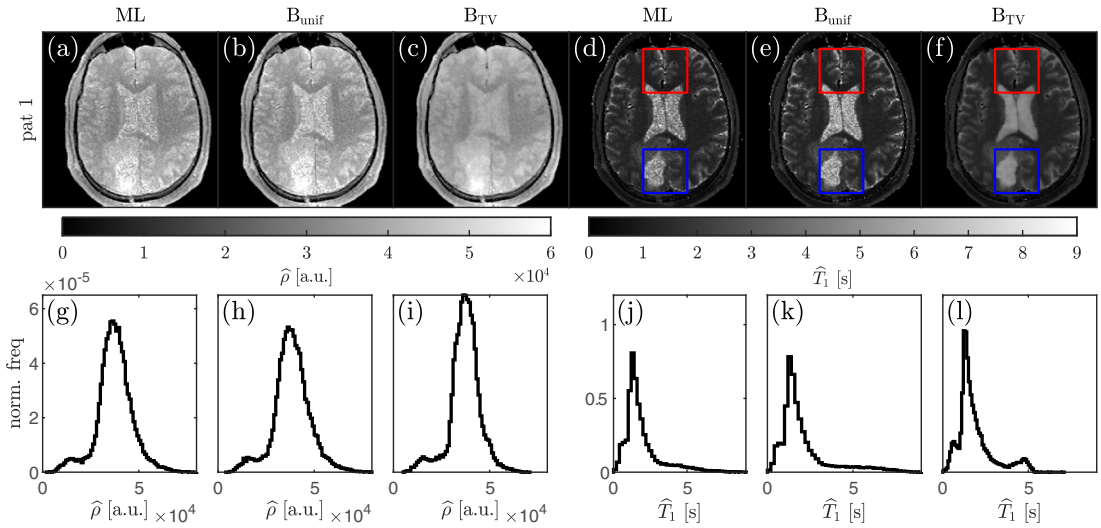

Figure 2. Parameter estimates obtained by using all three estimators for patient 1 in the *in-vivo* dataset. $\widehat{\rho}_v$ in ((a)–(c)) and $\widehat{T}_{1,v}$ in ((d)–(f)). The red and blue zoom-boxes in (d)–(f) are presented in greater detail in figure 3. The tissue parameter values in the parameter maps in (a)–(f) is also presented as normalised histograms in the subplot below.

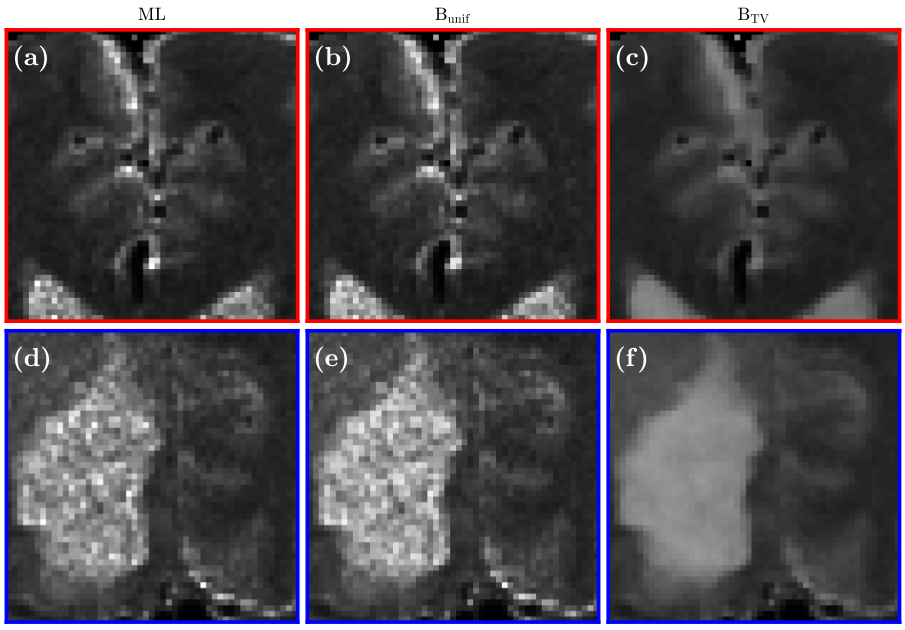

Figure 3. A 50 × 50 voxels zoomed-in view of figure 2. The red zoom-boxes in figure 2 are presented in the top row, and the blue zoom-boxes in the bottom row.

P-P plots of observed versus theoretical probabilities as well as the normalised standard deviation in the posteriors are presented in figure 5. A slight deviation from the ideal one-to-one P-P ratio is visible in both estimators, most evident in B_{TV} at higher probabilities.

The normalised bias and standard deviations, as defined in equations 14 through 17, are shown in figure 6. The left 3 × 3 array plot corresponds to the ρ_v estimate, and the right one corresponds to the T_1 estimate. Comparing the coefficient of variation in 6(a–c) shows an increase in precision when applying B_{TV}. This is most evident in regions with high parameter values and low signal, such as in the ventricles. In 6(d–f), we see that B_{unif} and B_{TV} introduce some level of estimation bias. The effect of the TV term when using B_{TV} is clearly visible in regions with sharp edges, where it introduces a negative estimation bias (see 6(f)). This bias is also presented in 6(g–i), but in terms of parameter values without spatial information.

3.2. Sampling parameters

The hyperparameter search for κ were conducted for both datasets prior to parameter estimation.

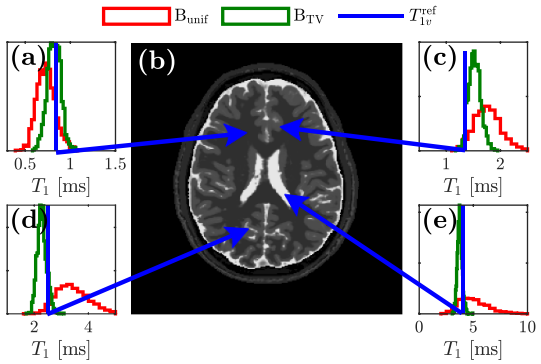

Figure 4. Presentation of the T_1 distribution consisting of 1 280 samples in each voxel, acquired with B_{unif} and B_{TV} on one randomly chosen element of the synthetic dataset. This is presented at four different voxels as normalised histograms, B_{unif} in red, B_{TV} in green and reference T_1 value (T_{1v}^{ref}) is marked with a blue line. The mean value of each sampling distribution corresponds to the T_1 estimate in that voxel, i.e. $\widehat{T}_{1v}$. Each histogram is linked to its corresponding voxel in the reference slice (b) with a blue arrow. (a) A voxel in a white matter region with low T_1 value and gradient. (c) A voxel in a grey matter region. (d) A voxel in a high-gradient region, which causes the most prominent estimation bias when applying B_{TV}. (e) A voxel in a low-gradient ventricle region, high in CSF.

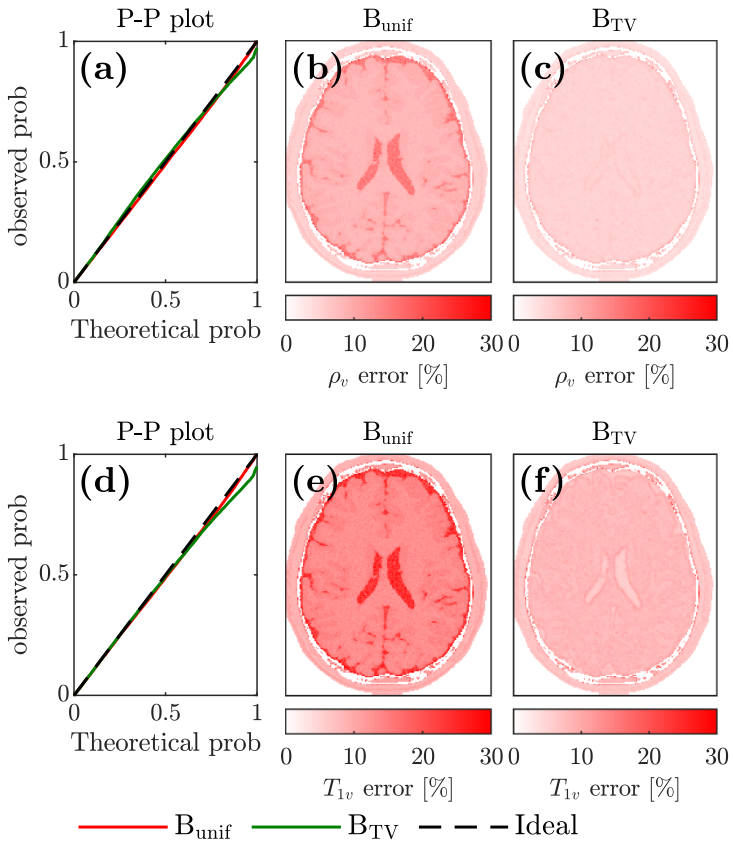

Figure 5. P-P plots of theoretical- versus observed probability when conducting tissue parameter estimation with B_{unif} and B_{TV}, both with $N_s = 1\,280$ samples (ρ in (a) and T_1 in (d)). (b), (c), (e), (f) Normalised standard deviation (in percent) of the ρ and T_1 posterior distributions. This data was obtained from parameter estimation with B_{unif} and B_{TV} on one of the 100 generated multi-flip angle images in the synthetic dataset.

For the *in-vivo* dataset, κ_{min} varied in the range $[0.07, 0.18]$ for patient 1 to 8, with a mean value of 0.11 and a standard deviation of 0.037, see figure 7. Since all searches ended up in the range $[0.07, 0.18]$ which is only about 7% of the total search window, the efficiency in this hyperparameter search can likely be increased by narrowing the search window. These hyperparameter computations took on average 9 hours per patient for the entire search window $[-0.5, 0.99]$ with a step size of 0.01.

For the synthetic dataset, a single search was conducted for all 100 images, which resulted in $\kappa_{min} = 0.06$. In addition to this, the effect of a difference in κ value was investigated by conducting parameter estimation with $\kappa = \kappa_{min} \pm \{0.05, 0.1\}$, i.e. comparing the parameter estimation result for slightly different κ values.

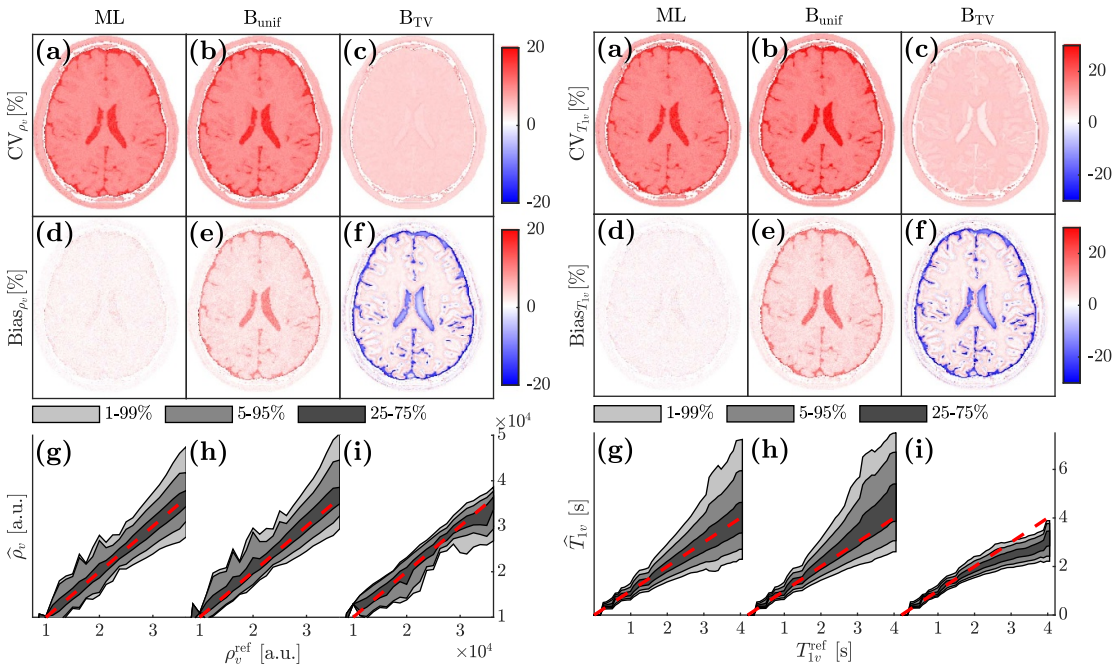

Figure 6. Coefficient of variation ((a)–(c)) and normalised estimation bias ((d)–(f)), in terms of ρ_v (left 3×3 array plots) and T_{1v} (right 3×3 array plots) estimated from the synthetic- and *in-vivo* datasets. The estimation bias is also presented in ((g)–(i)), where it is expressed as a function of the reference parameter value, grouped in the percentile ranges: 1–99%, 5–95%, and 25–75%. The dashed red line is the ideal 1:1 ratio.

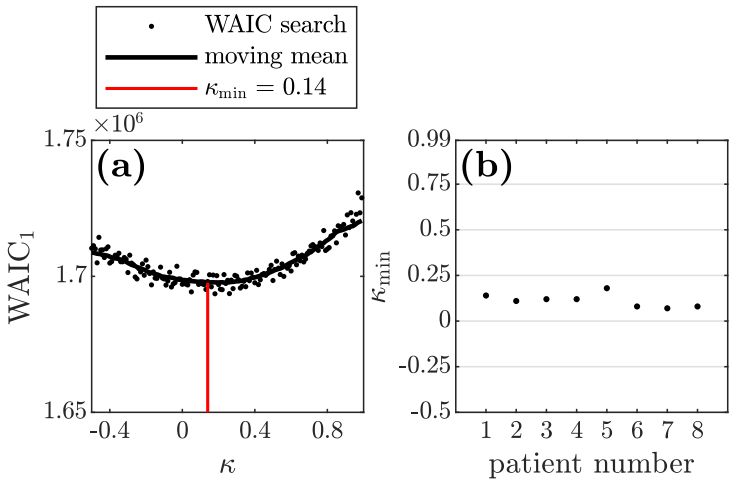

Figure 7. (a) Hyperparameter search for κ for patient 1. The selected κ value, κ_{min}, was chosen as the minimum of the smoothed WAIC_1 values over the evaluated range. (b) Result of the hyperparameter search for all patients in the *in-vivo* dataset. All κ_{min} values were obtained in the range [0.07, 0.18]. Therefore, the search window can probably be narrowed in order to increase computational efficiency.

The largest voxelwise difference was observed at $\kappa = \kappa_{\text{min}} - 0.1$ ($+$ 6.1 % in ρ_v and $+$ 4.0 % in T_{1v}). This is presented in figure 8.

Maps of the calculated required thinning, based on the autocorrelation, were investigated for both datasets, where it was evident that the amount of required thinning varies spatially. Further, it peaks at about 4 using B_{unif} and 8 when using B_{TV}. Therefore, the peak values 4 and 8 were chosen when conducting parameter estimation with B_{unif} and B_{TV}, in order to ensure a sufficient amount of uncorrelated samples in all voxels.

4. Discussion

The purpose of this work was to develop a method that can reduce and estimate uncertainty in quantitative MR parameter maps without the need for hand-tuning of hyperparameters. To this end, a Bayesian

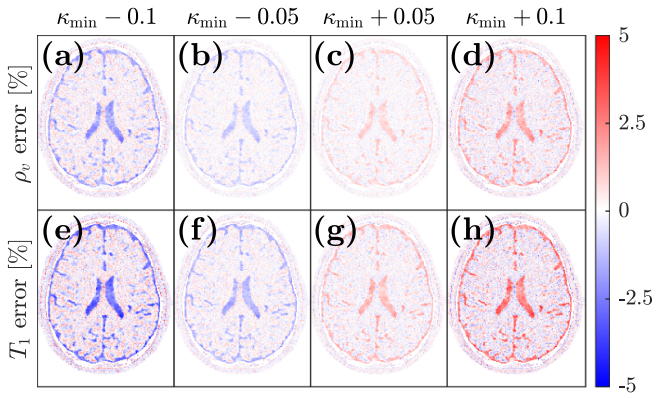

Figure 8. The error in estimation result for different κ values, in relation to $\kappa = \kappa_{\min}$. This is based on ρ_v and T_{1v} estimations conducted with the synthetic dataset. The largest positive voxelwise errors were (in subplot (a)–(h)): (6.1, 3.1, 2.6, 5.2, 4.0, 1.8, 2.0, 3.8)%. The corresponding largest negative voxelwise error were -(5.2, 2.8, 2.4, 4.8, 3.7, 1.9, 1.5, 3.1)%.

hierarchical non-linear regression model with a total variation prior was used to enable noise reduction. Uncertainties were obtained through sampling from the corresponding posterior distribution, and WAIC was used to automate the selection of hyperparameters. The method was evaluated for parameter estimation of proton density and T_1 relaxation time on VFA data. The main finding in this work is that noise reduction, uncertainty estimation, and automatic hyperparameter tuning can be combined, and that the method is applicable to VFA based T_1 estimation.

An important consideration when applying Bayesian methods for parameter estimation is the choice of prior. This choice entails both selecting the type of prior to use and also to what degree one should trust the prior or the data. With sufficient amounts of data, uninformative priors can be used, e.g. Orton *et al* (2014). However, in this work, we used the WAIC to compare models and based on that select the hyperparameters for the prior. This approach is novel in the context of qMRI and is much more objective than varying the hyperparameters until the images *look good*, or finding an optimal hyperparameter using e.g. synthetic data and using that value for subsequent estimations. A TV-based prior was used in this work due to its previous success in image denoising, including in the context of T_1 estimation using VFA data (Wang and Cao 2012). As can be observed in figures 2, 3 and 6, the TV-prior is quite successful in reducing uncertainties in T_1 and proton density values, and from a qualitative point of view, sharp borders at tissue interfaces, e.g. in figure 3(f) and the fine-grained details in figure 3(c) seem to be preserved. However, when looking more closely at the result in figure 6, it is clear that bias is introduced by the TV-based prior, in particular at interfaces between different types of tissues. This causes slightly incorrect tissue parameter values in interface-regions and fine-grained details. This bias is expected since the prior makes assumptions about differences between neighbouring voxels. This is also a sign that there likely is room for improvements. For instance, one could imagine that the use of two separate hyperparameters for the strength of the prior—one for each parameter, or using a weighted TV-based prior (e.g. such that the data is trusted more near borders)—could result in further improvements. Using other priors, e.g. exploiting wavelet compressibility (Ji *et al* 2008), could potentially improve the results. However, the MCMC algorithm used in this work is likely to perform poorly in that particular case since the developed algorithm's speed relies on each voxel being conditionally dependent only on its immediate neighbours.

It is important to mention that this estimation bias does not include contributions from inaccuracies caused by errors in the B1 map, insufficient signal spoiling, and other physical factors from the data acquisition itself. Although beyond the scope of this study, these factors could be included in a Bayesian model to incorporate their effects on the estimated uncertainty.

Obtaining the uncertainty of the estimated parameters was a key goal of this work. Several important statistical questions may be posed when uncertainty information is provided, such as, what the probability is that a parameter is above or below a certain value of clinical relevance. One particular strength of Bayesian statistics and sampling of parameter maps is that the spatially correlated noise will be properly captured, which is very important in order to obtain correct uncertainty estimates in ROI analyses, for instance.

In this work, we utilised P-P plots for quantitative evaluation of the parameter uncertainty estimation. Looking at the results in the P-P plots in figure 5(a), (d), we see a near-ideal match between observed and theoretical probability when using $\mathrm{B}_{\mathrm{unif}}$. When using B_{TV}, we see that the observed probability is slightly below the ideal 1-1 ratio. e.g. in figure 5(d) where only 91 % of the true T_1 values are in the 95th percentile. The cause of this can be either a biased posterior distribution or one with too small tails.

Although very powerful and flexible, Bayesian methods also have some downsides. Using MCMC for inference can be very computationally demanding, and may introduce complexity in terms of parameters that must be tuned for good performance, e.g. the burn-in length and amount of thinning. To reduce this complexity, we implemented a method that automatically determines the burn-in length. To reduce the computation time, we exploited the fact that parameters at different voxels can be updated in parallel in the Gibbs sampling scheme when the structure of the posterior allows for it (as was the case here). In this work, we used a TV-based prior, which assumes that voxels are conditionally dependent only on their immediate neighbours. Hence, the voxel could be updated in blocks as depicted in figure 1. This results in a substantial speed-up when using languages such as MATLAB that heavily rely on vectorisation for speed. However, even after vectorising the computations, the method cannot be considered fast. Obtaining a single slice (approximately 27 000 non-masked voxels) requires 18 minutes of burn-in and an additional 20 minutes of sampling to obtain 1 280 uncorrelated samples when using the TV-based prior. If several slices are to be computed, the estimation may take hours, which can be prohibitive—especially in clinical applications.

In addition to the estimation time, the search for the hyperparameters must also be added. A very fine grid was used in this work, requiring about 9 hours to find the hyperparameter κ for a single slice. This is a clear limiting factor that can be improved in several ways. When observing the results in figure 7, the most straight-forward approach to increase the efficiency in the hyperparameter search is to empirically narrow the search window. The maximum spread in κ_{min} we observed was 0.11, i.e. very small in comparison to the entire range we tested in this work. Simply narrowing the search grid to the range [0, 0.25] would decrease the computational time to about 1.5 hours per slice. Further, a coarser grid could likely be used since a difference in κ less than 0.1 had a relatively small effect, see figure 8, on the produced images. This small difference also implies that one likely only need to do the search for a single slice in a volume. Another promising approach is to investigate if it is possible to use a fixed κ value for all patients undergoing imaging with the same protocol. Our results in figure 7(b) and figure 8 suggests that it might be possible due to low patient variability in κ_{min}. This would drastically improve the effectiveness of our method.

To further speed up the estimation and hyperparameter search, the most obvious next step would be to move to more powerful hardware, e.g. GPUs instead of CPUs. Based on our experience of applying GPUs for computations, we expect such a migration to give improvements of at least one order of magnitude. Although this would greatly facilitate the practical applicability of the method in many situations, the computation time may still be a problem—in particular for clinical applications. Thus, there is a need for improvements of the algorithm used for sampling, perhaps using other MCMC algorithms such a e.g. Hamiltonian MCMC (Duane *et al* 1987). For the hyperparameter search, the grid-search employed in this work is quite inefficient and better methods are needed in particular if more than one hyperparameter is used. Bayesian optimisation (Snoek *et al* 2012), for instance, is commonly used for efficient hyperparameter search in other machine learning applications and would thus likely be a valuable tool here as well.

This work focused on T_1 estimation using VFA data. The proposed method is likely also applicable to other types of qMRI, e.g. diffusion and T_2/T_2^* quantification and possibly also DCE-MRI, since those estimations also imply finding a few parameters at each voxel location. Other possible generalisations and improvements could be to use other priors as mentioned above, utilise spatial information in 3D (e.g. 3D total variation), improve the gradient approximation method, and to use a larger family of priors and select the optimal one using WAIC, e.g. use WAIC to select between different functional forms for the prior and using a prior with more than a single hyperparameter, for instance, a separate κ value for T_1 and the proton density. Other possible improvements for this work would be to model the noise as spatially varying in order to properly describe the noise distribution when using e.g. parallel imaging techniques.

5. Conclusion

We have presented a novel framework for parameter estimation with associated error estimates in qMRI, applied for tissue parameter estimations of T_1 relaxation time. Due to automated hyperparameter selection, our method can deliver noise reduction without incorporating end-user subjectivity. Key features to address in future developments are to refine the design of the spatial priors, as well as to address the computational efficiency to enable more widespread use.

Conflict of interest

Tommy Löfstedt and Anders Garpebring are co-owners of NONPI Medical AB, the developer of MICE Toolkit—a software for medical image analysis that was used in this work.

Funding info

This research was in part funded by grants from The Swedish Research Council (Grant No. 2019-0432), Region Västerbotten (Central ALF, Grant No. RV-738491), Cancerforskningsfonden i Norrland (Grant No. AMP 18-912), and Lions Cancerforskningsfond (Grant No. LP 18-2182).

Appendix A. Derivation of the total variation-based prior

In order to impose the total variation function through a prior distribution on the parameters, we utilise the Boltzmann distribution (Gibbs distribution) which is a well-tried method in Bayesian image restoration (Geman and Geman 1984). The corresponding probability density function then has the form

$$p(\boldsymbol{\theta} \,|\, \lambda) = \frac{1}{Z(kT)} e^{\frac{-\phi(\boldsymbol{\theta})}{kT}} = \frac{1}{Z(\lambda)} e^{-\lambda \phi(\boldsymbol{\theta})}, \tag{A1}$$

where $\lambda = (kT)^{-1} > 0$ is a hyperparameter controlling the *temperature* of the system (k would be Boltzmann's constant, and T the temperature), $Z(\lambda)$ is the partition function ensuring that the probability density integrates to one, $\boldsymbol{\theta} \in \mathbb{R}^N$ for N parameters, and $\phi : \mathbb{R}^N \to \mathbb{R}$ is an *energy* function. In our case the energy function would be the total variation function, defined in equation (4) as

$$\begin{aligned} \mathrm{TV}(\boldsymbol{\theta}) &:= \sum_{i,j} \left\| \nabla_{i,j} \boldsymbol{\theta} \right\|_2 \\ &\approx \sum_{i,j} \sqrt{\left(\boldsymbol{\theta}_{v(i+1,j)} - \boldsymbol{\theta}_{v(i,j)}\right)^2 + \left(\boldsymbol{\theta}_{v(i,j+1)} - \boldsymbol{\theta}_{v(i,j)}\right)^2}, \end{aligned}$$

where $v(i, j)$ maps the spatial coordinates to the linear indices of the input parameter vector, $\boldsymbol{\theta}$, and where $\nabla_{i,j}\boldsymbol{\theta}$ denotes the spatial gradient in the image at location (i, j). The spatial gradient is approximated using a first-order approximation (the forward difference).

We note that the total variation function, TV, is absolutely homogeneous, i.e. it obeys the scaling rule

$$\mathrm{TV}(\gamma \boldsymbol{\theta}) = |\gamma| \mathrm{TV}(\boldsymbol{\theta}),$$

for any $\gamma \in \mathbb{R}$. Hence, for $\lambda > 0$, integrating equation (A1) with $\phi = \mathrm{TV}$ yields

$$\begin{aligned} 1 &= \frac{1}{Z(\lambda)} \int_{\mathbb{R}^N} e^{-\lambda \mathrm{TV}(\boldsymbol{\theta})} \, \mathrm{d}\boldsymbol{\theta} \\ &= \frac{1}{Z(\lambda)} \int_{\mathbb{R}^N} e^{-\mathrm{TV}(\lambda \boldsymbol{\theta})} \, \mathrm{d}\boldsymbol{\theta} \\ &= \frac{\lambda^{-N}}{Z(\lambda)} \int_{\mathbb{R}^N} e^{-\mathrm{TV}(\boldsymbol{\psi})} \, \mathrm{d}\boldsymbol{\psi} \\ &= \frac{\lambda^{-N} I_{\boldsymbol{\psi}}}{Z(\lambda)}, \end{aligned}$$

where $I_{\boldsymbol{\psi}}$ is the value of the integral over $\boldsymbol{\psi}$. Inserting this result into equation (A1) yields

$$\begin{aligned} p(\boldsymbol{\theta} \,|\, \lambda) &= I_{\boldsymbol{\psi}}^{-1} \lambda^N e^{-\lambda \phi(\boldsymbol{\theta})} \\ &\propto \lambda^N e^{-\lambda \mathrm{TV}(\boldsymbol{\theta})}. \end{aligned}$$

Appendix B. Determining the burn-in length

We adapted the method of Betancourt (2010) to determine when the burn-in phase is over, i.e. when the chains have converged to the posterior in a sufficient manner for practical use. In this approach, we look at the coefficients within a window of the last 100 sampled parameter values for each voxel, split these into two consecutive windows with 50 samples each, and compute the means within these two windows over all samples in the window and all chains. Hence, we obtain two means for each voxel, and thus have two 'mean images'. The means are approximately normally distributed, and the question is whether these two 'mean images' are sufficiently similar.

The voxels in the 'mean images' are modelled by a Bayesian mixture model, where the mixture is between a model that assumes the voxels have the same mean, and one where we assume they have different means. We then examine the posterior distribution of the mixture coefficient, to determine whether the chains have converged or not.

Formally, the posterior over the mixture coefficient is

$$p(\alpha|\boldsymbol{I},\boldsymbol{J}) = \frac{p(\boldsymbol{I},\boldsymbol{J}|\alpha)p(\alpha)}{\int_0^1 p(\boldsymbol{I},\boldsymbol{J}|\alpha)p(\alpha)\,\mathrm{d}\alpha}, \tag{B2}$$

where $\boldsymbol{I}$ and $\boldsymbol{J}$ are the 'mean images', and α is the mixture coefficient. The likelihood is a Gaussian mixture, and is defined as

$$p(\boldsymbol{I},\boldsymbol{J}|\alpha) = \prod_{v=1}^{V} \alpha p(I_v,J_v|\mathcal{S}) + (1-\alpha)p(I_v,J_v|\bar{\mathcal{S}}),$$

where $\mathcal{S}$ is the event that the voxels I_v and J_v have the same means, and $\bar{\mathcal{S}}$ is the event that the means are different. The prior for the mixture coefficient is uniform over $[0,1]$, such that

$$p(\alpha) = 1, \quad \forall\alpha \in [0,1].$$

The evidence that they are the same is modelled by two Gaussian distributions, evaluated over the support of the prior, with the same mean, μ, and the same (known) variance (σ^2, estimated from the data). The prior is uniform,

$$p(\mu|\mathcal{S}) = \begin{cases} \frac{1}{t}, & \text{if } 0 \leq \mu \leq t, \\ 0, & \text{otherwise,} \end{cases}$$

the joint likelihood is

$$\begin{aligned} p(I_v,J_v|\mu,\mathcal{S}) &= p(I_v|\mu,\mathcal{S})p(J_v|\mu,\mathcal{S}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(I_v-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(J_v-\mu)^2}{2\sigma^2}}, \end{aligned}$$

giving the evidence

$$\begin{aligned} p(I_v,J_v|\mathcal{S}) &= \int_{-\infty}^{\infty} p(I_v,J_v|\mu,\mathcal{S})p(\mu|\mathcal{S})\,\mathrm{d}\mu \\ &= \frac{1}{t}\int_0^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(I_v-\mu)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(J_v-\mu)^2}{2\sigma^2}}\,\mathrm{d}\mu. \end{aligned} \tag{B3}$$

The evidence that they are different is also modelled by two Gaussian distributions, evaluated over the support of the prior, but with different means, μ_I and μ_J, and (known) variances (σ_I^2 and σ_J^2, estimated from the data) where the joint prior is uniform,

$$\begin{aligned} p(\mu_I,\mu_J|\bar{\mathcal{S}}) &= p(\mu_I|\bar{\mathcal{S}})p(\mu_J|\bar{\mathcal{S}}) \\ &= \begin{cases} \frac{1}{t^2}, & \text{if } 0 \leq \mu_I \leq t \text{ and } 0 \leq \mu_J \leq t, \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

the joint likelihood is

$$\begin{aligned} p(I_v,J_v|\mu_I,\mu_J,\bar{\mathcal{S}}) &= p(I_v|\mu_I,\bar{\mathcal{S}})p(J_v|\mu_J,\bar{\mathcal{S}}) \\ &= \frac{1}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{(I_v-\mu_I)^2}{2\sigma_I^2}} \frac{1}{\sqrt{2\pi\sigma_J^2}} e^{-\frac{(J_v-\mu_J)^2}{2\sigma_J^2}}, \end{aligned}$$

giving the evidence

$$\begin{aligned} p(I_v,J_v|\bar{\mathcal{S}}) &= \int_{-\infty}^{\infty} p(I_v|\mu_I,\bar{\mathcal{S}})p(\mu_I|\bar{\mathcal{S}})\,\mathrm{d}\mu_I \int_{-\infty}^{\infty} p(J_v|\mu_J,\bar{\mathcal{S}})p(\mu_J|\bar{\mathcal{S}})\,\mathrm{d}\mu_J \\ &= \frac{1}{t\sqrt{2\pi\sigma_I^2}} \int_0^t e^{-\frac{(I_v-\mu_I)^2}{2\sigma_I^2}}\,\mathrm{d}\mu_I \cdot \frac{1}{t\sqrt{2\pi\sigma_J^2}} \int_0^t e^{-\frac{(J_v-\mu_J)^2}{2\sigma_J^2}}\,\mathrm{d}\mu_J. \end{aligned} \tag{B4}$$

The evidences, equations (B3) and (B4), can easily be computed analytically (using the error function). The posterior evidence (the denominator in equation (B2)) was computed numerically using Simpson's rule.

The chain was deemed having converged to the posterior when the mode of the posterior of the mixture coefficient was larger than 0.99. In that case, the model has found that the overwhelming majority of the voxels in the 'mean images' are better described by a single Gaussian than by two independent Gaussians.

Appendix C. The widely applicable information criterion (WAIC)

We used the criterion denoted WAIC_1 in Watanabe (2009), which is defined through the Bayesian and Gibbs training losses as

$$\begin{aligned}\mathrm{WAIC}_1 &:= \mathrm{BL}_t + 2\beta(\mathrm{GL}_t - \mathrm{BL}_t)\\ &= \mathrm{BL}_t + \frac{\beta}{N}V + o_p\left(\frac{1}{N}\right),\end{aligned}$$

where BL_t and GL_t are the Bayesian and Gibbs training losses, respectively, V is the empirical variance, o_p denotes convergence (to zero) in probability, and $\beta > 0$ is an inverse temperature parameter that was set to $\beta = 1$ in our estimations—corresponding to Bayesian estimation. The criterion is related to the Bayes generalization loss as

$$\mathbb{E}[\mathrm{BL}_g] = \mathbb{E}[\mathrm{WAIC}_1] + o\left(\frac{1}{N}\right),$$

where BL_g is the Bayes generalization loss, N is the number of data samples, and o is the little-o notation. The WAIC_1 can be estimated by

$$\mathrm{WAIC}_1 \approx \widehat{\mathrm{BL}}_t + \frac{\beta}{N}\widehat{V},$$

with

$$\begin{aligned}V &= \sum_{i=1}^{N}\mathrm{Var}\big(\log p(x_i\,|\,\theta)\big)\\ &\approx \sum_{i=1}^{N}\frac{1}{N_s-1}\sum_{s=1}^{N_s}\big(\log p(x_i\,|\,\theta^{(s)}) - \mu_i\big)^2\\ &=: \widehat{V}\end{aligned}$$

in which

$$\mu_i = \frac{1}{N_s}\sum_{s=1}^{N_s}\log p(x_i\,|\,\boldsymbol{\theta}^{(s)}),$$

the data points are $\{x_i\}_{i=1}^{N}$, and $\{\boldsymbol{\theta}^{(s)}\}_{s=1}^{N_s}$ are parameter samples from the posterior; and

$$\begin{aligned}\mathrm{BL}_t &= -\frac{1}{N}\sum_{i=1}^{N}\log\mathbb{E}_{\boldsymbol{\theta}}\big[p(x_i\,|\,\boldsymbol{\theta})\big]\\ &\approx -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{1}{N_s}\sum_{s=1}^{N_s}p(x_i\,|\,\boldsymbol{\theta}_s)\right)\\ &=: \widehat{\mathrm{BL}}_t.\end{aligned}$$

Appendix D. *In-vivo* dataset

Tissue parameter estimates, calculated with the three estimation models ML, $\mathrm{B}_{\mathrm{unif}}$, and B_{TV} are presented in figure D1. Each row in figure D1 presents the estimated parameters for a different patient the *in-vivo* dataset (patients 1 to 8).

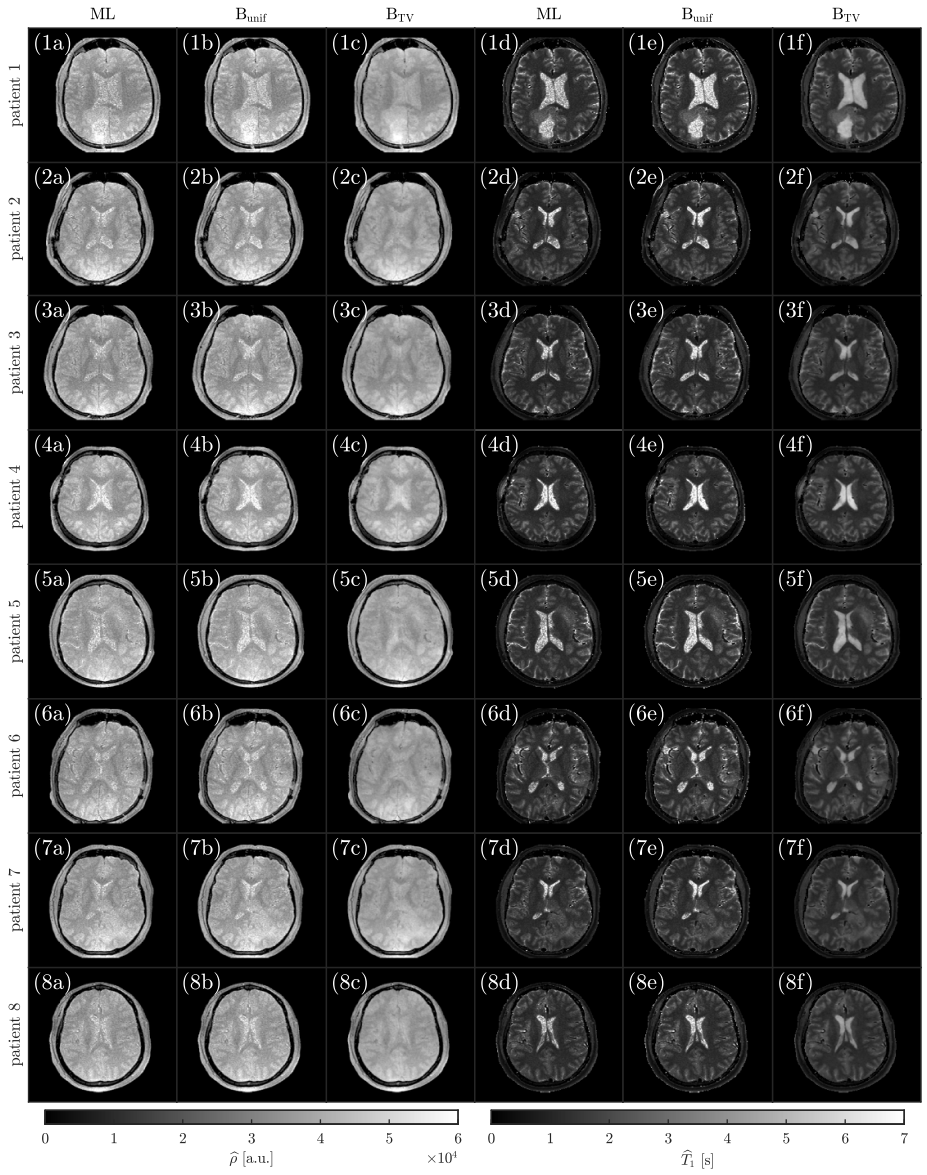

Figure D1. Tissue parameter estimation results of ρ (col (a)–(c)) and T_1 (col (d)–(f)) for each of the eight patients in the *in-vivo* dataset (row 1–8).

ORCID iDs

Tommy Löfstedt https://orcid.org/0000-0001-7119-7646
Max Hellström https://orcid.org/0000-0002-0200-6567
Mikael Bylund https://orcid.org/0000-0001-7539-2262
Anders Garpebring https://orcid.org/0000-0002-0532-232X

References

Bartos M, Rajmic P, Šorel M, Mangová M, Keunen O and Jiřík R 2019 Spatially regularized estimation of the tissue homogeneity model parameters in DCE-MRI using proximal minimization *Magn. Reson. Med.* **82** 2257–72

Baselice F, Ferraioli G and Shabou A 2010 Field map reconstruction in magnetic resonance imaging using Bayesian estimation *Sensors* **10** 266–79

Baselice F, Ferraioli G and Pascazio V 2016 A Bayesian approach for relaxation times estimation in MRI *Magn. Reson. Imaging* **34** 312–25

Betancourt M 2010 A bayesian approach to histogram comparison Massachusetts Institute of Technology, Cambridge, MA, USA (arXiv:1009.5604) [physics.data-an]

Bitsch A, Kuhlmann T, Stadelmann C, Lassmann H, Lucchinetti C and Brück W 2001 A longitudinal MRI study of histopathologically defined hypointense multiple sclerosis lesions *Ann. Neurol.* **49** 793–6
Cocosco C A, Kollokian V, Kwan R K S and Evans A C 1997 Brain web: Online interface to a 3D MRI simulated brain database *Neuroimage* **5** 425
Collins D L, Zijdenbos A P, Kollokian V, Sied J G, Kabani N J, Holmes C J and Evans A C 1998 Design and construction of a realistic digital brain phantom *IEEE Trans. Med. Imaging* **17** 463–8
De Pasquale F, Sebastiani G, Egger E, Guidoni L, Luciani A M, Marzola P, Manfredi R, Pacilio M, Piermattei A, Viti V and Barone P 2000 Bayesian estimation of relaxation times T_1 in MR images of irradiated Fricke-agarose gels *Magn. Reson. Imaging* **18** 721–31
Duane S, Kennedy A D, Pendleton B J and Roweth D 1987 Hybrid monte carlo *Phys. Lett.* **195** 216–22
Freiman M, Perez-Rossello J M, Callahan M J, Voss S D, Ecklund K, Mulkern R V and Warfield S K 2013 Reliable estimation of incoherent motion parametric maps from diffusion-weighted MRI using fusion bootstrap moves *Med. Image Anal.* **17** 325–36
Fütterer J J 2016 Multiparametric mri in the detection of clinically significant prostate cancer *Korean J. Radiol.* **18** 597–606
Garpebring A, Brynolfsson P, Yu J, Wirestam R, Johansson A, Asklund T and Karlsson M 2013 Uncertainty estimation in dynamic contrast-enhanced MRI *Magn. Reson. Med.* **69** 992–1002
Gelman A, Hwang J and Vehtari A 2014 Understanding predictive information criteria for Bayesian models *Stat. Comput.* **24** 997–1016
Gelman A, Carlin J B, Stern H S, Dunson D B, Vehtari A and Rubin D B 2014 *Bayesian Data Analysis* (Boca Raton, FL: CRC Press)
Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-6 721–41
Glad I K and Sebastiani G 1995 A Bayesian approach to synthetic magnetic resonance imaging *Biometrika* **82** 237–50
Goodman J and Weare J 2010 Ensemble samplers with affine invariance *Commun. Appl. Math. Comput. Sci.* **5** 65–80
Gudbjartsson H and Patz S 1995 The Rician distribution of noisy MRI data *Magn. Reson. Med.* **34** 910–14
Ji S, Xue Y and Carin L 2008 Bayesian compressive sensing *IEEE Trans. Signal Process.* **56** 2346–56
Kelm B M, Menze B H, Nix O, Zechmann C M and Hamprecht F A 2009 Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge *IEEE Trans. Med. Imaging* **28** 1534–47
Kumar D, Nguyen T D, Gauthier S A and Raj A 2012 Bayesian algorithm using spatial priors for multiexponential T 2 relaxometry from multiecho spin echo MRI *Magn. Reson. Med.* **68** 1536–43
Kwan R K S, Evans A C and Pike B 1999 MRI simulation-based evaluation of image-processing and classification methods *IEEE Trans. Med. Imaging* **18** 1085–97
Kwan R K S, Evans A C and Pike G B 1996 An extensible MRI simulator for post-processing evaluation *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* pp 135–40
Nürnberger L, Gracien Re-M, Hok P, Hof S-M, Rüb U, Steinmetz H, Hilker Rudiger, Klein J C, Deichmann R and S 2017 Longitudinal changes of cortical microstructure in Parkinson's disease assessed with T1 relaxometry *Neuroimage Clin.* **13** 405–14
Orton M R, Collins D J, Koh D-M and Leach M O 2014 Improved intravoxel incoherent motion analysis of diffusion weighted imaging by data driven Bayesian modeling *Magn. Reson. Med.* **71** 411–20
Pham T T, Liney G P, Wong K and Barton M B 2017 Functional MRI for quantitative treatment response prediction in locally advanced rectal cancer *Br. J. Radiol.* **90** 20151078
Raj A, Pandya S, Shen X, LoCastro E, Nguyen T D and Gauthier S A 2014 Multi-compartment T2 relaxometry using a spatially constrained multi-Gaussian model *PLoS One* **9** 1–13
Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
Schabel M C and Parker D L 2008 Uncertainty and bias in contrast concentration measurements using spoiled gradient echo pulse sequences *Phys. Med. Biol.* **53** 2345–73
Schmid V J, Whitcher B, Padhani A R, Jane Taylor N and Yang G-Z 2006 Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging *IEEE Trans. Med. Imaging* **25** 1627–36
Sjölund J, Eklund A, Özarslan E, Herberthson M, Baankestad M and Knutsson H 2018 Bayesian uncertainty quantification in linear models for diffusion MRI *Neuroimage* **175** 272–85
Snoek J, Larochelle H and Adams R P 2012 Practical bayesian optimization of machine learning algorithms *Advances in Neural Information Processing Systems 25* eds Pereira F, Burges C J C, Bottou L and Weinberger K Q (Curran Associates, Inc.) pp 2951–9
Sommer J C and Schmid V J 2014 Spatial two-tissue compartment model for dynamic contrast-enhanced magnetic resonance imaging *J. R. Stat. Soc.: Ser. C Appl. Stat.* **63** 695–713
Tofts P S (ed) 2003 *Quantitative MRI of the Brain: Measuring Changes Caused by Disease* (Chichester: Wiley)
Tozer D J, Davies G R, Altmann D R, Miller D H and Tofts P S 2005 Correlation of apparent myelin measures obtained in multiple sclerosis patients and controls from magnetization transfer and multicompartmental T 2 analysis *Magn. Reson. Med.* **53** 1415–22
Wang H and Cao Y 2012 Spatially regularized T1 estimation from variable flip angles MRI *Med. Phys.* **39** 4139–48
Watanabe S 2009 *Algebraic Geometry and Statistical Learning Theory* (Cambridge: Cambridge University Press)
Watanabe S 2010 Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory *J. Mach. Learn. Res.* **11** 3571–94
While P T 2017 A comparative simulation study of bayesian fitting approaches to intravoxel incoherent motion modeling in diffusion-weighted MRI *Magn. Reson. Med.* **78** 2373–87
Wood J C 2014 Guidelines for quantifying iron overload *Hematology* **2014** 210–15