32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Bayesian optimization of perfusion and transit time estimation in PASL-MRI

Nuno Santos, João Sanches and Patrícia Figueiredo

***Abstract*— Pulsed Arterial Spin Labeling (PASL) techniques potentially allow the absolute, non-invasive quantification of brain perfusion and arterial transit time. This can be achieved by fitting a kinetic model to the data acquired at a number of inversion time points (*TI*). The intrinsically low SNR of PASL data, together with the uncertainty in the model parameters, can hinder the estimation of the parameters of interest. Here, a two-compartment kinetic model is used to estimate perfusion and transit time, based on a *Maximum a Posteriori* (MAP) criterion. A priori information concerning the physiological variation of the multiple model parameters is used to guide the solution. Monte Carlo simulations are performed to compare the accuracy of our proposed Bayesian estimation method with a conventional Least Squares (LS) approach, using four different sets of *TI* points. Each set is obtained either with a uniform distribution or an optimal sampling strategy designed based on the same MAP criterion. We show that the estimation errors are minimized when our proposed Bayesian estimation method is employed in combination with an optimal set of sampling points. In conclusion, our results indicate that PASL perfusion and transit time measurements would benefit from a Bayesian approach for the optimization of both the sampling strategy and the estimation algorithm, whereby prior information on the parameters is used.**

I. INTRODUCTION

Perfusion measures the rate at which nutrients are delivered by the blood to the tissues in the capillary bed and its accurate measurement is important in the diagnosis and monitoring of different pathological conditions. Pulsed Arterial Spin Labeling Magnetic Resonance Imaging (PASL-MRI) techniques offer a non-invasive way of measuring perfusion, by magnetically labeling the water protons in the arterial blood through magnetization inversion and then measuring the magnetization of the tissues after a certain period of time, the inversion time (*TI*). The magnetization difference ΔM between a labeled image and a control image, as a function of *TI*, can be described by a general kinetic model [1]. Here, we use an extension of this single-compartment model to a two-compartment model, by accounting for the contributions of labeled water molecules in the capillary blood, H_{blood}, as well as in the brain tissue, H_{tissue} [2]

$$\begin{aligned}\Delta M(\boldsymbol{t},\boldsymbol{\theta}) &= \tfrac{2\alpha f}{\lambda}(H_{tissue}(\boldsymbol{t},\boldsymbol{\theta}) \\ &\quad +H_{blood}(\boldsymbol{t},\boldsymbol{\theta})e^{-TED_2^{(*)}}) \ ,\end{aligned} \tag{1}$$

where $\boldsymbol{t} = [t_1, t_2, ..., t_N]$ is the vector of *TI*, $\boldsymbol{\theta} = [f, \Delta t, \tau, \tau_{ex}, T_{1t}, T_{1b}]$ is the vector of parameters: f is the perfusion, Δt is the arterial transit time (ATT), τ is the label bolus time width, τ_{ex} is the brain-blood water exchange time, T_{1t} is the tissue longitudinal relaxation time, T_{1b} is the blood longitudinal relaxation time, α is the labeling efficiency, λ is the brain-blood partition coefficient of water, and

$$D_2^{(*)} = \frac{1}{T_{2b}^{(*)}} - \frac{1}{T_{2t}^{(*)}} \ , \tag{2}$$

with $T_{2b}^{(*)}$ and $T_{2t}^{(*)}$ being the blood and tissue transverse relaxation times and TE the acquisition echo time. The ΔM curve as a function of *TI* is illustrated in Fig. 1. In principle, if the values of the other model parameters are available, then the acquisition of data at a single *TI* point is sufficient for the estimation of perfusion. However, there is considerable uncertainty regarding the values of various model parameters, particularly in respect to the arterial transit time and in some pathological conditions. Therefore, in order to correctly estimate f, Δt should also be estimated by fitting the PASL model to ΔM data sampled at multiple *TI* points [3]. On the other hand, the acquisition of more sampling points can lead to undesirably long scanning times. This is especially critical because PASL measurements require substantial signal averaging as a consequence of their low Signal to Noise Ratio (SNR). Therefore, a compromise between the number of *TI* points and scanning time has to be made. Previous studies have shown that the distribution of the *TI* sampling points along time has a strong effect on the accuracy of the estimation of the parameters. Optimal

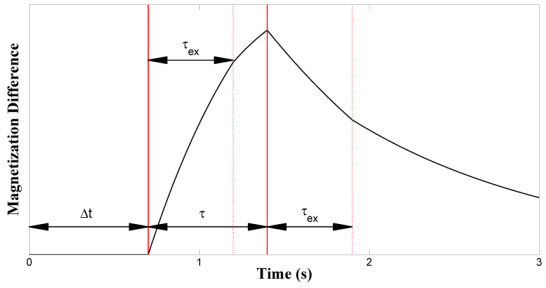


Fig. 1. PASL signal ΔM as a function of the inversion time TI, according to the two compartment kinetic PASL model.

Nuno Santos is with Siemens SA, Healthcare Sector and with Instituto Superior Técnico, Lisboa, Portugal. njgsantos@gmail.com

João Sanches and Patrícia Figueiredo are with Institute for Systems and Robotics / Instituto Superior Técnico, Lisboa, Portugal. jmrs@ist.utl.pt, patricia.figueiredo@ist.utl.pt

This work was partially supported by Fundação para a Ciência e a Tecnologia (ISR/IST pluriannual funding) through the POS Conhecimento Program that includes FEDER funds, and by the PTDC/SAU-BEB/65977/2006 Project.

sampling strategies were designed based on the Fisher information matrix criterion , in order to minimize scanning time while maximizing estimation accuracy [4]. A Bayesian extension of this criterion has also been proposed to take into account the variability of multiple model parameters [5].

In this paper, we implement a Bayesian algorithm based on the *Maximum a posteriori* (MAP) criterion for the estimation of perfusion, f, and transit time, Δt, from PASL data. A priori assumptions about the physiological variation of the model parameters are incorporated in the framework to better guide the algorithm to a more accurate solution. We then test whether the proposed Bayesian estimation approach, in combination with an optimal sampling strategy, provides the most accurate measurements of perfusion and ATT, relative to the standard Least Squares estimation method and uniformly distributed sampling points.

II. PROBLEM FORMULATION

Let $\boldsymbol{Y} = [y_1, y_2, ..., y_N]$ be a set of observations

$$y_i = \Delta M(t_i, \boldsymbol{\theta}) + \eta_i \ , \ i = 1...N \ \ , \tag{3}$$

where $\eta_i \sim N(0, \sigma_y^2)$ is assumed to be Additive White Noise (AWGN) and $\Delta M(t_i, \boldsymbol{\theta})$ is the magnetization difference predicted by the PASL model (1) at time *i* (both α, λ and $D_2^{(*)}$ are assumed constant). Because this model is highly non-linear, it is necessary to use appropriate techniques to solve this ill-posed problem. A common approach to estimate the parameter vector $\boldsymbol{\theta}$ is the Least Squares (LS) method, formulated as the following optimization task

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}) \ \ , \tag{4}$$

where $E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta})$ is the following energy function

$$E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}) = \|\boldsymbol{Y} - \Delta M(\boldsymbol{t}, \boldsymbol{\theta})\|_2^2 \ \ . \tag{5}$$

In this work, a second method is also considered. A Bayesian framework is proposed to estimate the parameters using the MAP criterion which is equivalent to minimize the following energy function

$$E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}) = -\log\left[p(\boldsymbol{Y}|\boldsymbol{t}, \boldsymbol{\theta}) p(\boldsymbol{\theta})\right] \ \ , \tag{6}$$

where $p(\boldsymbol{Y}|\boldsymbol{t}, \boldsymbol{\theta})$ represents the acquisition process (the observations are assumed to be statistically independent) and $p(\boldsymbol{\theta})$ models the a priori knowledge of the parameters to be estimated. All the parameters of $\boldsymbol{\theta}$ are assumed to be independent and Gaussian distributed around a mean value with different standard deviations according to the level of uncertainty associated with each one. Then, the distribution of $\boldsymbol{\theta}$ is a multivariate Normal distribution $N(\overline{\boldsymbol{\theta}}, \boldsymbol{C})$ with a diagonal covariance matrix $\boldsymbol{C} = diag(\{\sigma_1^2, \sigma_2^2, ..., \sigma_P^2\})$ (where P is the number of parameters), and the energy function (6) may be written as follows

$$\begin{aligned} E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}) \ \ = & \underbrace{\frac{1}{2}\|\boldsymbol{Y} - \Delta M(\boldsymbol{t}, \boldsymbol{\theta})\|_2^2}_{\text{Data Fidelity term}} \\ & + \underbrace{\frac{1}{2}\boldsymbol{\sigma}_y^2 (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})^T \boldsymbol{C}^{-1} (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})}_{\text{Prior term}} \ \ . \end{aligned} \tag{7}$$

Once again, to find the optimal parameters with the Bayesian approach we need to determine the $\boldsymbol{\theta}$ that minimize the energy function (equivalently given by (4)).

In both estimation methods (LS and Bayesian), the optimization is accomplished by using the Levenberg-Marquardt (LM) algorithm [6]

$$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n + \boldsymbol{D}^{-1} \, . \, \nabla_{\boldsymbol{\theta}} E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}^n) \ \ , \tag{8}$$

where $\nabla_{\boldsymbol{\theta}} E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta})$ is the gradient of $E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{D}$ is given as

$$\boldsymbol{D} = \boldsymbol{H}(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}^n) + \mu \, . \, diag(\boldsymbol{H}(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}^n)) \ \ , \tag{9}$$

where $\boldsymbol{H}(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta}) = \frac{\partial^2 E(\boldsymbol{Y}, \boldsymbol{t}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ is the Hessian matrix and μ is a damping factor of the LM algorithm.

When the iterative optimization LM algorithm does not converge, a Continuous Variation Method [7] is used to enforce a priori information about the parameters and regularize the solution. In this strategy, a fudge factor greater than 1 is introduced in (7), multiplying the prior term. This factor converges to 1 along the iterative process, guarantying the convergence of (8) in the initial iterations. However the right solution is reached at the end when the factor becomes 1.

In this work, using synthetic data, the parameters $\alpha_i = \sigma_y^2/\sigma_i^2$ are computed and used in the estimation of $\boldsymbol{\theta}$ with the Bayesian approach. When using real data, an accurate estimation of the amount of noise corrupting the data, σ_y^2, is used. The uncertainty associated with the parameters, σ_i^2, is assumed to be known.

III. MONTE CARLO SIMULATIONS

Monte Carlo simulations [8] were performed in order to test the performance of the proposed algorithm in the estimation of the parameters f and Δt. For each noise level, 1000 synthetic datasets were generated using the two-compartment PASL kinetic model in Eq.(1) . Six different noise levels were obtained as a fraction of the maximum signal

$$\sigma_Y = \beta \times \max\left[\Delta M(\boldsymbol{t}, \boldsymbol{\theta})\right] \ \ , \tag{10}$$

where $\beta = \{10, 50, 75, 100, 125, 150\}[\%]$.

For each set of *TI* points, the process of estimation of the parameters was then performed using both a standard LS method and our proposed Bayesian approach.

The a priori knowledge of the parameters was obtained from the literature [4][9][10][11] where the parameters are described with a Normal distribution. Typical gray matter values at 3T drawn from their physiological variations were considered (see Table I). This information was also used to randomly generate the true values of the parameters for each estimation. The remaining variables used in the two compartment model were considered constant where $\alpha = 0.9$, $\lambda = 0.9$ ml of blood per g of tissue, $T_{2b}^{(*)} = 0.1$ s and $T_{2t}^{(*)} = 0.05$ s [2].

The simulations were performed with four different sets of 100 inversion time points (*TI*), shown in Fig. 2:

(i) **Uniform**; *TI* points uniformly distributed in the interval $[100, 3000] ms$;

TABLE I

PRIOR INFORMATION. THE PARAMETERS ARE RULED BY NORMAL DISTRIBUTIONS WITH DIFFERENT MEANS AND VARIANCES.

Parameter	Mean	Standard Deviation
f (s^{-1})	0.012	0.004
Δt (s)	0.7	0.3
τ (s)	0.7	0.1
τ_{ex} (s)	0.7	0.35
T_{1t} (s)	1.3	0.1
T_{1b} (s)	1.6	0.1

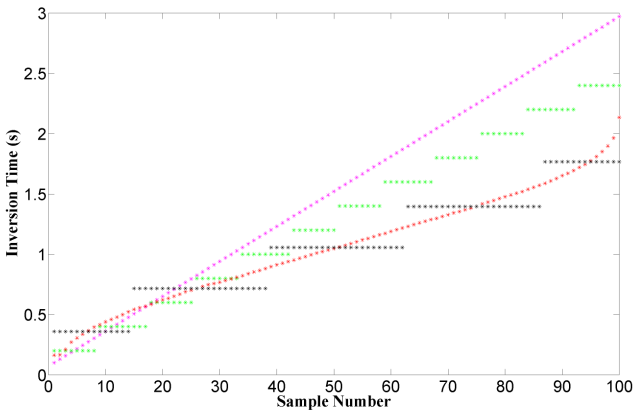


Fig. 2. Accumulation curves for the different sets of *TI* points considered in the simulations (with 125% of noise level): (i) **Uniform** (purple), (ii) **Uniform Sets** (green), (iii) **Optimal** (red) and (iv) **Clustered Optimal** (black).

(ii) **Uniform Sets**; 12 Sets of *TI* points uniformly distributed in the interval $[200, 2400]ms$, which was obtained from the literature [12];

(iii) **Optimal**; *TI* points optimally distributed in the interval $[100, 3000]ms$ [5];

(iv) **Clustered Optimal**; *TI* points optimally organized in 5 clusters in the interval $[100, 3000]ms$.

The optimal sampling strategy (OSS) was obtained according to the algorithm proposed in [5], using the two-compartment model described here, the parameter distributions shown in Table I and the corresponding noise level. A clustered optimal sampling strategy was designed in order to fulfill the experimental requirements of ASL data acquisition, imposed by the minimum temporal resolution feasible and the minimum number of repetitions per *TI* to achieve a reasonable SNR. A k-means clustering procedure was applied to distribute the points in the optimal set among a fixed number of mutually exclusive clusters [13][14]. For comparison, a uniform sampling strategy based on the same interval as the OSS was tested, as well as a slightly different uniform sampling strategy according to the literature [12] in which approximately the same number of inversion times are divided in sets of 12 points in the interval $[200, 2400]ms$.

IV. RESULTS

For each Monte Carlo simulation, the parameter estimation errors, $\epsilon_{\theta^{(i)}}$, and the Signal to Noise Ratio of both data ($SNR_{initial}$) and estimated curves (SNR_{final}) were calculated as follows

$$\begin{aligned} \epsilon_{\theta^{(i)}} &= \frac{\left|\theta^{(i)}_{true} - \theta^{(i)}_{est}\right|}{\theta^{(i)}_{true}} \\ SNR_{initial} &= 10\ \log\left\{\frac{\|\boldsymbol{\Delta M}\|^2}{\|\boldsymbol{Y}-\boldsymbol{\Delta M}\|^2}\right\} \\ SNR_{final} &= 10\ \log\left\{\frac{\|\boldsymbol{\Delta M}\|^2}{\|\hat{\boldsymbol{Y}}-\boldsymbol{\Delta M}\|^2}\right\}\ , \end{aligned} \tag{11}$$

where $\theta^{(i)}_{true}$ and $\theta^{(i)}_{est}$ are the true and estimated parameter values, respectively, while $\boldsymbol{\Delta M}$ is the theoretical curve obtained with (1) for $\boldsymbol{\theta}_{true}$, $\boldsymbol{Y}$ is the noisy data obtained with (3) and $\hat{\boldsymbol{Y}}$ is the estimated curve. An improved SNR ($ISNR$) was determined as the difference between the final and the initial mean values of the SNR.

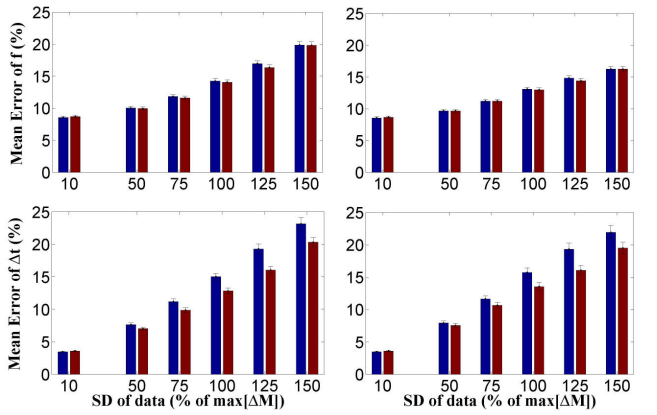
A repeated measures Analysis of Variance (ANOVA) was performed in order to test for any significant effects on these measures of the factors estimation algorithm (*LS*, *Bayesian*), sampling strategy (*Uniform*, *Uniform Sets*, *Optimal*, *Clustered Optimal*) and noise. A significant main effect of the three factors was observed for all measures ($p < 0.001$). Moreover, the interactions between noise and both sampling strategy and estimation algorithm were also significant ($p < 0.001$).

Specifically, the two optimal sampling strategies are more accurate than the two uniform strategies, as expected. In terms of the estimation algorithm, the Bayesian method generally provides more accurate results than the LS. We observe that the two uniform strategies were not significantly different from each other, nor were the two optimal strategies. In particular, it is interesting to notice that clustering the optimal set of TI points around 5 values only does not impair the accuracy of the results. Therefore, in order to better understand the main effects of the estimation algorithm and the sampling strategy, the results are shown only for the data obtained with the *clustered optimal* and the *uniform sets* strategies. The mean and the standard error (SE) of the absolute values of the errors ϵ_f and $\epsilon_{\Delta t}$ are shown in Fig. 3(a) and Fig. 4(a) and those of the *ISNR* are shown in Fig. 3(b) and Fig. 4(b).

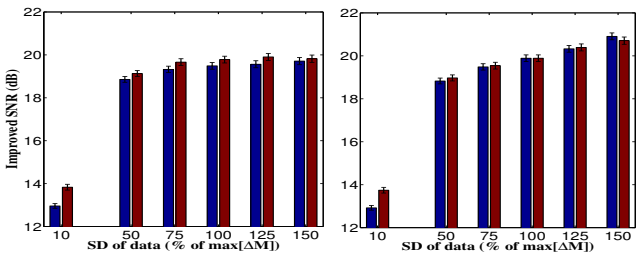
As expected, the absolute values of the errors ϵ_f and $\epsilon_{\Delta t}$ in both parameters increase with the level of noise in the data. Additionally, the improvement in accuracy observed for Bayesian *vs* LS methods and for optimal *vs* uniform strategies also increases with the noise level. For the lowest noise level, the absolute values of the estimation errors are relatively high. This is a consequence of the variability of the fixed model parameters, which is evident for low noise levels but becomes dominated by the noise in the data as the noise level increases. For the same reason, the *ISNR* is considerably smaller for the lowest noise level relative to higher noise levels as a consequence of the decreased SNR_{final}.

V. CONCLUSION

A Bayesian estimation algorithm was implemented and validated for the measurement of perfusion and arterial transit time, based on a two-compartment kinetic model of

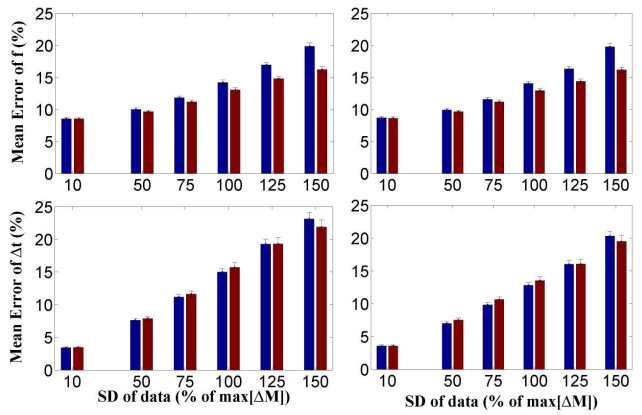


(a)

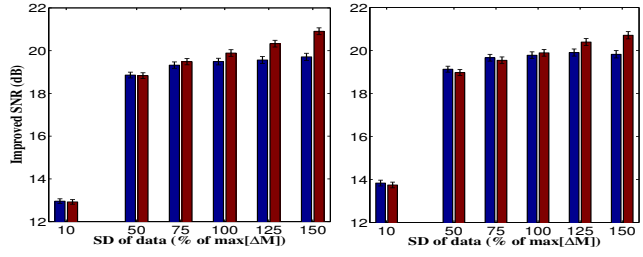


(b)

Fig. 3. **(a)** Absolute estimation errors (%) of f (top) and Δt (bottom) (mean $\pm$ SE) and **(b)** Improved SNR (mean $\pm$ SE), for Uniform Sets (left/blue columns) and Clustered Optimal sets (right/red columns) of *TI* sampling points, considering the LS (left) and the Bayesian (right) estimation algorithms and for each level of noise.



(a)



(b)

Fig. 4. **(a)** Absolute estimation errors (%) of f (top) and Δt (bottom) (mean $\pm$ SE) and **(b)** Improved SNR (mean $\pm$ SE), for Uniform Sets (left) and Clustered Optimal sets (right) of *TI* sampling points, considering the LS (left/blue columns) and the Bayesian (right/red columns) estimation algorithms and for each level of noise.

PASL data and incorporating a priori knowledge about the physiological variation of the parameters. Monte Carlo simulations were performed to compare the estimation accuracy obtained with the Bayesian algorithm relative to a standard LS method, using four different sets of *TI* points. We showed that both the estimation algorithm and the sampling strategy used have an effect on the results, particularly for high noise levels. Most importantly, we found that the estimation errors are minimized when optimization with a Bayesian approach is used both in the estimation algorithm and in the design of the sampling strategy. Interestingly, we also showed that clustering the optimal sampling points around a limited number of distinct *TI* values can be used in practice without impairing the accuracy of the results.

REFERENCES

[1] R.B. Buxton, L.R. Frank, E.C. Wong, B. Siewert, S. Warach, and R.R. Edelman, A general kinetic model for quantitative perfusion imaging with arterial spin labeling, *Magnetic Resonance in Medicine*, vol. 40, no. 3, 1998, pp. 383-396.

[2] P.M. Figueiredo, *Measuring Brain Perfusion using Arterial Spin Labeling by Magnetic Resonance Imaging*, Ph.D. Dissertation, Wolfson College - University of Oxford, Oxford; 2003.

[3] P.M. Figueiredo, S. Clare, and P. Jezzard, Quantitative perfusion measurements using pulsed Arterial Spin Labeling: effects of large region-of-interest analysis, *Journal of Magnetic Resonance Imaging*, vol. 21, no. 6, 2005, pp. 676-682.

[4] J. Xie, D. Gallichan, R.N. Gunn, and P. Jezzard, Optimal design of Pulsed Arterial Spin Labeling MRI experiments, *Magnetic Resonance in Medicine*, vol. 59, no. 4, 2008, pp. 826-834.

[5] J. Sanches, I. Sousa, and P. Figueiredo, Bayesian Fisher Information Criterion for Sampling Optimization in ASL-MRI, *in ISBI*, 2010.

[6] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, Modeling of Data - Nonlinear Models, *in Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 1988-1992, pp. 45-58.

[7] S.Z. Li, *Markov random field modeling in image analysis*, Springer-Verlag New York, Inc., 2001.

[8] R.Y. Rubenstein and D.P. Kroese, *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York, 2007.

[9] W.M. Luh, E.C. Wong, P.A. Bandettini, and J.S. Hyde, QUIPSS II with thin-slice TI1 periodic saturation: a method for improving accuracy of quantitative perfusion imaging using pulsed arterial spin labeling, *Magn Reson Med*, vol. 41, no. 6, 1999, pp. 1246-54.

[10] H. Lu, C. Clingman, X. Golay, and P.C.V. Zijl, Determining the longitudinal relaxation time T1 of blood at 3.0 Tesla, *Magnetic Resonance in Medicine*, vol. 52, 2004, pp. 679-682.

[11] W. Shin, H. Gu, and Y. Yang, Fast high-resolution T1 mapping using inversion-recovery Look-Locker echoplanar imaging at steady state: optimization for accuracy and reliability, *Magnetic Resonance in Medicine*, vol. 61, 2009, pp. 899-906.

[12] M. Günther, K. Oshio, and D.A. Feinberg, Single-shot 3D imaging techniques improve arterial spin labeling perfusion measurements, *Magnetic Resonance in Medicine*, vol. 54, 2005, pp. 491-498.

[13] G.A.F. Seber, *Multivariate Observations*, Hoboken, John Wiley & Sons, NJ, 1984.

[14] H. Spath, *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*, Translated by J. Goldschmidt, Halsted Press, New York, 1985.