Optimal Sampling and Estimation in PASL Perfusion Imaging

Nuno Santos, J. Miguel Sanches, Inês Sousa, and Patrícia Figueiredo*

***Abstract*—Pulsed arterial spin labeling (PASL) techniques potentially allow the absolute, noninvasive quantification of brain perfusion using MRI. This can be achieved by fitting a kinetic model to the data acquired at a number of sampling times. However, the intrinsically low signal-to-noise ratio of PASL measurements usually requires substantial signal averaging, which may result in undesirably long scanning times. A judicious choice of the sampling points is, therefore, crucial in order to minimize scanning time, while optimizing estimation accuracy. On the other hand, *a priori* information regarding the model parameters may improve estimation performance. Here, we propose a Bayesian framework to determine an optimal sampling strategy and estimation method for the measurement of brain perfusion and arterial transit time (ATT). A Bayesian Fisher information criterion is used to determine the optimal sampling points and a MAP criterion is employed for the estimation of the model parameters, both taking into account the uncertainty in the model parameters as well as the amount of noise in the data. By Monte Carlo simulations, we show that using optimal compared to uniform sampling strategies, as well as the Bayesian estimator relative to a standard least squares approach, improves the accuracy of perfusion and ATT measurements. Moreover, we also demonstrate the applicability of the proposed approach to real data, with the advantage of reduced intersubject variability relative to conventional sampling and estimation approaches.**

***Index Terms*—Bayesian, estimation, Fisher, MRI, pulsed arterial spin labeling (PASL), sampling.**

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 (PASL) MRI techniques offer a noninvasive 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) [1]. The magnetization difference between a labeled image and a control image, as a function of the TI, can be described by a kinetic model [2], which makes PASL a potentially quantitative technique.

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 (ATT) and in some pathological conditions such as cerebrovascular disease. Therefore, in order to correctly estimate perfusion, the ATT should also be estimated by fitting the PASL model to magnetization difference 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 intrinsically low signal-to-noise ratio (SNR). Therefore, a compromise between the number of TIs sampled and the total scanning time has to be made. Previous studies have shown that the distribution of the sampling points along time has a strong effect on the accuracy of the estimation of the parameters [4], [5]. Therefore, a correct choice of the value and number of sampling points is decisive in order to minimize scanning time, while optimizing estimation accuracy.

Optimal sampling strategies have previously been designed based on the Fisher information matrix optimality criterion for the simultaneous estimation of perfusion and ATT [4], [5]. However, the uncertainty associated with the remaining model parameters was not taken into account. Here, we propose a Bayesian framework for PASL perfusion quantification based on the MAP criterion [6], [7] for model estimation, where the *a priori* knowledge of the physiological distributions of the model parameters, as well as the amount of noise in the data, are taken into account. The purpose of the framework is twofold. First, the sampling points for optimal perfusion and ATT estimation based on the MAP criterion are determined (following our previous work [8]). Second, both parameters are estimated from the data using the same criterion (as previously presented by Santos *et al.* [9]). Using both simulated and empirical data, the proposed methodologies are compared with uniformly distributed sampling points and a standard *least squares* (LS) estimation method.

Manuscript received February 2, 2011; revised May 4, 2011 and July 14, 2011; accepted August 7, 2011. Date of publication August 15, 2011; date of current version October 19, 2011. This work was supported in part by Hospital da Luz, in part by the Portuguese Science Foundation under Grant PTDC/SAU-BEB/65977/2006 and Grant SFRH/BDE /33365/2008, and in part by the ISR/IST plurianual funding through the PIDDAC Program funds. *Asterisk indicates corresponding author.*

N. Santos and I. Sousa are with the Institute for Systems and Robotics, Lisbon 1049-001, Portugal, and also with the Department of Bioengineering, Instituto Superior Técnico, Lisbon 1049-001, Portugal, and also with the Healthcare Sector, Siemens S.A., Amadora 2720-093, Portugal (e-mail: njgsantos@gmail.com; inesnsousa@gmail.com).

J. M. Sanches is with the Institute for Systems and Robotics, Lisbon 1049-001, Portugal, and also with the Department of Bioengineering, Instituto Superior Técnico, Lisbon 1049-001, Portugal (e-mail: jmrs@ist.utl.pt).

*P. Figueiredo is with the Institute for Systems and Robotics, Lisbon 1049-001, Portugal, and also with the Department of Bioengineering, Instituto Superior Técnico, Lisbon 1049-001, Portugal (e-mail: patricia.figueiredo@ist.utl.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2011.2164916

II. Methods

The two main contributions of this paper are first mathematically formulated: 1) a Bayesian algorithm to estimate model parameters from a limited number of observations; and 2) a strategy to determine the optimal set of sampling time points, which minimizes the variance of the previously defined estimator. The kinetic model used to describe the PASL signal is then presented and the identifiability of the model parameters of interest is analyzed.

The estimation procedure is designed in a Bayesian framework where an additive model for the noise is assumed and prior information about the parameters is taken into account. The prior term is twofold. First, the poor SNR and the nonlinear dependence of the PASL magnetization difference on the model parameters makes the maximum-likelihood estimation problem (based only on the likelihood function) ill-posed [10]. The prior term is, therefore, important to regularize the solution [11]–[13]. Second, a prior term is also used to guide the estimation toward a physiologically plausible solution, by incorporating knowledge about the parameter distributions.

A. Bayesian Estimator

Let us consider the function $\Delta M(t,\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of P unknown parameters $\boldsymbol{\theta}=[\theta_1,\theta_2,\ldots,\theta_P]$, to be estimated from a set of observations $Y=[y_1,y_2,\ldots,y_N]$ taken at the time points $\boldsymbol{t}=[t_1,\ldots,t_N]$. Assuming an additive white Gaussian noise (AWGN) model to describe the data generation process [14], the observation model is given by

$$y_i=\Delta M(t_i,\boldsymbol{\theta})+\eta\ ,\ i=1,\ldots,N \tag{1}$$

where $\eta\sim\mathcal{N}(0,\sigma_y^2)$, with σ_y being the standard deviation of the noise. Therefore, $p(y_i|t_i,\boldsymbol{\theta})=\mathcal{N}(\Delta M(t_i,\boldsymbol{\theta}),\sigma_y^2)$.

The estimation of the vector $\boldsymbol{\theta}$ with the MAP criterion can be formulated as the following optimization task [15]:

$$\hat{\boldsymbol{\theta}}=\arg\min_{\boldsymbol{\theta}}E(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta}). \tag{2}$$

The energy function to be minimized is defined by

$$E(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta})=-\log\left[p(\boldsymbol{Y}|\boldsymbol{t},\boldsymbol{\theta})p(\boldsymbol{\theta})\right] \tag{3}$$

where $p(\boldsymbol{Y}|\boldsymbol{t},\boldsymbol{\theta})$ is the likelihood function describing the acquisition process, given by $p(\boldsymbol{Y}|\boldsymbol{t},\boldsymbol{\theta})=\prod_{i=1}^{N}p(y_i|t_i,\boldsymbol{\theta})$ if statistic independence of the observations is assumed.

Each of the P elements of $\boldsymbol{\theta}$, θ_i, is assumed to be independent and Gaussian distributed with mean θ_{0i}, and standard deviation σ_i reflecting the known level of uncertainty associated with it. Thus, the parameter vector $\boldsymbol{\theta}$ is described by a multivariate normal distribution $\mathcal{N}(\overline{\boldsymbol{\theta}},\boldsymbol{C})$, where $\boldsymbol{C}=\mathrm{diag}(\{\sigma_1^2,\sigma_2^2,\ldots,\sigma_P^2\})$ is a diagonal covariance matrix.

Although a Gaussian distribution does not accurately describe the ATT, this was considered a convenient parsimonious approximation, in order to well pose the estimation problem [10]. In fact, approximately, Gaussian distributions of ATTs have been observed (e.g., [5], [16]) and such approximation has also been used in previous approaches to PASL model estimation [14], [17].

The energy function, defined by (3), hence, contains two terms, the data fidelity term, which depends on the observations, and the prior term, which describes the prior knowledge about the parameters, and it can be written as follows:

$$E(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta})=\underbrace{\frac{1}{2}\|\Delta M(\boldsymbol{t},\boldsymbol{\theta})-\boldsymbol{Y}\|_2^2}_{\text{Data Fidelity term}}+\underbrace{\frac{1}{2}\boldsymbol{\sigma}_y^2(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})^T\boldsymbol{C}^{-1}(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})}_{\text{Prior term}}. \tag{4}$$

The minimizer of (4) is computed by finding its stationary point with respect to $\boldsymbol{\theta}$, $\nabla_{\boldsymbol{\theta}}E(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta})=0$.

In order to assess the effectiveness of the prior information in the estimation process, the common LS method is also applied to estimate θ from the same set of time points. The corresponding energy function to be minimized contains only the data fidelity term and is given by

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

In these two estimation procedures (LS and Bayesian), the optimization is accomplished by using the Levenberg–Marquardt (LM) algorithm [18]. For a correct comparison between the two methods, the algorithm was implemented as described in the literature [19]. The LM algorithm is an iterative process, whereby for each step n of the iteration the estimation of the parameters is given by

$$\boldsymbol{\theta}^{n+1}=\boldsymbol{\theta}^n+\boldsymbol{D}^{-1}\ \cdot\ \nabla_{\boldsymbol{\theta}}E(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta}^n) \tag{6}$$

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

$$\boldsymbol{D}=\boldsymbol{H}(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta}^n)+\mu\ \cdot\ \mathrm{diag}(\boldsymbol{H}(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta}^n)) \tag{7}$$

with $\boldsymbol{H}(\boldsymbol{t},\boldsymbol{\theta},\boldsymbol{Y})=\{H_{k,r}(\boldsymbol{t},\boldsymbol{\theta},\boldsymbol{Y})\}$ being the Hessian matrix of $E(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta})$ with respect to the parameters

$$H_{k,r}(\boldsymbol{t},\boldsymbol{\theta},\boldsymbol{Y})=\frac{\partial^2E(\boldsymbol{Y},\boldsymbol{t},\boldsymbol{\theta})}{\partial\theta_k\partial\theta_r} \tag{8}$$

and μ is the damping factor of the LM algorithm.

When the iterative optimization LM algorithm does not converge, a continuous variation method [20] 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 (4), multiplying the prior term. This factor converges to 1 along the iterative process, guaranteeing the convergence of (6) in the initial iterations. However, the correct solution is reached at the end when it becomes 1.

For the 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. For the empirical data, an 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, based on published data (see Table I).

TABLE I
PHYSIOLOGICAL PRIOR INFORMATION

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

B. Optimal Sampling Strategy

The selection of the N sampling time points $\boldsymbol{t}$ can be optimally performed in order to minimize the variance of the estimator [6]. These optimal time points depend on the function $\Delta M(\boldsymbol{t},\boldsymbol{\theta})$ and the distribution of the parameters $p(\boldsymbol{\theta})$. The minimum variance estimator is obtained by maximizing the determinant of the Fisher information matrix [21], $\boldsymbol{G}(\boldsymbol{t},\boldsymbol{\theta}) = \{G_{k,r}\}$, defined as

$$G_{k,r}(\boldsymbol{t},\boldsymbol{\theta}) = -\mathcal{E}_{\boldsymbol{Y}}\left[H_{k,r}(\boldsymbol{t},\boldsymbol{\theta},\boldsymbol{Y})\right] \tag{9}$$

where $\mathcal{E}_{\boldsymbol{Y}}(.)$ denotes the expectation operator with respect to the multivariate random variable $\boldsymbol{Y}$, which gives

$$\boldsymbol{t}^*(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{t}} \mathcal{J}(\boldsymbol{t},\boldsymbol{\theta}) \tag{10}$$

where $\mathcal{J}(\boldsymbol{t},\boldsymbol{\theta}) = \det \boldsymbol{G}(\boldsymbol{t},\boldsymbol{\theta})$.

Since the AWGN model is adopted here, the data are described by a normal distribution with mean $\Delta M(t_i,\boldsymbol{\theta})$ which means $\mathcal{E}_{\boldsymbol{Y}}(y_i) = \Delta M(t_i,\boldsymbol{\theta})$. Taking (4) into account

$$\begin{aligned}\frac{\partial^2 E}{\partial\theta_k\partial\theta_r} &= \sum_{i=1}^{N}\left[\frac{\partial\Delta M(t_i,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial\Delta M(t_i,\boldsymbol{\theta})}{\partial\theta_r}\right] \\ &\quad + \sum_{i=1}^{N}\left[(\Delta M(t_i,\boldsymbol{\theta}) - y_i)\frac{\partial^2\Delta M(t_i,\boldsymbol{\theta})}{\partial\theta_r\partial\theta_k}\right] \\ &\quad + \delta_{k,r}\frac{\sigma_y^2}{\sigma_k^2}\end{aligned} \tag{11}$$

where $\delta_{k,r}$ is the Kronecker delta function. By computing the expectation of (11), (9) gives

$$G_{k,r} = \sum_{i=1}^{N}\left[\frac{\partial\Delta M(t_i,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial\Delta M(t_i,\boldsymbol{\theta})}{\partial\theta_r}\right] + \frac{\sigma_y^2}{\sigma_k^2}\delta_{k,r}. \tag{12}$$

Here, we propose an incremental strategy to solve (10) whereby each time point is computed as a function of the previously computed time points. The cost function used to compute a new optimal time point at the nth iteration is given by $\mathcal{J}_n(t,\boldsymbol{t}_{n-1},\boldsymbol{\theta}) = \det G_n(t,\boldsymbol{t}_{n-1},\boldsymbol{\theta})$, where $\boldsymbol{t}_{n-1} = [t_1, t_2, \ldots, t_{n-1}]$ are the $n-1$ optimal time points estimated up to the moment and G_n is the incremental Fisher information matrix given by

$$\begin{aligned}G_{k,r}(t,\boldsymbol{t}_{n-1},\boldsymbol{\theta}) &= G_{k,r}(t_{n-1},\boldsymbol{t}_{n-2},\boldsymbol{\theta}) \\ &\quad + \frac{1}{\sigma_y^2}\frac{\partial\Delta M(t,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial\Delta M(t,\boldsymbol{\theta})}{\partial\theta_r}\end{aligned} \tag{13}$$

where

$$\begin{aligned}G_{k,r}(t_{n-1},\boldsymbol{t}_{n-2},\boldsymbol{\theta}) &= \frac{1}{\sigma_k^2}\delta_{k,r} \\ &\quad + \frac{1}{\sigma_y^2}\sum_{i=1}^{n-1}\left[\frac{\partial\Delta M(t_i,\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial\Delta M(t_i,\boldsymbol{\theta})}{\partial\theta_r}\right]\end{aligned} \tag{14}$$

was incrementally estimated in the previous $n-1$ steps. Then, the nth optimum time point t_n is determined by solving the following 1-D optimization problem:

$$t_n = \arg\max_{t} \mathcal{J}_n(t,\boldsymbol{t}_{n-1},\boldsymbol{\theta}). \tag{15}$$

The optimization problem (15) is solved L times, in order to obtain the $N \ll L$ optimum time points needed to design the acquisition experiment. This set of points satisfies (10) if the second term of (13), $(\partial\Delta M(t_i,\boldsymbol{\theta})/\partial\theta_k)(\partial\Delta M(t_i,\boldsymbol{\theta})/\partial\theta_r) \geq 0$, is nonnegative [22], which is the case.

The N points obtained from (10), or equivalently from (15), depend on the parameter vector θ to be estimated, which is obviously unknown in this step. However, its distribution $p(\theta)$ is assumed to be known. Hence, the procedure for the identification of the N optimal time points is the following.

1) A large number of $L \gg N$ parameter vectors $\boldsymbol{\theta}_j$ are obtained by randomly sampling $p(\boldsymbol{\theta})$.
2) For each parameter vector $\boldsymbol{\theta}_j$, (13), (14), and (15) are used to obtain a set of $M \gg N$ optimal points $\boldsymbol{T}_j$.
3) Each set of M optimal time points, obtained for each $\boldsymbol{\theta}_j$, are appended to a running histogram of optimal time points $h(t,\boldsymbol{T})$.
4) The final N optimal time points, covering the full distribution of vector $\boldsymbol{\theta}$, are obtained from the final histogram, by partitioning the area under the curve in N equal intervals (through the cumulative curve).

C. Kinetic Model

The magnetization difference measured in PASL is described by the standard kinetic model introduced by Buxton *et al.* [2]

$$\begin{aligned}\Delta M(t,\boldsymbol{\theta}) &= 2\alpha f M_{0b}\frac{\exp\{-R_{1t}t\}\exp\{-D_1\Delta t\}}{D_1} \\ &\quad \times \begin{cases} 0 & t < \Delta t \\ 1-\exp\{-D_1(t-\Delta t)\} & \Delta t \leq t < \Delta t+\tau \\ 1-\exp\{-D_1\tau\} & \Delta t+\tau \leq t \end{cases}\end{aligned} \tag{16}$$

with

$$D_1 = R_{1b} - R_{1t}^{\text{app}},\quad R_{1t}^{\text{app}} = \frac{1}{T_{1t}} + \frac{f}{\lambda},\quad R_{1b} = \frac{1}{T_{1b}} \tag{17}$$

where $\boldsymbol{\theta} = [f, \Delta t]$ is the vector of parameters to be estimated, including f, the perfusion, and Δt, the ATT; and the remaining constant parameters are τ, the label bolus time width; T_{1t}, the tissue longitudinal relaxation time; T_{1b}, the blood longitudinal relaxation time; α, the labeling efficiency; and M_{0b}, the arterial blood equilibrium magnetization. This is given by $M_{0b} = M_{0t}/\lambda$, where λ is the brain–blood partition coefficient

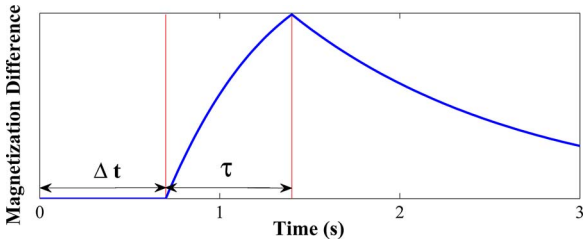

Fig. 1. PASL magnetization difference signal as a function of the TI, according to the standard kinetic model.

of water and M_{0t} is the brain tissue equilibrium magnetization, measured experimentally [23]. As illustrated in Fig. 1, ΔM is 0 until the labeled arterial blood water arrives at the imaging site, at time Δt, and it then increases for a period of time corresponding to the time width of the labeled bolus τ. After $\Delta t + \tau$, ΔM decays back to 0 as the label magnetization relaxes toward equilibrium with time constant T_{1t}.

D. Parameter Identifiability

A structural model analysis is performed here, in order to ascertain the parameter identifiability, for the specific case of the estimation of the parameters of interest, f and Δt, i.e., $\boldsymbol{\theta} = \{f, \Delta t\}$. The model is structurally identifiable if the Fisher information matrix is nonsingular [24] and enough data are available. However, this approach only guarantees local identifiability [24]. In this particular case, $\mathcal{J}(\boldsymbol{t}, \boldsymbol{\theta})$ in (10) is

$$\mathcal{J}(\boldsymbol{t},\boldsymbol{\theta}) = \left(\sum_{i=1}^{\mathcal{N}} u_i^2 + \alpha_u\right)\left(\sum_{i=1}^{\mathcal{N}} v_i^2 + \alpha_v\right) - \left(\sum_{i=1}^{\mathcal{N}} u_i v_i\right)^2 \tag{18}$$

where $u_i = \frac{\partial \Delta M(t_i, f, \Delta t)}{\partial f}$, $v_i = \frac{\partial \Delta M(t_i, f, \Delta t)}{\partial \Delta t}$, $\alpha_u = \sigma_y^2/\sigma_f^2$, and $\alpha_v = \sigma_y^2/\sigma_{\Delta t}^2$. Defining $\boldsymbol{u} = \{u_1, u_2, \ldots, u_N\}$ and $\boldsymbol{v} = \{v_1, v_2, \ldots, v_N\}$, (18) may be written as follows:

$$\begin{aligned}\mathcal{J}(\boldsymbol{t},\boldsymbol{\theta}) &= (\|\boldsymbol{u}\|^2 + \alpha_u)(\|\boldsymbol{v}\|^2 + \alpha_v) - <\boldsymbol{u},\boldsymbol{v}>^2 \\ &= \underbrace{\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - <\boldsymbol{u},\boldsymbol{v}>^2}_{\geq 0} + \underbrace{\alpha_v\|\boldsymbol{u}\|^2 + \alpha_u\|\boldsymbol{v}\|^2}_{>0}\end{aligned} \tag{19}$$

where $<,>$ denotes the vector inner product operation.

Under the assumption of enough data, *structural identifiability* holds because $\mathcal{J}(\boldsymbol{t}, \boldsymbol{\theta}) > 0$. The first term in (19), $\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - <\boldsymbol{u}, \boldsymbol{v}>^2$, is nonnegative by the Cauchy–Schwarz inequality [6] and the second, $\alpha_v\|\boldsymbol{u}\|^2 + \alpha_u\|\boldsymbol{v}\|^2$, is strictly positive if at least one of u_i or v_i is not zero.

Practical identifiability, however, may not hold if there are not enough data or if data were not acquired at the appropriated locations [25]. In order for $\mathcal{J}(\boldsymbol{t}, \boldsymbol{\theta})$ in (19) to be strictly positive, at least one sampling time point t_k must be acquired at a time instant where the partial derivatives of $\Delta M(t_k, \theta)$ are not null. Fig. 2 displays the partial derivatives with respect to f and Δt as a function of time. Both derivatives are nonzero for $t > \Delta t$, which means that practical identifiability is guaranteed locally

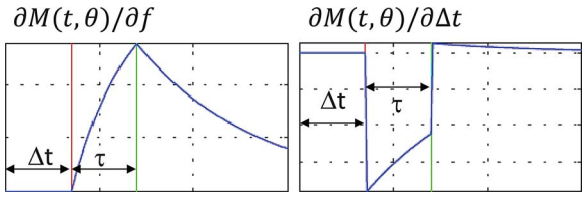

Fig. 2. Partial derivatives of $\Delta M(t, \theta)$ with respect to (left) f and (right) Δt as a function of time.

if at least one sampling time point is greater than Δt for a particular experiment.

III. Results

In this section, the results obtained with both synthetic and empirical data will be presented. First, sets of optimal sampling points are obtained, as a function of the amount of noise in the data and the uncertainty of the model parameters. Second, Monte Carlo simulations at different noise levels are performed to test the performance of the Bayesian estimation algorithm, relative to a standard LS method, using the optimal sampling strategy, relative to uniform sampling. Finally, empirical data acquired with the two sampling schemes are analyzed using both the proposed and the standard methods for estimation of perfusion and ATT.

A. Optimal Sampling Strategy

Sets of optimal sampling points were obtained for the estimation of f and Δt, using different amounts of noise corrupting the data σ_Y as well as different levels of parameter uncertainty σ_f and $\sigma_{\Delta t}$ in order to evaluate the sensitivity of the optimal sampling points to errors in these factors. The noise level is considered as a fraction β of the maximum signal (according to empirical observation): $\sigma_Y = \beta \times \max[\Delta M(\boldsymbol{t}, \boldsymbol{\theta})]$.

Physiologically plausible distributions of the model parameters, for gray matter (GM) of healthy subjects at 3 T, were obtained from the literature [4], [26]–[28]. For pathological populations, the parameter distributions would have to be adapted according to the clinical hypothesis. Parameters described by normal distributions are shown in Table I, while α and λ are assumed to be known: $\alpha = 0.9$ and $\lambda = 0.9$.

The set of optimal sampling points obtained with our proposed method, using these parameter values and a noise level of $\beta = 125\%$, is illustrated by the cumulative curve shown in Fig. 3 (red curve). In comparison with a uniform sampling strategy over the same interval (black curve), the optimized sampling set has a higher density of points around the values $\Delta t_0 = 0.7$ s and $\Delta t_0 + \tau_0 = 1.4$ s, as observed before [4].

The optimal sampling, therefore, depends essentially on the values of Δt and τ. In our case, the value of τ can be assumed with a high degree of certainty because we use several saturation pulses to limit the length of the bolus, such as in the Q2TIPS (quantitative imaging of perfusion using a single subtraction, second version, with interleaved periodic saturation) PASL sequence [26]. The optimal sampling strategy is, hence, mainly determined by Δt. Its sensitivity to this parameter is illustrated

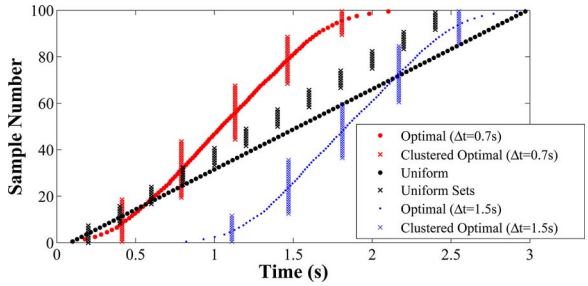

Fig. 3. Cumulative curves of the four sampling sets considered in the simulations (with $\beta = 125\%$ noise level). The optimal sampling sets that would be obtained for $\Delta t = 1.5$ s are also shown to illustrate the sensitivity to this parameter.

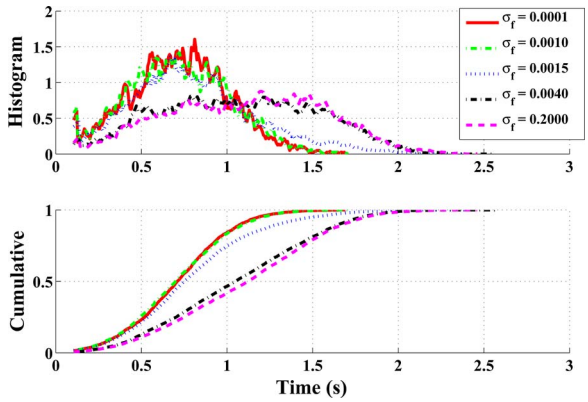

Fig. 4. Histograms and cumulative curves of optimal sampling sets with varying σ_f, $\sigma_{\Delta t} = 0.3$ s, and $\beta = 125\%$.

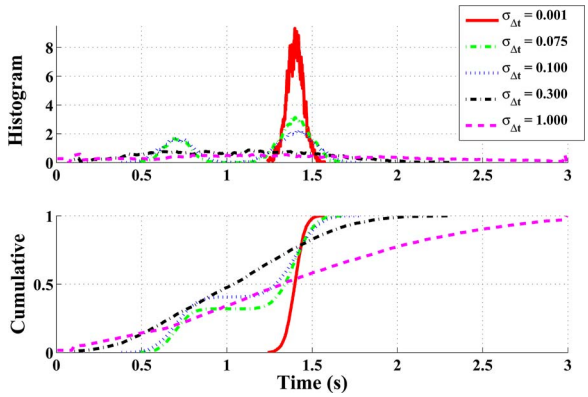

Fig. 5. Histograms and cumulative curves of optimal sampling sets with varying $\sigma_{\Delta t}$, $\sigma_f = 0.004$ s^{-1}, and $\beta = 125\%$.

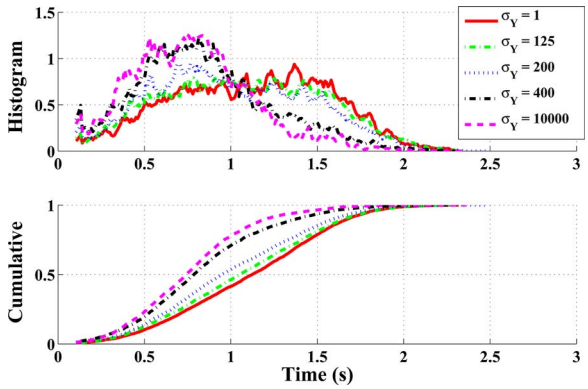

Fig. 6. Histograms and cumulative curves of optimal sampling sets with varying β, $\sigma_f = 0.004$ s^{-1}, and $\sigma_{\Delta t} = 0.3$ s.

by comparison with the optimal sampling sets obtained with $\Delta t = 1.5$ s, as shown in Fig. 3.

First, sensitivity to f uncertainty was analyzed by considering five uncertainty levels, $\sigma_f = [0.0001, 0.001, 0.0015, 0.004, 0.2]\text{s}^{-1}$, while keeping Δt uncertainty level constant at $\sigma_{\Delta t} = 0.3$ s, and the noise level constant at $\beta = 125\%$. The resulting histograms and corresponding cumulative curves are shown in Fig. 4. The values of σ_f were chosen so as to contain the reference value $\sigma_f = 0.004$ s^{-1}, and to obtain a significant variation of the resulting curves. It can be observed that, as σ_f decreases, the distribution of the optimal sampling points moves toward $\Delta t_0 = 0.7$ s and away from the point $\Delta t_0 + \tau_0 = 1.4$ s. Hence, better knowledge of the value of parameter f will concentrate estimation efforts on the other unknown parameter Δt.

Second, sensitivity to Δt uncertainty was analyzed by considering five uncertainty levels, $\sigma_{\Delta t} = [0.001, 0.075, 0.1, 0.3, 1]$s, while keeping $\sigma_f = 0.004$ s^{-1} and $\beta = 125\%$. The resulting histograms and corresponding cumulative curves are shown in Fig. 5. In this case, better knowledge of Δt moves the distribution of optimal sampling points away from $\Delta t_0 = 0.7$ s and toward $\Delta t_0 + \tau_0 = 1.4$ s. It should be noted that changes in the distribution of sampling points as $\sigma_{\Delta t}$ decreases are more abrupt in this case relative to what was observed for σ_f. It is also noted that the optimal sampling strategy approaches the uniform scheme as $\sigma_{\Delta t}$ increases.

Finally, sensitivity to β was analyzed by considering five noise levels, $\beta = [1, 125, 200, 400, 10000]\%$, while keeping $\sigma_f = 0.004$ s^{-1} and $\sigma_{\Delta t} = 0.3$ s. The resulting histograms and corresponding cumulative curves are shown in Fig. 6. It can be observed that, in this range, a change in the noise level has a small influence on the distribution of optimal sampling points. Nevertheless, as the level of noise increases, the distribution of the optimal sampling points moves toward the point $\Delta t_0 = 0.7$ s and away from the point $\Delta t_0 + \tau_0 = 1.4$ s.

For the Monte Carlo simulations, sets of 100 TI sampling time points were determined, for six noise levels, $\beta = [10, 50, 75, 100, 125, 150]\%$, and considering the reference σ_f and $\sigma_{\Delta t}$ values presented in Table I, using four different strategies, shown in Fig. 4

1) *Optimal:* TIs optimally distributed in the interval $[100, 3000]$ms, using the proposed method.
2) *Clustered Optimal:* TI points optimally distributed in the interval $[100, 3000]$ms, using the proposed method, and subsequently organized into five clusters. A *k*-means clustering procedure was applied to distribute the points in the optimal set among a fixed number, 5, of mutually exclusive clusters [29], [30]. This strategy was designed in order to fulfill the experimental requirements of arterial spin labeling (ASL) data acquisition, imposed by the minimum

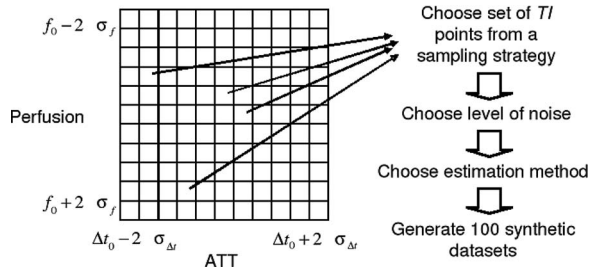

Fig. 7. Generation of synthetic data: for each pair of parameters f and Δt, 100 synthetic datasets are generated for each sampling strategy, noise level, and estimation method.

temporal resolution feasible and the minimum number of repetitions required to achieve a reasonable SNR per TI.

3) *Uniform:* TI points uniformly distributed in the interval $[100, 3000]$ms. This strategy was considered for comparison between our informed method and a completely uninformed approach.

4) *Uniform Sets:* 12 sets of TI points uniformly distributed in the interval $[200, 2400]$ms, according to a common approach in the literature [31]. This slightly modified uniform sampling strategy was considered due to its relevance in the literature.

B. Monte Carlo Simulations

For the Monte Carlo simulations, the space of the parameters under study, f and Δt, was sampled in the range of $\theta - 2 \times \sigma_\theta$ to $\theta + 2 \times \sigma_\theta$, where σ_θ is the standard deviation of the parameter θ, as illustrated in Fig. 7. For each pair of parameters, 100 synthetic datasets were generated using the PASL kinetic model described in (16), with each of the six noise levels, $\beta = [10, 50, 75, 100, 125, 150]\%$, and with each of the four sampling strategies (uniform, uniform sets, optimal, and clustered optimal). In each case, the model parameters f and Δt were then estimated using both the standard LS method (LS) and our proposed Bayesian approach (Bayesian).

Two measures were obtained to evaluate the performance of the different sampling and estimation techniques: the parameter estimation normalized mean square errors $\epsilon^{\mathrm{NMS}}_{\theta^{(i)}} = \sum_v \left(\theta^{(v,i)}_{\mathrm{true}} - \theta^{(v,i)}_{\mathrm{est}}\right)^2 / \sum_v \left(\theta^{(v,i)}_{\mathrm{true}}\right)^2$ where $\theta^{(i)}_{\mathrm{true}}$ and $\theta^{(i)}_{\mathrm{est}}$ are the true and estimated parameter values, respectively, and the improved SNR, $\mathrm{ISNR} = \left[\|\boldsymbol{Y} - \boldsymbol{\Delta M}\| / \|\hat{\boldsymbol{Y}} - \boldsymbol{\Delta M}\|\right]_{dB}$ where $\boldsymbol{\Delta M}$ is the theoretical curve obtained with (16) for $\boldsymbol{\theta}_{\mathrm{true}}$, $\boldsymbol{Y}$ represents the noisy data obtained with (1), and $\hat{\boldsymbol{Y}}$ is the estimated curve. The mean NMSE and ISNR of f and Δt estimation, at a noise level of $\beta = 125\%$, are shown in Fig. 8, for each pair of parameter values in parameter space.

As expected, the Bayesian method generally provides reduced estimation errors and increased SNR improvement relative to the LS approach. In Fig. 8, it can be seen that the NMSE values produced by the Bayesian method are relatively lower compared with the LS estimator at the center of the parameter space. However, they increase toward the periphery as the real values of

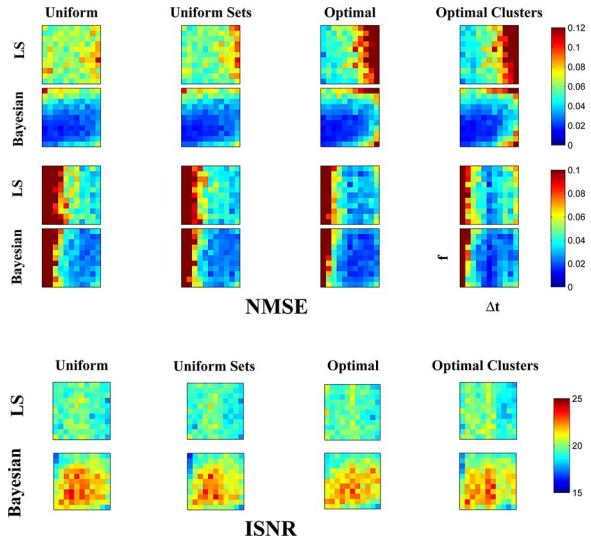

Fig. 8. NMSE and ISNR of (top) f and Δt and (bottom) ISNR, for each pair of parameter values, obtained with each method.

the parameters are deviated from the expected central values, as a consequence of the influence of the prior term in the Bayesian estimator. In general, the Bayesian approach produces greater ISNR values than the LS approach.

In terms of the sampling strategy, the optimal and clustered optimal strategies tend to outperform the uniform and uniform sets strategies, in what concerns the estimation of the parameter Δt. In fact, the NMSE values obtained for Δt with the two optimal sampling strategies are lower relative to the values obtained with the two uniform strategies. As far the f is concerned, however, the errors produced using optimal strategies become degraded as Δt increases, away from the central value assumed *a priori*. This observation indicates that it is critical do adjust the prior information when transit times are expected to be prolonged beyond normal values.

The average values of the NMSE and ISNR measurements, across parameter space, are shown in Fig. 9, as a function of the noise level. A repeated measures analysis of variance was performed in order to test for any significant effects on these measures of the three factors under study: 1) the estimation method (LS, Bayesian); 2) the sampling strategy (uniform, uniform sets, optimal, and clustered optimal); and 3) the noise. A significant main effect of each of the three factors was observed for all measures ($p < 0.001$). Moreover, the interaction between noise and estimation algorithm was also significant ($p < 0.001$), with an increasing effect of the estimation method as noise increases. The interaction between noise and sampling strategy was only significant for the NMSE values of Δt. As expected, parameter estimation errors increase with the noise level in the data. Additionally, the improvement observed for Bayesian versus LS methods and optimal versus uniform strategies, in the case of Δt, also increases with the noise level.

We further observe that the results obtained with two uniform strategies are not significantly different from each other, nor are the ones obtained with the two optimal strategies. In particular, it is interesting to notice that clustering the optimal set of TI points

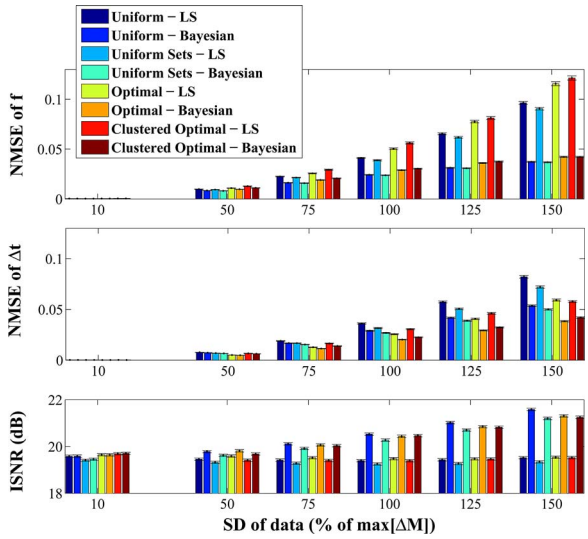

Fig. 9. (Top) NMSE of f and Δt (mean $\pm$ SE) and (bottom) ISNR (mean $\pm$ SE), averaged across parameter space, obtained each method, as a function of noise level.

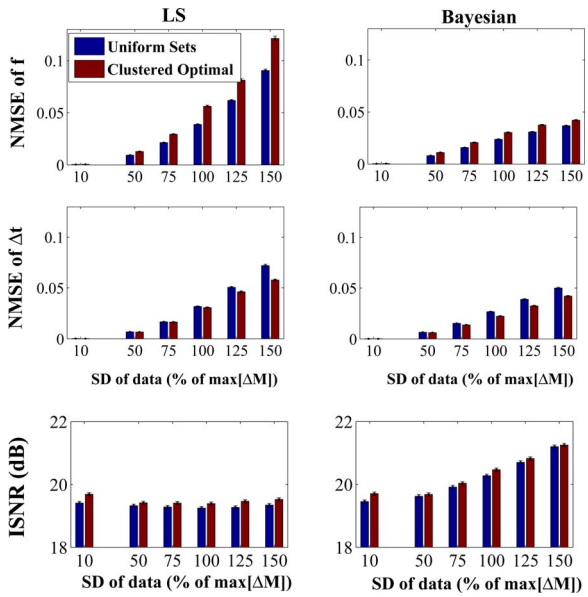

Fig. 10. (Top) NMSE of f and Δt (mean $\pm$ SE) and (bottom) ISNR (mean $\pm$ SE), as a function of noise level: comparison between estimation methods for each sampling strategy.

around five values only does not significantly impair the optimality of the results. Thus, in order to more clearly understand the main effects of the estimation algorithm and the sampling strategy, the results of the Monte Carlo simulations obtained with the clustered optimal and the uniform sets strategies were selected for further analysis. The average values of *NMSE* and *ISNR*, grouped in terms of the sampling strategy and estimation method, are shown in Figs. 10 and 11, respectively.

In Fig. 10, it can be seen that the clustered optimal strategy outperforms the uniform sets strategy in terms of the NMSE values for the parameter Δt, especially at high noise levels. Although slightly greater NMSE values are obtained for f us-

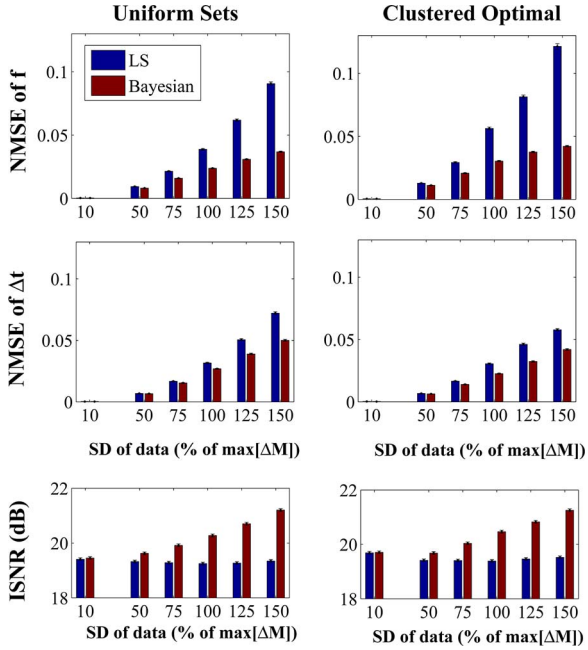

Fig. 11. (Top) NMSE of f and Δt (mean $\pm$ SE) and (bottom) ISNR (mean $\pm$ SE), as a function of noise level: comparison between sampling strategies for each estimation method.

ing the clustered optimal relative to the uniform sets strategy, with the standard LS estimation method, when our proposed Bayesian estimation method is employed this is no longer the case. In Fig. 11, it can be seen that the proposed Bayesian estimation method systematically outperforms the standard LS estimator, in terms of both measurements.

C. Empirical Results

In order to demonstrate the applicability of the proposed sampling and estimation methods, the physiological parameters, perfusion, and ATT were estimated from real multi-TI PASL data. The uniform sets and the optimal clusters sampling strategies were considered only, because of their greater practical feasibility and our previous observation through simulation that the results obtained with these strategies are not significantly different from the corresponding uniform and optimal sampling strategies.

The PASL data were collected from seven healthy volunteers (2 males and 5 females, aged 23–26 years old) on a Siemens Verio 3T system using a 12-channel head RF coil. Pulsed ASL data were acquired at multiple TIs, using saturation pulses to limit the bolus width, with a Q2TIPS-PICORE sequence [26] ($\mathrm{TI}_1/\mathrm{TI}_{1s}/\mathrm{TI}_2 = 750\ \mathrm{ms}/900\ \mathrm{ms}/1700\ \mathrm{ms}$), with gradient-echo echo-planar imaging readout ($\mathrm{TR}/\mathrm{TE} = 2500\ \mathrm{ms}/19\ \mathrm{ms}$), from nine contiguous axial slices positioned parallel to the anterior commissure–posterior commissure line (spatial resolution of $3.5\ \mathrm{mm} \times 3.5\ \mathrm{mm} \times 7.0\ \mathrm{mm}$). Two multiple TI datasets were collected from each subject, at a set of TI_2 points defined by the uniform sets and the optimal clusters sampling strategies, as shown in Fig. 3. Data were preprocessed using FSL tools (www.fmrib.ox.ac.uk/fsl), including motion

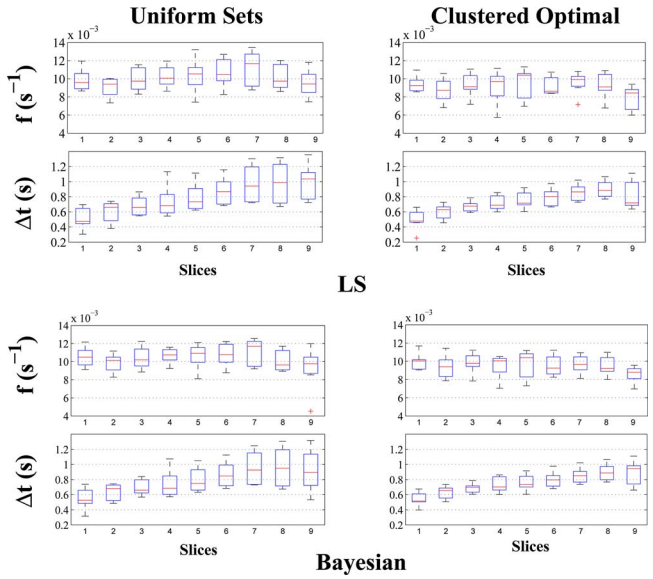

Fig. 12. Box plots of parameter estimates obtained in the subject group for each slice, using the different sampling and estimation methods. Boxes are delimited by 25th and 75th percentiles, whiskers extend to minimum and maximum nonoutliers, + indicates outliers, and red lines indicate the median.

TABLE II
GROUP MEAN ± SD OF EMPIRICAL ESTIMATED PARAMETERS: f VALUES AVERAGED ACROSS SLICES AND Δt VALUES PRESENTED FOR EACH SLICE: Δt_i FOR SLICE i

	Uniform Sets		Clustered Optimal	
$\times 10^{-3}$	LS	Bayesian	LS	Bayesian
$f[s^{-1}]$	10.1±1.4	10.3±1.3	9.1±1.3	9.5±1.1
$\Delta t_1[s]$	512±140	544±141	501±132	529±122
$\Delta t_2[s]$	595±138	633±110	599±096	630±083
$\Delta t_3[s]$	677±125	696±104	672±067	684±062
$\Delta t_4[s]$	743±205	749±179	722±098	735±100
$\Delta t_5[s]$	798±180	801±159	761±112	763±106
$\Delta t_6[s]$	865±181	859±169	794±116	799±103
$\Delta t_7[s]$	955±248	937±222	853±108	850±099
$\Delta t_8[s]$	972±265	959±254	907±110	902±109
$\Delta t_9[s]$	992±228	915±278	829±186	895±165

correction and high-pass temporal filtering with a 70 s frequency cutoff. A T_1-weighted, high-resolution structural image was also obtained from each subject, using an MPRAGE sequence (TR/TE = 2250 ms/2.26 ms, 1 mm × 1 mm × 1 mm). This image was segmented into GM, white matter (WM) and cerebrospinal fluid, using the FSL tool FAST [32] and was subsequently coregistered with the PASL images using FLIRT [33].

Magnetization difference data ΔM were obtained as the differences between consecutive control-tag pairs of images, averaged across GM in each brain slice, normalized by the mean of all control images. The amount of noise corrupting the data σ_Y was obtained as the standard deviation of the ΔM data in a region defined in the background outside the brain, with a number of pixels comparable to that of the respective GM region. Both the LS and Bayesian methods were employed for the estimation of the parameters f and Δt from the data collected using the two sampling strategies. The distributions of the results obtained for the group of subjects are presented in Fig. 12 and the corresponding average values are shown in Table II.

TABLE III
INTERSUBJECT CVS

	Uniform Sets		Clustered Optimal	
	LS	Bayesian	LS	Bayesian
$CV_f[\%]$	14.5	13.2	14.9	11.8
$CV_{\Delta t}[\%]$	31.0	28.3	22.4	20.1

In general, the values of the estimated parameters are consistent across sampling strategies and estimation methods. The perfusion results are around the expected value of $f = 0.012\ \text{s}^{-1}$ and do not vary significantly across slice. Regarding the ATT, the results are also around the expected value of $\Delta t = 0.7$ s and increase from inferior to superior slices, due to the fact that the blood is traveling upward across the brain. In terms of the estimation method, the Bayesian approach produced slightly higher perfusion values relative to the LS approach, as a consequence of the influence of the prior value for this parameter (which was higher than the values found in the data). In terms of the sampling strategy, it can be seen that the ATT values vary more widely when using the uniform sets approach compared with the optimal clusters approach. This is likely the result of the finer sampling of the curve around the relevant TI values for the estimation of Δt, in the case of an optimal sampling scheme.

In order to evaluate the variability of the measurements obtained with the various methodologies, the intersubject coefficient of variation (CV) was determined, for each parameter θ_i, sampling strategy, and estimation method, as $\text{CV}_{\theta_i} = \frac{\sigma_{\theta_i}}{\overline{\theta_i}}$, where σ_{θ_i} and $\overline{\theta_i}$ are the standard deviation and mean of parameter θ_i estimates across subjects, and also across slices in the case of parameter f, respectively. The final value of $\text{CV}_{\Delta t}$ was calculated as the mean $\text{CV}_{\Delta t}$ values across slices, because the value of Δt is dependent on the slice. The CV measure represents the dispersion of a probability distribution, giving the information about the variability of the measurement [34].

The CV values obtained for the subject group are shown in Table III. All CV values are below 33%, which is considered to represent an acceptable variability in a normal distribution [34]. It can be observed that CV_f is lower than $\text{CV}_{\Delta t}$, which likely results from the variation of Δt with the slice. Most interestingly, the variability of Δt values is considerably reduced when using the clustered optimal sampling strategy relative to the uniform sets sampling strategy. Moreover, the variability is, in general, reduced for the estimations using the Bayesian method relative to the LS method.

IV. DISCUSSION AND CONCLUSION

A Bayesian framework was proposed for the optimization of perfusion and ATT measurement using PASL, by selecting an optimal sampling strategy and a parameter estimation method, based on the structure of a standard kinetic model as well as on the information about the physiological distribution of the parameters and the measurement noise.

We showed that the optimal distribution of sampling points depends critically on the prior knowledge of the ATTs, on the uncertainty of the model parameters, and, to a lesser extent, on the level of noise in the data. Through Monte Carlo simulations,

we showed that parameter estimation errors were reduced by combining the optimal sampling strategy with Bayesian parameter estimation, compared with uniform sampling and standard least squares. When applied to real PASL data, the proposed methodology produced the lowest intersubject CVs, which may have great impact on the clinical applicability of the technique, by improving its sensitivity to detect pathology and also as a probe in longitudinal studies monitoring disease progression.

This paper adds to previous reports on the optimization of PASL sampling schemes [4], [5], by incorporating *a priori* knowledge of the model parameters within a Bayesian framework. Furthermore, it adds to existing Bayesian approaches for PASL model estimation [14], [17], [35], by addressing the sampling strategy optimization within the same framework.

A major limitation of our method is that the optimal sampling strategy is critically dependent on the prior knowledge of the parameters that may change significantly in disease. Here, the strategy is applied to a healthy population. For pathological populations, the parameter distributions would have to be adapted according to the specific clinical hypothesis. In order to overcome this limitation, Xie *et al.* implemented a real-time adaptive process that iteratively updates the parameter estimates and adjusts the optimal sampling schedule accordingly, as data are collected [5]. In this way, their method becomes applicable in clinical situations, when no prior knowledge on perfusion and ATT can be assumed. Our approach could also be implemented in a similar way with the appropriate adaptations.

Another limitation is the assumption of a Gaussian distribution for the ATTs, based on considering a single tissue type. In fact, ATTs are known to vary between GM and WM, between watershed and central areas of the arterial territories, as well as in pathological regions. In order to consider more than one tissue type, a mixture of distributions could be used. In this case, the optimal sampling strategy obtained would be a compromise between the optimal schemes for each of the individual tissue type. Although considering a more realistic mixture model could improve the methodology, we believe that in some cases it would still be desirable to target the experimental design to one specific tissue type. In fact, the main distinctions are between GM and WM, or between healthy and pathological tissue. Because PASL data from WM are well known to exhibit a critically low SNR due to lower perfusion and longer arrival times [36], [37], in case WM perfusion is of interest it would be recommendable to perform a separate acquisition specifically aimed at the longer transit times of this tissue in order to optimize detection sensitivity. A similar approach could be taken for patient populations exhibiting very prolonged transit times.

In conclusion, optimal sampling and estimation methodologies were proposed for improved perfusion quantification using PASL imaging relative to conventional approaches. It is expected that these methodologies will contribute in enhancing the applicability of PASL perfusion imaging.

REFERENCES

[1] E. T. Petersen, I. Zimine, Y.-C. Ho, and X. Golay, "Non-invasive measurement of perfusion: A critical review of arterial spin labelling techniques," *Br. J. Radiol.*, vol. 79, no. 944, pp. 688–701, 2006.

[2] 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," *Magn. Reson. Med.*, vol. 40, no. 3, pp. 383–396, 1998.

[3] P. M. Figueiredo, S. Clare, and P. Jezzard, "Quantitative perfusion measurements using pulsed arterial spin labeling: Effects of large region-of-interest analysis," *J. Magn. Reson. Imag.*, vol. 21, no. 6, pp. 676–682, Jun. 2005.

[4] J. Xie, D. Gallichan, R. N. Gunn, and P. Jezzard, "Optimal design of pulsed arterial spin labeling MRI experiments," *Magn. Reson. Med.*, vol. 59, no. 4, pp. 826–834, Apr. 2008.

[5] J. Xie, S. Clare, D. Gallichan, R. N. Gunn, and P. Jezzard, "Real-time adaptive sequential design for optimal acquisition of arterial spin labeling MRI data," *Magn. Reson. Med.*, vol. 64, no. 1, pp. 203–210, 2010.

[6] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise.* Melbourne, FL: Krieger, 1992.

[8] J. Sanches, I. Sousa, and P. Figueiredo, "Bayesian fisher information criterion for sampling optimization in ASL-MRI," in *Proc. IEEE Int. Symp. Biomed. Imag. 2010*, pp. 880–883.

[9] N. Santos, J. Sanches, and P. Figueiredo, "Bayesian optimization of perfusion and transit time estimation in PASL-MRI," in *Proc. IEEE Eng. Med. Biol. Conf.*, Aug. 31–Sep. 4, 2010, pp. 4284–4287.

[10] C. R. Vogel, *Computational Methods for Inverse Problem*, (Frontiers in Applied Mathematics), 1st ed. Philadelphia, PA SIAM, 2002.

[11] A. Bouhamidi and K. Jbilou, "Sylvester Tikhonov-regularization methods in image restoration," *J. Comput. Appl. Math.*, vol. 206, no. 1, pp. 86–98, 2007.

[12] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[13] J. M. Sanches, J. C. Nascimento, and J. S. Marques, "Medical image noise reduction using the Sylvester–Lyapunov equation," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1522–1539, Sep. 2008.

[14] A. R. Groves, M. A. Chappell, and M. W. Woolrich, "Combined spatial and non-spatial prior for inference on MRI time-series," *Neuroimage*, vol. 45, no. 3, pp. 795–809, 2009.

[15] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.

[16] B. J. MacIntosh, N. Filippini, M. A. Chappell, M. W. Woolrich, C. E. Mackay, and P. Jezzard, "Assessment of arterial arrival times derived from multiple inversion time pulsed arterial spin labeling MRI," *Mag. Reson. Med.*, vol. 63, no. 3, pp. 641–647, 2010.

[17] M. A. Chappell, B. J. MacIntosh, M. J. Donahue, M. Gnther, P. Jezzard, and M. W. Woolrich, "Separation of macrovascular signal in multi-inversion time arterial spin labelling MRI," *Magn. Reson. Med.*, vol. 63, no. 5, pp. 1357–1365, 2010.

[18] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1988.

[19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Chapter Modeling of Data—Nonlinear Models*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1988–1992, pp. 45–58.

[20] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York: Springer-Verlag, 2001.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. 2, *Detection Theory*, 1st ed. Englewood Cliffs, NJ: Prentice-Hall, Jan. 1998.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, Mar. 2004.

[23] R. B. Buxton, "Quantifying CBF with arterial spin labeling," *J. Magn. Reson. Imaging*, vol. 22, no. 6, pp. 723–726, 2005.

[24] H. P. Wynn and N. Parkin, "Sensitivity analysis and identifiability for differential equation models," in *Proc. 40th IEEE Conf. Decision Control*, 2001, vol. 4, pp. 3116–3121.

[25] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, pp. 1923–1929, Aug. 2009.

[26] 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," *Mag. Reson. Med.*, vol. 41, no. 6, pp. 1246–1254, 1999.

[27] H. Lu, C. Clingman, X. Golay, and P. C. V. Zijl, "Determining the longitudinal relaxation time (T1) of blood at 3.0 Tesla," *Magn. Reson. Med.*, vol. 52, pp. 679–682, 2004.

[28] 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," *Magn. Reson. Med.*, vol. 61, pp. 899–906, 2009.

[29] G. A. F. Seber, *Multivariate Observations*. Hoboken, NJ: Wiley, 1984.

[30] H. Spath, *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*. New York: Halsted, 1985.

[31] M. Günther, K. Oshio and D. A. Feinberg, "Single-shot 3D imaging techniques improve arterial spin labeling perfusion measurements," *Magn. Reson. Med.*, vol. 54, pp. 491–498, 2005.

[32] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.

[33] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, 2001.

[34] S. Magon, G. Basso, P. Farace, G. K. Ricciardi, A. Beltramello, and A. Sbarbati, "Reproducibility of BOLD signal change induced by breath holding," *NeuroImage*, vol. 45, pp. 702–712, 2009.

[35] M. A. Chappell, A. R. Groves, B. Whitcher, and M. W. Woolrich, "Variational Bayesian inference for a nonlinear forward model," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 223–236, Jan. 2009.

[36] M. J. P. van Osch, W. M. Teeuwisse, M. A. A. van Walderveen, J. Hendrikse, D. A. Kies, and M. A. van Buchem, "Can arterial spin labeling detect white matter perfusion signal?" *Magn. Reson. Med.*, vol. 62, no. 1, pp. 165–173, 2009.

[37] R. Pohmann, "Accurate, localized quantification of white matter perfusion with single-voxel ASL," *Magn. Reson. Med.*, vol. 64, no. 4, pp. 1109–1113, 2010.

Authors' photographs and biographies not available at the time of publication.