ROBUST T_2 RELAXOMETRY WITH HAMILTONIAN MCMC FOR MYELIN WATER FRACTION ESTIMATION

Thomas Yu[⋆†] *Marco Pizzolato*[⋆†] *Erick Jorge Canales-Rodríguez* [‡] *Jean-Philippe Thiran*[⋆‡◇]

[⋆]Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[‡]Radiology Department, CHUV, Lausanne, Switzerland, [◇]UNIL, Lausanne, Switzerland
[†]*Authors contributed equally*

ABSTRACT

We present a voxel-wise Bayesian multi-compartment T_2 relaxometry fitting method based on Hamiltonian Markov Chain Monte Carlo (HMCMC) sampling. The T_2 spectrum is modeled as a mixture of truncated Gaussian components, which involves the estimation of parameters in a completely data-driven and voxel-based fashion, i.e. without fixing any parameters or imposing spatial regularization. We estimate each parameter as the expectation of the corresponding marginal distribution drawn from the joint posterior obtained with Hamiltonian sampling. We validate our scheme on synthetic and *ex vivo* data for which histology is available. We show that the proposed method enables a more robust parameter estimation than a state of the art point estimate based on differential evolution. Moreover, the proposed HMCMC-based myelin water fraction calculation reveals high spatial correlation with the histological counterpart.

***Index Terms*—** truncated-gaussian, myelin, validation, variable-projection, MRI, MCMC

1. INTRODUCTION

Multi Echo T_2 experiments [1] in Magnetic Resonance Imaging (MRI) allow measuring a signal, for each voxel, that is the superposition of the contributions of different microstructural water compartments, such as myelin, the combination of intra and extra axonal space (IE), and cerebrospinal fluid (CSF) [2]. The signal amplitude is measured at varying echo times, TE, producing a characteristic decay due to the different transverse relaxation times, T_2, of the different water pools. Typically, to disentangle the different compartment contributions – such as the myelin water fraction (MWF) – the signal is represented as a sum of decaying exponentials weighted by the corresponding water volume fractions of each compartment [2, 3]. Accurate and robust disentangling is important, for example, because an accurate measure of MWF can be used to assess neurological damage and disease [3]. However, this model assumes that each compartment has a single, well defined T_2, an assumption made to simplify the modeling so as not to consider the heterogeneity of the tissue. We propose to characterize T_2 decay with a multi-compartment model that accounts for a distribution of T_2 values within each compartment through truncated Gaussian distributions. To obtain the tissue parameters, we use a Hamiltonian Markov Chain Monte Carlo (HMCMC) sampler [4] called the No U Turn sampler (NUTS) [5] within a Bayesian framework. This, allows us to easily incorporate prior knowledge and constraints concerning the tissue parameters as well as to account for the noise in the data. Further, instead of simply obtaining best fit parameters, we obtain probability distributions for each parameter, allowing us to quantitatively examine the level of stability and uncertainty in the parameter estimates. Using HMCMC rather than conventional samplers, such as Metropolis [6] or Gibbs [7], allows us to use fewer samples with greater inferential power due to the use of Hamiltonian dynamics. There are a few prior attempts to use MCMC for parameter estimation in MRI, particularly diffusion MRI [8, 9, 10], which are focused on modifications to Metropolis or Gibbs samplers. Our choice is motivated by the fact that as the number of parameters in a model or the complexity of the model increases, point estimates like maximum likelihood or maximum a posteriori can become increasingly ill-suited for inference due to, for example, the phenomenon of the concentration of measure [4]. This may in part contribute to the instability and inaccuracy of estimating all parameters of complex relaxometry models in traditional approaches [11]. We show that in synthetic and in real data, estimates coming from taking the means of the marginal posterior distributions provided by HMCMC are more robust than a state of the art point estimate. Furthermore, a comparison of the MWF calculated with our method with respect to histological myelin measurements reveals a high spatial correlation.

2. MODELING

A recent approach [12] uses a gamma distribution for each compartment to model the T_2 distribution and fits parameters

[†]Corresponding authors: thomas.yu@epfl.ch, marco.pizzolato@epfl.ch

through a least squares fitting using variable projection; however, all except one of the means and standard deviations of the T_2 distributions are fixed to values coming from the literature in order to avoid instability in the fitting. In another recent approach [13], authors use a Gaussian distribution to characterize each compartment and also estimate all volume and T_2 parameters; the authors resort to a multi-voxel least squares approach with spatial regularization in order to obtain a stable fitting. Since physically there are clear lower and upper bounds on the T_2 due to non-negativity and the chemical properties of water, we propose to use a truncated Gaussian distribution to describe the myelin and IE compartments; we note that [13] shows that a variety of distributions such as the normal and gamma distributions produce the same decay curve given a fixed mean and standard deviation. Due to the homogeneity of CSF, we expect it to have a very narrow T_2 distribution which we represent with a delta function [3]. Hence, we model the voxel signal at echo time TE_i as

$$\begin{aligned} S(TE_i) =&(v_{mye}\int_0^\infty \exp(-\frac{TE_i}{T_2})TG\left(T_2,\mu_{mye},\sigma_{mye}\right)\mathrm{d}T_2 \\ &+ v_{IE}\int_0^\infty \exp(-\frac{TE_i}{T_2})TG(T_2,\mu_{IE},\sigma_{IE})\mathrm{d}T_2 \\ &+ v_{csf}\exp(-\frac{TE_i}{T_2^{csf}}))M_0 \end{aligned} \tag{1}$$

where TG is the truncated Gaussian distribution, M_0 is the proton density, v_i are the water fractions , μ_i, σ_i are the means, and standard deviations of the T_2 distribution of the ith compartment, and T_2^{csf} is the decay constant of CSF. Thus, the parameters of the model to be estimated are

$$\mathbf{x} = (v_{mye}, v_{IE}, v_{csf}, \mu_{mye}, \sigma_{mye}, \mu_{IE}, \sigma_{IE}, T_2^{csf}, M_0). \tag{2}$$

3. PARAMETER FITTING

For each voxel, we have a signal acquired at $N = 32$ equally spaced echo times, from $10ms$ to $320ms$. We use a voxel-wise Bayesian approach where, denoting the parameters of the model as $\mathbf{x}$ and the signal data of a voxel as $\mathbf{s} = \{s_1, \ldots, s_N\}$, we have

$$p(\mathbf{x}|\mathbf{s}) \propto p(\mathbf{s}|\mathbf{x})p(\mathbf{x}) \tag{3}$$

where we omit a normalization constant depending on $\mathbf{s}$. Given a set of parameters $\mathbf{x}$, we model $\mathbf{s}_i$ as arising from the model evaluated at $(\mathbf{x}, TE_i)$ with additive white Gaussian noise such that

$$\mathbf{s}_i \sim \mathcal{N}(Model(TE_i, \mathbf{x}), \sigma^2) \tag{4}$$

where σ is the noise standard deviation estimated [14] from the image in real data, or set in synthetic data. Due to the assumption of independence, we have that the likelihood function satisfies

$$p(\mathbf{s}|\mathbf{x}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{(s_i - Model(TE_i, \mathbf{x}))^2}{2\sigma^2}). \tag{5}$$

Since the volume fractions must sum to one, we use an uninformative, three dimensional, symmetric Dirichlet prior

$$(v_{mye}, v_{oth}, v_{csf}) \sim \mathbf{Dir}(1.0, 1.0, 1.0). \tag{6}$$

For the other parameters, we use uniformly distributed priors supported on intervals which correspond to reasonable bounds for the parameters taken from literature values [3] while considering *ex vivo* data:

$$\begin{aligned} &T_2^{CSF} \sim \mathcal{U}([200ms, 500ms]),\ M_0 \sim \mathcal{U}([0.95, 1.05]) \\ &\mu_{mye} \sim \mathcal{U}([1ms, 50ms]),\ \sigma_{mye} \sim \mathcal{U}([1ms, 20ms]) \\ &\mu_{IE} \sim \mathcal{U}([50ms, 200ms]),\ \sigma_{IE} \sim \mathcal{U}([1ms, 60ms]). \end{aligned}$$

Note that we normalize the signal using a proton density estimated with a previous NNLS fit [2]; M_0 is thus a factor which corrects for mis-estimation of this parameter in the model. Given the likelihood and the priors, HMCMC approximates the whole posterior and marginal posterior distributions through sampling of the support of these distributions by following the trajectories of points under the following dynamical system:

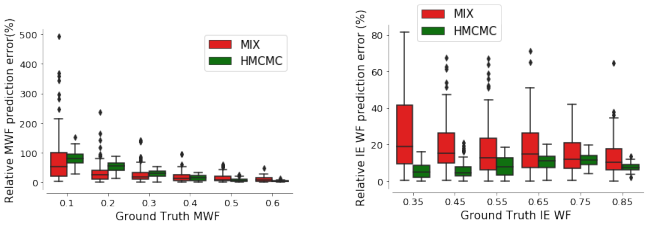
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\partial H}{\partial l} \tag{7}$$

$$\frac{\mathrm{d}l}{\mathrm{d}t} = \frac{-\partial H}{\partial x} \tag{8}$$

where H is a function constructed from the likelihood and the priors, x represents the parameters as variables, and l is a set of auxiliary variables which are not used in terms of the sampling. In practice, this system is integrated numerically using a leapfrog scheme [5]. We run the sampler for 10,000 samples, 5000 samples for burn-in to help the sampler reach convergence, and 5000 samples used for inference. We then obtain best fit parameters through an expectation:

$$\mathbf{x}^* = \int \mathbf{x} p(\mathbf{x}|\mathbf{s})\mathrm{d}\mathbf{x}, \tag{9}$$

which we calculate through the standard sample approximation. We note that as the number of parameters increases, this expectation rarely is similar to the maximum a posteriori estimate or other point estimates obtained from traditional fitting methods. A readable but comprehensive overview of HMCMC as well the pitfalls of using traditional methods is available in [4]. To initialize the HMCMC sampler, we use Automatic Differentiation Variational Inference(ADVI) [15], where instead of sampling the posterior, the parameters of a tractable distribution are fit to match the posterior. We use



(a) Myelin (b) IE

Fig. 1: Relative errors from ground truth for MWF and IEWF calculated using HMCMC and MIX. Note the significantly higher variance of errors in MIX as compared to HMCMC.

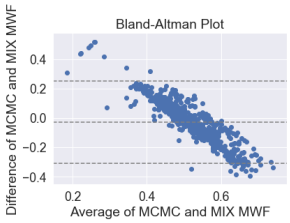


Fig. 2: Bland-Altman Plot showing the clear difference between the methods.

	In WM	In Tot.
HMCMC	0.45	0.79
MIX	0.08	0.67
Standard [20]	0.2	0.72

Fig. 3: Pearson Correlation to Histology. Note that HMCMC significantly outperforms the other methods in white matter.

the open source probabilistic programming library PyMC3 [16] in Python to carry out the HMCMC fitting. To compare, we also fit parameters using an in-house implementation of variable projection [17] called MIX [18] which solves the standard least squares minimization problem using an initial guess generated from solving reduced problems using differential evolution and convex optimization, resulting in a point estimate. We used the same bounds on the parameters in MIX as in the HMCMC fitting.

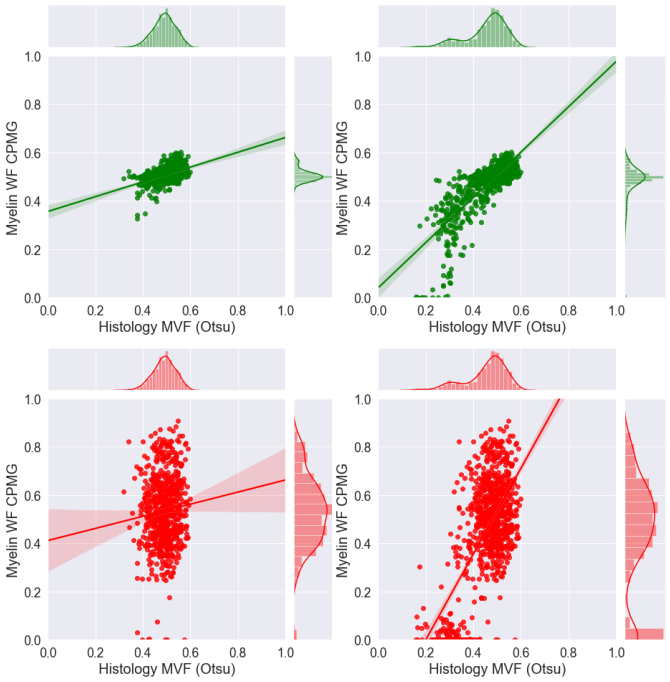
4. EXPERIMENTS AND RESULTS

We applied HMCMC and MIX to both synthetic and *ex vivo* spinal cord data with histology [19]. The estimated SNR of the real data acquisition was 500; hence, we generated the synthetic data from the model discussed in the previous section with additive zero mean Gaussian noise in order to match this SNR. We fix the T_2 values of the myelin, IE, and CSF compartments, while varying myelin and IE volume fractions. For each volume fraction set, we generate 100 realizations of zero mean Gaussian noise and add it to the synthetic signal. In particular, the simulated parameters are $T_2^{CSF} = 0.3s, M_0 = 1.0, \mu_{mye} = 0.02s, \sigma_{mye} = 0.005s, \mu_{IE} = 0.1s$, $\sigma_{IE} = 0.01s, v_{mye} \in \{0.3, 0.4, 0.5, 0.6\}, v_{CSF} = 0.05, v_{IE} \in \{0.65, 0.55, 0.45, 0.35\}$. In Fig. 1, for each ground truth myelin and IE volume, we took the absolute error between the predicted volumes and the ground truth, divided by the ground truth, and plotted the corresponding boxplot which accounts for 100 noisy repetitions. We can see that the variance of the relative errors of MIX estimates is much larger than that of HMCMC for myelin and IE over all ground truth volumes. In addition, MIX had more outliers with large error values. Further, in general, the median and mean relative errors of HMCMC is equal to or lower than that of MIX.

Regarding the real data, in Fig. 4 we plot the volumes calculated from histology versus the predicted volumes from the different fitting procedures voxelwise. We calculate the Pearson correlation coefficients for each method and also show a Bland-Altman plot in Fig. 2 and Fig. 3. The table also reports the correlation coefficients from a standard parameter fitting previously done on the same dataset [20] with different modeling. In Fig 5 we show some parameter maps from HMCMC. Moreover, in Fig. 6 we show the marginal distributions of the volumes from HMCMC for two representative voxels selected randomly in white and gray matter. We use cocor [21] to do a statistical comparison of the correlation coefficients of MIX and HMCMC and find that the null hypothesis of the coefficients being the same can be rejected at a 95 percent confidence level using a variety of statistical tests. Further, we can see from the Bland-Altman plot that the two methods clearly differ significantly, indicating a fundamental difference in prediction quality.

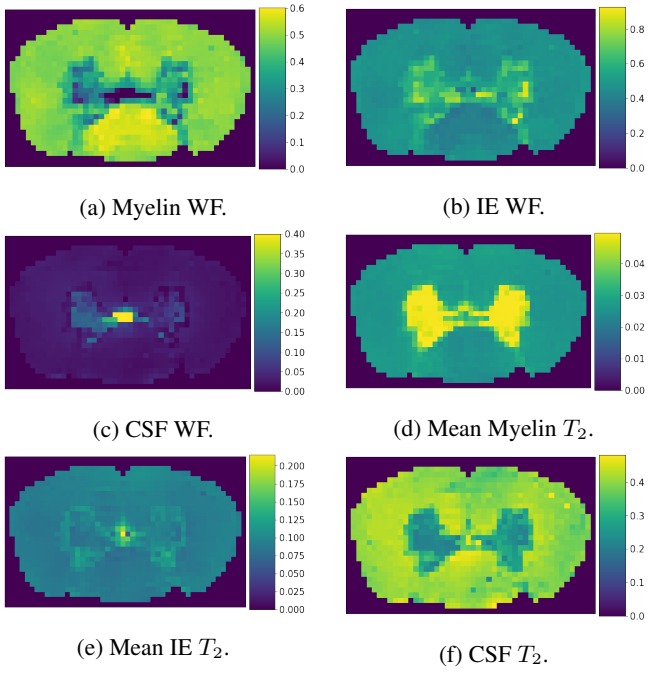
5. DISCUSSION

From the results, we can see that using a voxel-wise approach, even at a fairly high SNR, volume point estimates are much less robust to noise and less accurate than distributional estimates, in both synthetic and real data. HMCMC provides robust parameter estimates, allowing for better performance on synthetic data, Fig. 1, high spatial correlation with histology for myelin water fraction estimates, Fig. 4b, and smooth parameter maps, Fig. 5. As for future perspectives, we note that in our model we do not account for imperfect rephasing of nuclear spins in the MRI acquisition [22] which leads to the signal decay not being perfectly exponential. This can have a significant impact on the fitting, particularly if the signal at low TE are affected. This phenomenon might in part explain the performance of MIX on real data, since the global optimum of the squared differences might change due to the model inconsistency; we conjecture that our method is more robust to model inconsistency since it computes posterior distributions and uses averaging to obtain the estimates. Further, we omitted discussing the statistical properties of HMCMC such as convergence and optimal sample numbers as well as tunable parameters of the NUTS sampler and statistical validation of sampling. In addition, while HMCMC provides more robust estimates, it requires much more computational time compared to MIX or other point estimators. However, using parallel CPU processing on clusters and GPU based MCMC sampling, this problem can be mitigated. We em-



(a) White matter (b) White and Gray matter

Fig. 4: Voxelwise plots of predicted Myelin Water Fraction (MWF) for HMCMC (Green) and MIX (Red) vs. Histological Myelin Volume Fraction (MVF). HMCMC provides a better spatial correlation to histology, more precise MWF estimates, and a plausible linear trend with respect to histology.



(a) Myelin WF. (b) IE WF.

(c) CSF WF. (d) Mean Myelin T_2.

(e) Mean IE T_2. (f) CSF T_2.

Fig. 5: Water Fraction and T_2 Maps obtained from HMCMC.

phasize that the parameters of Dirichlet prior were chosen to be weakly informative i.e. to encode the sum constraint and to not favor low or high volumes for each compartment; tuning of these parameters based on prior estimates of the ranges of water fractions seen for each compartment could lead to better results. Further, while we use an ADVI initialization, there are many other initialization strategies one can use; for instance, one could also run HMCMC on a simpler model and use those results to initialize the complex model used in this work. Finally, we note that modifying the Bayesian approach to incorporate spatial regularization and/or a multi-voxel approach and using HMCMC could also stabilize the fitting even further.

6. CONCLUSION

We proposed a T_2 relaxometry model which takes into account heterogeneity of T_2 decay within compartments by modeling distributions of T_2 for each compartment using truncated Gaussian distributions. We proposed to fit this complex model's parameters to synthetic and real data using a voxel-wise Bayesian approach and Hamiltonian MCMC sampling. To the best of our knowledge, it is the first time that such a framework has been used for Multi Echo T_2 relaxometry, enabling the fitting of all the parameters of the model without resorting to spatial regularization or multi-voxel methods. We validated our method on synthetic and real data and compared it to a state of the art point estimate, finding that the use of HMCMC yields more robust and accurate estimates of water fractions, with significantly better correlations to histology.

7. ACKNOWLEDGMENTS

Thomas Yu is supported by the European Union's Horizon 2020 program under the Marie Sklodowska-Curie project TRABIT (agreement No 765148). Marco Pizzolato is supported by the SNSF under Sinergia grant number CRSII5_170873.

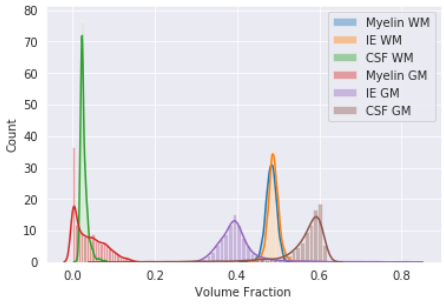


Fig. 6: Sample volume fraction distributions from HMCMC in a WM and GM voxel

8. REFERENCES

[1] Saul Meiboom and David Gill, "Modified spin-echo method for measuring nuclear relaxation times," *Review of scientific instruments*, vol. 29, no. 8, pp. 688–691, 1958.

[2] Whittall and MacKay, "Quantitative interpretation of nmr relaxation data.," *Journal of Magnetic Resonance(1969)*, vol. 84(1), pp. 134–152, 1989.

[3] MacKay et al, "Magnetic resonance of myelin water: an in vivo marker for myelin," *Brain Plasticity*, vol. 2(1), pp. 71–91, 2016.

[4] Betancourt et al, "A conceptual introduction to hamiltonian monte carlo.," *arXiv preprint*, vol. arXiv:1701.02434, 2017.

[5] Hoffman et al, "The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo.," *Journal of Machine Learning Research*, vol. 15(1), pp. 1593–1623, 2014.

[6] Metropolis et al, "Equation of state calculations by fast computing machines.," *The journal of chemical physics*, vol. 21(6), pp. 1087–1092, 1943.

[7] Geman et al, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on pattern analysis and machine intelligence*, vol. (6), pp. 721–741, 1984.

[8] Melie-Garcia et al, "A bayesian framework to identify principal intravoxel diffusion profiles based on diffusion-weighted mr imaging.," *Neuroimage*, vol. Vol 42 Issue 2, 2008.

[9] Harms and Roebroeck, "Robust and fast monte carlo markov chain sampling of diffusion mri microstructure models.," *bioRxiv*, 2018.

[10] Clayden et al, "Microstructural parameter estimation in vivo using diffusion mri and structured prior information.," *Magnetic resonance in medicine*, vol. 75(4), pp. 1787–1796, 2016.

[11] Layton et al, "Modelling and estimation of multicomponent t2 distributions.," *IEEE TMI*, vol. vol. 32, no. 8, pp. 1423–1434, 2013.

[12] Chatterjee et al, "Multi-compartment model of brain tissues from t2 relaxometry mri using gamma distribution," in *15th International Symposium on Biomedical Imaging(ISBI)*. IEEE, 2018, pp. 141–144.

[13] Raj et al, "Multi-compartment t2 relaxometry using a spatially constrained multi-gaussian model.," *PLoS One*, vol. 9(6), pp. e98391, 2014.

[14] Veraart et al, "Denoising of diffusion mri using random matrix theory," *NeuroImage*, vol. 142, pp. 394–406, 2016.

[15] Kucukelbir et al, "Automatic differentiation variational inference," *Journal of Machine Learning Research*, vol. 18(1), pp. 430–474, 2017.

[16] Salvatier et al, "Probabilistic programming in python using pymc3," *PeerJ Computer Science 2*, vol. e55, 2016.

[17] Golub and Pereyra, "Separable nonlinear least squares: the variable projection method and its applications.," *Inverse problems*, vol. 19(2), pp. R1, 2003.

[18] Farooq et al, "Microstructure imaging of crossing (mix) white matter fibers from diffusion mri.," *Scientific reports*, vol. 6, pp. 38927, 2016.

[19] Cohen-Adad et al, "White matter microscopy database.," *OSF, Web*, 2018.

[20] Vuong et al, "On the precision of myelin imaging: Characterizing ex vivo dog spinal cord.," 2017, p. 3760.

[21] Diedenhofen and Musch, "A comprehensive solution for the statistical comparison of correlations.," *PLoS ONE*, vol. 10(4), pp. doi:10.1371/journal.pone.0121945, 2015.

[22] Hennig et al, "Calculation of flip angles for echo trains with predefined amplitudes with the extended phase graph (epg)-algorithm: Principles and applications to hyperecho and traps sequences.," *MRM*, vol. vol. 51, no. 1, pp. 68–80, 2004.