HAL open science

Bayesian inverse regression for vascular magnetic resonance fingerprinting

Fabien Boux, Florence Forbes, Julyan Arbel, Benjamin Lemasson, Emmanuel L. Barbier

To cite this version:

Fabien Boux, Florence Forbes, Julyan Arbel, Benjamin Lemasson, Emmanuel L. Barbier. Bayesian inverse regression for vascular magnetic resonance fingerprinting. 2020. hal-02314026v2

HAL Id: hal-02314026

https://hal.science/hal-02314026v2

Preprint submitted on 12 May 2020 (v2), last revised 17 Mar 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian inverse regression for vascular magnetic resonance fingerprinting

Fabien Boux, Florence Forbes, Julyan Arbel, Benjamin Lemasson and Emmanuel L. Barbier *†‡

May 12, 2020

Abstract

Standard parameter estimation from vascular magnetic resonance fingerprinting (MRF) data is based on matching the MRF signals to their best counterparts in a grid of coupled simulated signals and parameters, referred to as a dictionary. To reach a good accuracy, the matching requires an informative dictionary whose cost, in terms of design, storage and exploration, is rapidly prohibitive for even moderate numbers of parameters. In this work, we propose an alternative dictionary-based learning (DBL) approach made of three steps: 1) a quasi-random sampling strategy to produce efficiently an informative dictionary, 2) an inverse statistical regression model to learn from the dictionary a correspondence between fingerprints and parameters, and 3) the use of this mapping to provide both parameter estimates and their confidence indices. Our DBL method is first compared to MRF matching on two types of synthetic signals: scalable and vascular MRF signals. On scalable signals, quasi-random sampling outperforms the grid when using DBL. Dictionaries up to 100 times smaller than in MRF matching, yield a 12 % decreased error on parameter estimates. The confidence indices match the parameter estimation errors ($R^2 = 0.95$). Then, on vascular signals, dictionary-based methods yield more accurate estimates than the conventional, closed-form expression fitting method with significantly smaller errors on vessel size estimates. On real vascular MRF signals acquired from tumor bearing rats, the DBL method shows less noisy maps than MRF matching. Our DBL proposal effectively reduces the number of simulations required and speeds up parameter estimation, while providing more accurate estimates with their confidence indices.

1 Introduction

Magnetic resonance fingerprinting (MRF) is a novel approach to quantitative magnetic resonance imaging that allows the estimation of multiple tissue properties in a single acquisition [1, 2]. The acquisition, which consists in repeating measurements with varying experimental conditions, generates a signal evolution (or *fingerprint*) that depends on the parameters of the studied tissue. To

*This work was supported by the French National Research Agency - project cerebrovascular dynamics in epilepsy: endothelial-pericyte interface – Epicyte, ANR-16-CE37-0013. This work was performed on the IRMaGe platform member of France Life Imaging network (grant ANR-11-INBS-0006). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

†F. Boux, B. Lemasson and E. L. Barbier are with Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, France (phone: +33-4-56-52-05-88, e-mail: emmanuel.barbier@univ-grenoble-alpes.fr).

‡F. Boux, F. Forbes and J. Arbel are with Univ. Grenoble Alpes, Inria, CNRS, G-INP, 38000 Grenoble, France.

estimate these parameters, a large database, referred to as a *dictionary* and containing a large number of possible signal evolutions, is simulated from biophysical models. A comparison is performed between an acquired signal and the signals in the dictionary to find the best match according to an objective function. The tissue parameters are then estimated to the values that generated the best signal evolution match. In MRF, parameter estimation accuracy therefore depends on the number of dictionary entries, which increases exponentially with the number of parameters. For applications with many parameters such as vascular MRF [3], the required memory size and simulation time as well as the parameter estimation time (or reconstruction time) quickly become a limit.

To compress the dictionary while limiting the loss of information, several authors have used singular value decomposition to project the dictionary in a well-chosen subspace [4–8]. However, this compression procedure generally decreases parameter accuracy. It has also been proposed to directly find a mapping from the fingerprints to the parameter space using kernel regression [9], maximum likelihood approach [10] or neural network approaches [11–18]. The resulting compact representation offers the advantage over the discrete MRF grid of a continuous exploration of parameter values. These approaches significantly reduce the reconstruction time, but not the simulation time due to the need to span a high dimensional fingerprint space. To limit the simulation time, Cohen *et al.* [14] studied a mapping obtained from a sparse set of dictionary entries. The study, carried out with only two parameters, led to a modest reduction of dictionary entries (up to 60). Consider a dictionary of 10×10 entries simulated in 1 hour. If the number of parameters increases from 2 to 7 parameters, and always considering 10 values per parameter, then the dictionary computation time increases from 1 hour to more than 11 years. In this case, an approach that greatly reduces the need for simulation, continuously represents the parameters without loss of precision, relies on an explainable model and reduces the reconstruction time becomes highly desirable [19].

To reach this goal, we adopt in this work a mapping approach that circumvents the difficulty of learning a *high-to-low* mapping from a high dimensional fingerprint space to a low dimensional parameter space, learning instead the much less problematic *low-to-high* reverse mapping from parameters to fingerprints. More specifically, we use the Gaussian locally linear mapping (GLLiM) model [20], which allows both a tractable learning of the low-to-high mapping and a subsequent analytical expression of the high-to-low or signal-to-parameter mapping. Furthermore, unlike most other regression methods that focus on pointwise predictions, GLLiM provides a full posterior distribution per fingerprint. This distribution can then be used to compute an estimated value and a confidence index for each parameter, using respectively the posterior expectation and standard deviation.

In this vascular MRF study, the proposed dictionary-based learning (DBL) method and the standard dictionary-based matching (DBM) method are compared. Synthetic scalable signals are first used to assess quantitatively the methods' performance while increasing the number of parameters. Vascular MRF signals are then considered both through simulations and real data acquired in tumor bearing rats.

2 MRF as an inverse problem

In inverse problems, the overall issue is to provide information on some parameters of interest $\mathbf{x}$ given an observed signal $\mathbf{y}$, using a known *direct* or *forward model* that describes how the parameters $\mathbf{x}$ translate into a signal $\mathbf{y}$. Among inverse problems, MRF exhibits the following difficulties: 1) the direct model is (highly) non-linear, as a (complex) series of equations or simulation tools; 2)

the $\mathbf{y}$'s are high-dimensional signals and 3) many $\mathbf{y}$'s need to be inverted (one for each voxel in an image); 4) the vector of parameters $\mathbf{x}$ is multidimensional and predicting each component of $\mathbf{x}$ independently is likely to be sub-optimal.

To account for possible sources of uncertainty, we focus on a statistical modeling assuming that the forward model is described by a likelihood and a prior distribution. The likelihood function is linking parameter values $\mathbf{x}$ to a probability of observing signal $\mathbf{y}$, $\mathcal{L}_{\mathbf{x}}(\mathbf{y}) = p(\mathbf{y}|\mathbf{x})$. A natural assumption is that $\mathcal{L}_{\mathbf{x}}(\mathbf{y})$ is a Gaussian distribution $\mathcal{N}(\mathbf{y}; f(\mathbf{x}), \mathbf{\Sigma})$ centered at $f(\mathbf{x})$ where f is the known simulation function that links the physical and physiological parameters to the fingerprint and $\mathbf{\Sigma}$ is a covariance matrix accounting for measuring or modelling imperfections. The parameter prior distribution, denoted by $p(\mathbf{x})$, encodes in turns information on the possible parameter values. Standard MRF uses a finite grid of values, which corresponds to a very particular discrete prior. This probabilistic point of view allows, with the Bayesian framework, to derive a posterior distribution $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$, which provides for any given $\mathbf{y}$, a characterization of $\mathbf{x}$ by a probability density function more informative than a single point prediction of $\mathbf{x}$. It corresponds to a richer inverse model but is not usually available in closed-form and requires approximations to be usable in practice.

More generally, most methods to solve inverse problems can be classified into two main categories, optimization-based and learning-based methods. In the next section, we refer to standard MRF as a matching method. We show that it can be seen as a penalized optimization, which does not require statistical modeling, while the method we propose next belongs to statistical learning approaches.

2.1 Dictionary-based matching (DBM) method

MRF [1] requires a large database $\mathcal{D}_f$, referred to as a dictionary. It is made of N entries of coupled fingerprint and parameters $(\mathbf{x}, \mathbf{y})$. The S-dimensional fingerprints $\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ are generated by running the simulation model f for N different values of the P-dimensional magnetic and physiological parameters $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. In the DBM method, a P-dimensional grid is generated with sampled values in a pre-set interval for each parameter. Then, to invert an observed $\mathbf{y}_{\text{obs}}$, it is compared with the signals in $\mathcal{D}_f$ to find the best match according to an objective function $\mathrm{d}(\cdot, \cdot)$, usually a standard distance or dissimilarity measure (*e.g.* in MRF, the dot product). With $\mathcal{D}_f = \{(\mathbf{x}_n, \mathbf{y}_n = f(\mathbf{x}_n)), n = 1{:}N\}$, $\mathbf{x}$ is thus estimated as the argument of the following minimization:

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathcal{D}_f}{\arg\min}\ \mathrm{d}(\mathbf{y}_{\text{obs}}, f(\mathbf{x})). \tag{1}$$

Solutions are sought in $\mathcal{D}_f$ only, while in a non-constrained optimization the minimization is over the whole continuous space of parameter values. The performance of the method depends directly on the space discretization *i.e.* the choice of the number of dictionary entries and the number of parameters. The larger the number N of entry $(\mathbf{x}_n, \mathbf{y}_n)$, the more accurate the estimates but the larger the simulation time and memory requirement. Even for moderate number of parameters, the required number of elements in the dictionary renders grid search intractable on a desktop computer. In addition, each new $\mathbf{y}_{\text{obs}}$, requires the computation and comparison of N matching scores $\mathrm{d}(\mathbf{y}_{\text{obs}}, \mathbf{y}_n)$, which can be costly if N is very large and if many inversions are desired. The regression or learning method that we propose in the next section is more efficient with respect to these aspects.

2.2 Proposed dictionary-based learning (DBL) method

In contrast to the DBM method, regression and learning methods can adapt to handle massive inversions of high dimensional data. The main principle is to transfer the computational cost, from 2-signal matchings to the learning of an inverse operator $\mathcal{F}^{-1}$. Equivalently, the goal is to learn a mapping from the fingerprint space to the parameter space, for any $\mathbf{y}$, with cost-less evaluation of $\mathcal{F}^{-1}(\mathbf{y})$. The dictionary $\mathcal{D}_f$ can be used to estimate $\mathcal{F}^{-1}$. Learning or regression methods adapted to high dimensions include inverse regression methods, *i.e.* sliced inverse regression [21], partial least squares [22], approaches based on mixtures of regressions with different variants, *e.g.* Gaussian locally linear mapping (GLLiM) [20], mixtures of experts [23], cluster weighted models [24], and kernel methods [9]. Inverse regression methods are flexible in that they reduce the dimension in a way optimal to the subsequent mapping estimation task that can itself be carried out by any kind of standard regression tool. In that sense, the inverse regression methods are said to be non-parametric or semi-parametric. Similarly, in [9], the authors propose a regression with an appropriate kernel function to learn the non-linear mapping. The procedure has the advantage to be semi-parametric but a serious limit is that the components of f are optimized in each dimension separately. As regards application to MRF, the learning strategy has also been proposed by several groups using deep learning tools [11–18]. A major limitation of these methods is that they require a large number of training points to learn many model parameters without overfitting.

In the same vein as [9], and in contrast to deep learning approaches, we propose to use the GLLiM method that exploits Gaussian mixture models [20]. Compared to other regression methods that focus on providing point-wise estimates, GLLiM provides a full probability distribution selected in a family of parametric models, *e.g.* mixture of Gaussian distributions, where the parameters are denoted by $\boldsymbol{\theta}$. The inversion operator is defined as $\mathcal{F}^{-1}(\mathbf{y}) = p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is estimated from the dictionary. More specifically, GLLiM handles the modeling of non-linear relationships with a piecewise linear model. Each $\mathbf{y}$ is seen as the noisy image of $\mathbf{x}$ obtained from a K-component mixture of affine transformations. This is modeled by introducing a latent variable $z \in \{1, \dots, K\}$ such that

$$\mathbf{y} = \sum_{k=1}^{K} \delta_k(z)\,(\boldsymbol{A}_k\mathbf{x} + \boldsymbol{b}_k + \boldsymbol{\epsilon}_k), \tag{2}$$

where $\delta_k(z)$ indicates membership in the region k of $\mathbf{x}$, having the value 1 if it belongs to the region and the value 0 otherwise. $\boldsymbol{A}_k$ is a $P \times S$ matrix and $\boldsymbol{b}_k$ a vector in $\mathbb{R}^P$ that characterize an affine transformation. Variable $\boldsymbol{\epsilon}_k$ corresponds to an error term in $\mathbb{R}^P$ which is assumed to be zero-mean and not correlated with $\mathbf{x}$, capturing both the modelling noise and the reconstruction error due to the affine approximations. In GLLiM, $\boldsymbol{\epsilon}_k$ follows a Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_k)$ and $\mathbf{x}$ follows a mixture of K Gaussians defined by $p(\mathbf{x}|z=k) = \mathcal{N}(\mathbf{x}; \boldsymbol{c}_k, \boldsymbol{\Gamma}_k)$, and $p(z=k) = \pi_k$. It follows that

$$p(\mathbf{y}|\mathbf{x};\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k(\mathbf{x})\,\mathcal{N}(\mathbf{y}; \boldsymbol{A}_k\mathbf{x} + \boldsymbol{b}_k, \boldsymbol{\Sigma}_k) \tag{3}$$

$$\text{with } w_k(\mathbf{x}) = \frac{\pi_k\,\mathcal{N}(\mathbf{x}; \boldsymbol{c}_k, \boldsymbol{\Gamma}_k)}{\sum_{j=1}^{K} \pi_j\,\mathcal{N}(\mathbf{x}; \boldsymbol{c}_j, \boldsymbol{\Gamma}_j)}$$

and $\boldsymbol{\theta} = \{\pi_k, \boldsymbol{c}_k, \boldsymbol{\Gamma}_k, \boldsymbol{A}_k, \boldsymbol{b}_k, \boldsymbol{\Sigma}_k\}_{k=1:K}$ is the set of parameters defining the model. The conditional probability distribution of interest can be derived as

$$p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k^*(\mathbf{y})\,\mathcal{N}(\mathbf{x};\boldsymbol{A}_k^*\mathbf{y}+\boldsymbol{b}_k^*,\boldsymbol{\Sigma}_k^*) \tag{4}$$
$$\text{with } w_k^*(\mathbf{y}) = \frac{\pi_k\,\mathcal{N}(\mathbf{y};\boldsymbol{c}_k^*,\boldsymbol{\Gamma}_k^*)}{\sum_{j=1}^{K}\pi_j^*\,\mathcal{N}(\mathbf{y};\boldsymbol{c}_j^*,\boldsymbol{\Gamma}_j^*)}$$

and a new parameterization $\boldsymbol{\theta}^* = \{\boldsymbol{c}_k^*, \boldsymbol{\Gamma}_k^*, \boldsymbol{A}_k^*, \boldsymbol{b}_k^*, \boldsymbol{\Sigma}_k^*\}_{k=1:K}$ easily expressed as an analytical function of $\boldsymbol{\theta}$. The mixture setting provides some guaranties that when choosing K large enough it is possible to approximate any reasonable relationship [23]. Automatic model selection criteria can also be used to select K (see [20]).

The $p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta})$ distribution provides both estimates of the parameters $\mathbf{x}$ and information about the confidence to be placed in these estimates. In this work, estimates are defined through the expectation and the confidence indices as the square root of the covariance matrix diagonal element vector:

$$\hat{\mathbf{x}} = \mathbb{E}\left[\mathbf{x}|\mathbf{y};\boldsymbol{\theta}\right], \tag{5}$$
$$\text{CI} = \sqrt{\text{diag}\left(\text{Var}\left[\mathbf{x}|\mathbf{y};\boldsymbol{\theta}\right]\right)}, \tag{6}$$

with $\mathbb{E}\left[\mathbf{x}|\mathbf{y};\boldsymbol{\theta}\right] = \sum_{k=1}^{K} w_k^*(\mathbf{y})(\boldsymbol{A}_k^*\mathbf{y}+\boldsymbol{b}_k^*)$, and

$$\text{Var}\left[\mathbf{x}|\mathbf{y};\boldsymbol{\theta}\right] = \sum_{k=1}^{K} w_k^*(\mathbf{y})\left[\boldsymbol{\Sigma}_k^* + (\boldsymbol{A}_k^*\mathbf{y}+\boldsymbol{b}_k^*)(\boldsymbol{A}_k^*\mathbf{y}+\boldsymbol{b}_k^*)^T\right] - \left(\sum_{k=1}^{K} w_k^*(\mathbf{y})(\boldsymbol{A}_k^*\mathbf{y}+\boldsymbol{b}_k^*)\right)\left(\sum_{k=1}^{K} w_k^*(\mathbf{y})(\boldsymbol{A}_k^*\mathbf{y}+\boldsymbol{b}_k^*)\right)^T,$$

where diag$(\cdot)$ denotes the function returning the diagonal elements of a matrix. For the CI, computed from the estimated posterior $p(\mathbf{x}|\mathbf{y};\boldsymbol{\theta})$, to be a good indicator of the parameter estimation error, it is required that the inverted $\mathbf{y}$ follows the same model used to computed $\boldsymbol{\theta}$. The use of a unique $\boldsymbol{\theta}$ parameter for all inversions provides a great gain when massive inversions are required but it also assumes that the same model is valid for all fingerprints and that the dictionary $\mathcal{D}_f$ is a good representation of them. In practice, acquired fingerprints may come with different noise levels. An interesting feature of GLLiM is to adapt to this case at a very low cost. When the observed $\mathbf{y}$ comes with some covariance matrix $\boldsymbol{\Sigma}_\eta$ corresponding to a centered Gaussian noise variable $\boldsymbol{\eta}$, the initial dictionary $\mathcal{D}_f$ may not be fully adapted if it has not been generated with this same additional measurement error. Another training set should be simulated and used instead, with a corrected likelihood corresponding to $\mathcal{N}(\mathbf{y}; f(\mathbf{x}), \boldsymbol{\Sigma}+\boldsymbol{\Sigma}_{\boldsymbol{\eta}})$. Fortunately, it is straightforward to check that the structure of the Gaussian mixture approximation avoid the re-learning of the GLLiM model. Indeed, it suffices to change the estimated $\boldsymbol{\Sigma}_k$'s into $\boldsymbol{\Sigma}_k + \boldsymbol{\Sigma}_\eta$ and to report this change when computing $\boldsymbol{\theta}^*$.

Because S is much larger than P in MRF applications, it is important that the model (3) involving $\boldsymbol{\theta}$ is estimated first and then used to derive model (4) that has a similar structure. The number of model parameters $\boldsymbol{\theta}$ can be drastically reduced by choosing constraints on covariance matrices $\boldsymbol{\Sigma}_k$ without inducing oversimplifications on the target model (4). In this work, equal diagonal covariance matrices are used as they yield the best results: for $1 \leq k \leq K$, $\boldsymbol{\Sigma}_k = \mathbf{D}_S$, where $\mathbf{D}_S \in \mathbb{R}^{S\times S}$ is a diagonal matrix. For example, with $S = 100$, $P = 3$ and $K = 50$, the number of parameters $\boldsymbol{\theta}$ is equal to 20 600 while a direct estimation of $\boldsymbol{\theta}^*$ would involve 272 703 parameters (see [20] for more details).

2.3 Dictionary sampling strategy

The dictionary design depends on the sampling strategy of the parameter space. In MRF, regular grids of P-dimensional parameter values are generally considered. In [16], authors show that in a regression context, the random sampling strategy provides better estimation of the parameters than the use of a regular grid. However, this strategy entails a risk of imperfectly covering the parameter space coverage.

Fig. 1a shows a two-dimensional projection of $N = 1\,000$ points from a uniform grid in the 3D-hypercube ($P = 3$). Each parameter is described by 10 separate values. Note that with 1 000 points in 3D, only 100 distinct combinations appear in the 2D projection plane, each representing 10 different values of the third variable. This sampling scheme is not optimal in terms of information content. A significant improvement over the grid can be achieved by scrambled nets [25, 26]. In this paper, the Sobol sequence is generated [27] and scrambled [28]. We show the projection of $N = 1\,000$ points from the scrambled Sobol sequence (fig. 1c) referred to as quasi-random in the remainder of the manuscript.

3 Analysis framework

3.1 Signals

3.1.1 Synthetic scalable signals

The sensitivity of the MRF signals to each parameter is variable. In addition, parameters cannot readily be added to the simulation tool that produces the MRF or the vascular MRF signals. To produce signals that are equally sensitive to each parameter and dependent on a variable number of parameters (*i.e.* P may be set to any value), scalable signals that mimic MRF signals are introduced in equation (7). The parameters of the synthetic scalable signal have physical units to help understand their structure but no physical meaning,

$$\mathbf{y} = \left| \sum_{i=1}^{P} \sin\left(50\,\phi_i\,t\right) \exp\left(-\frac{t}{x_i}\right) \right|, \tag{7}$$

where x_i are the elements of $\mathbf{x}$, t varies from 10 to 1 000 ms in 10 ms steps ($S = 100$), the $\boldsymbol{\phi}_i$ values are between 0.1 and 1 and $|\cdot|$ is the absolute value function. The values of parameters $\mathbf{x}$ are in the range of 10 to 1 000 ms. The vector $\boldsymbol{\phi}$ is defined randomly such that none of the terms are equal. This makes the parameters x_i non-exchangeable: permutations of the $\mathbf{x}$ elements cannot lead to the same signal $\mathbf{y}$. Note that the relationship between $\mathbf{x}$ and $\mathbf{y}$ is non-linear. Examples of synthetic scalable signals are given in Supporting Information, fig. S2.

To create a noisy signal, a Gaussian zero-mean random variable with standard deviation σ_{noise} is added to the complex signal $\mathbf{y}$. The absolute value of the noisy signal is then considered. The signal-to-noise ratio is defined as: $\text{SNR} = I_{\max}/\sigma_{\text{noise}}$, where $I_{\max}$ is the maximum signal intensity. The same procedure is used to add noise to the following signals.

3.1.2 Synthetic vascular MRF signals

Vascular MRF signals are ratio of the gradient echo sampling of the free induction decay and spin echo (GESFIDSE) signals measured pre- and post-injection of ultrasmall superparamagnetic iron

oxide particles (USPIO) [3]. Eight sampled time points are obtained after the 90-degree pulse and 24 sampled time points after the 180-degree pulse ($S = 32$). These signals mainly depend on the vascular properties of the tissues, which in our application are specified by three parameters ($P = 3$): blood volume fraction (BVf), vessel size index (VSI) and tissue oxygen saturation (StO_2). The simulation tool [29] takes into account intrinsic relaxations, magnetic field perturbations induced by susceptibility interfaces (vessels), water proton diffusion and compartmentalization of the contrast agent in the vessels. Due to the complexity of the tool, simulations are extremely time-consuming. Simulation of a single synthetic vascular MRF signal takes about 10 seconds and a dictionary of 100 000 signals is generated on a 32-node high-performance computer (*Intel Xeon Gold 6130*, 2.1 GHz) in about 67 hours.

3.1.3 Acquired vascular MRF signals

Experimental data were acquired at 4.7 T (Bruker Biospin, Ettlingen, Germany) and have been introduced in [30]. The field of view was 30×30 mm^2 and the voxel size was $234 \times 234 \times 800$ μm^3. A turbo spin-echo sequence was acquired to identify anatomical structures and tumor tissues. Then, two GESFIDSE sequences ($S = 32$) were acquired, before and after injection of UPSIO. See Appendix I for more details on the conditions of animal preparation and data acquisition.

3.2 Analysis pipeline

The simulated and acquired data are processed using custom code developed in the Matlab environment (The MathWorks Inc., Natick, Ma, USA). This code and the numerical experiment scripts are available[1]. Data from tumor bearing rats are processed using the *Medical software for Processing multi-Parametric images Pipelines*[2].

3.2.1 Dictionary design

The dictionary is generated in two steps. First, combinations of parameter values in the parameter space are sampled using one of the sampling strategies in section 2.3. Then, for each combination of parameter values, the associated fingerprint is simulated using either equation (7) for synthetic scalable signals or the simulation tool described in section 3.1.2 for vascular MRF signals. For the DBL method, a low level, zero-mean Gaussian noise (typically SNR = 60) is added to the dictionary signals as this improved our results (see Supporting Information, fig. S3).

3.2.2 Dictionary-based analysis

The dictionary is fully stored for the DBM method or summarized by a parametric model $\boldsymbol{\theta}$ for the DBL method. To obtain this model, we use the GLLiM regression described in section 2.2. The model learning, a potentially time-consuming step, is performed only once, just after the production of the dictionary. The model requires only the setting of the K calibration value. In practice, the precise K value is not critical and different K values give similar results as long as they are sufficiently large compared to number of dictionary entries ($K \geq 50$ in our study).

In DBM, given an observed signal $\mathbf{y}_{\text{obs}}$, an estimate $\hat{\mathbf{x}}$ of the true $\mathbf{x}_{\text{obs}}$ is calculated as the minimization argument of equation (1) among the couples $(\mathbf{x}, \mathbf{y})$ in the dictionary. The observed

[1] https://github.com/nifm-gin/DBL-qMRI
[2] https://github.com/nifm-gin/MP3

signal and the signals in the dictionary are previously normalised to have unit Euclidean norm. The parameters are normalized to have a mean of zero and unit variance using scaling and translating factors that are then used to rescale the estimates.

In DBL, an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_{\text{obs}}$ is computed using equation (5) and a confidence index (CI) using equation (6). To obtain an accurate CI, an estimation of the signal noise variance is required. This estimate can be derived from the data SNR and then used as explained in section 2.2 to update $\boldsymbol{\theta}$ adequately.

3.2.3 Closed-form expression fitting (CEF) analysis

Vascular MRF signals can also be analyzed by fitting of a non linear biophysical model [3,31]. The closed-form expression fitting (CEF) analysis method refers to this multiple-operation procedure. First, relaxation rates are extracted by fitting the intensities of MRI signals (synthetic or acquired). Then, these relaxation rates are used to compute the BVf, VSI and StO_2 parameters using two equations, described in Appendix II.

3.2.4 Performance evaluation

To compare the methods performance in parameter estimation, a set of M test signals is generated in the same way as for the dictionaries. The parameters values are randomly sampled in the parameter space and then the associated signals are computed. For each parameter, we compute the root mean square error (RMSE) as the square root of the quadratic mean of the differences between the estimated and the true parameter values.

4 Results

4.1 Synthetic scalable signals

4.1.1 Effect of sampling strategy on parameter accuracy

We investigate the impact of three parameter sampling strategies, regular, random and quasi-random, using synthetic scalable signals and the DBL method. We consider successively $P = 3$, 5, and 7, for each sampling strategy, leading to a total of 9 conditions. The numbers of entries in the dictionary are $N = 216$, 1 024, and 2 187, respectively. For each value of P, $M = 1\,000$ test signals are generated from parameters randomly sampled in the parameter space. The RMSE between the estimated and the true parameter values is then computed (see section 3.2.4) and divided by the number of parameters to obtain the average RMSE. To characterize the distribution of the average RMSE, the whole procedure was repeated 500 times (fig. 1).

Regardless of the sampling strategy, the average RMSE increases with P, the number of parameters. As reported previously, for the same number of signals in the dictionary, random sampling gives a lower average RMSE than regular sampling, whatever the number of parameters [16]. This observation is also valid for other conditions presented in Supporting Information, fig. S4. Note that when using the DBM approach instead of DBL, regular sampling yields a lower average RMSE than the random or quasi-random sampling (Supporting Information, fig. S3). The quasi-random sampling further reduces the average RMSE up to 12.4 %. Altogether, there is a reduction of 12.3 %,

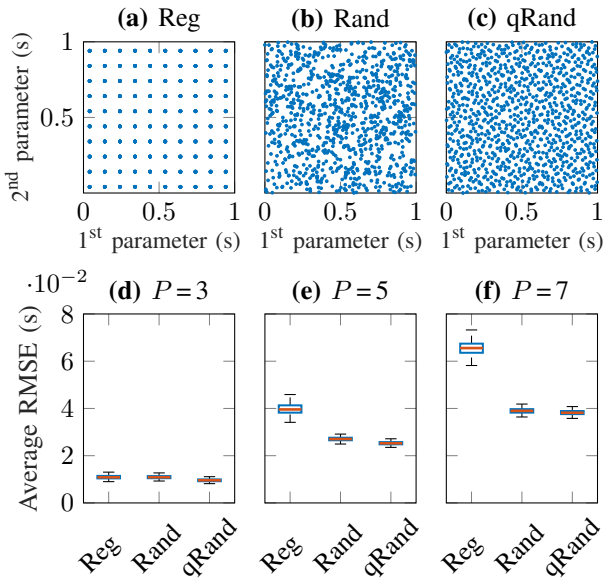

Figure 1: Effect of sampling strategy on the dictionary-based learning (DBL) method, using synthetic scalable signals: The first row shows the 2-dimensional projection of $N = 1\,000$ dictionary entries of the 3-dimensional parameter space ($P = 3$) obtained from **(a)** a regular grid sampling (Reg), **(b)** a random sampling (Rand) and **(c)** a quasi-random sampling (qRand) obtained from scrambled Sobol sequence. The second row shows the average RMSE ($M = 1\,000$ test signals) on the parameter estimates obtained using the DBL method for the three sampling strategies and **(d)** $P = 3$, **(e)** 5 and **(f)** 7 parameters. For each box, the red central mark indicates the median; the lower and upper edges indicate the 25th and 75th percentiles, respectively. The whiskers extend to the minimum and maximum values.

36.4 % and 41.7 % in average RMSE between regular and quasi-random sampling for 3, 5, and 7 parameters, respectively. Therefore, in the following, a regular sampling is used for the DBM method and a quasi-random sampling is used for the DBL method.

4.1.2 Impact of the dictionary size and SNR on parameter accuracy

To study this impact for DBM and DBL, we generate four scalable signals dictionaries for $P = 5$ and 7 parameters (a total of 8 conditions). The number of dictionary entries N was chosen so as to keep similar densities, *i.e.* a constant number of values per parameter (for $P = 5$: $N = 3^5$, 4^5, 5^5, and 6^5 and for $P = 7$: $N = 3^7$, 4^7, 5^7, and 6^7). For each condition, we evaluate the average RMSE, using $M = 10\,000$ signals. To characterize the impact of SNR, the procedure is repeated for test signals with SNR between 10 and 110 (fig. 2).

As expected for the DBM method, the average RMSE decreases as the number of entries N increases. The average RMSE decreases as the SNR increases to about SNR = 60 and then plateaus near the value obtained in absence of noise. For the DBL method, the average RMSE also decreases as the SNR increases but up to about SNR = 90. Again, the highest SNR yields an average RMSE close to that obtained in the absence of noise. For the DBM method, the average RMSE is comparable between 5 and 7 parameters. The average RMSE obtained with the DBL method are

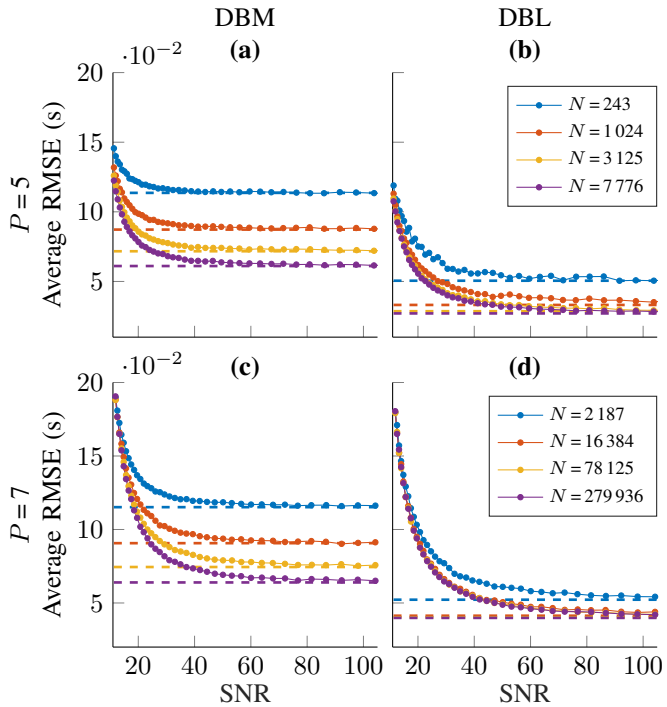

Figure 2: Impact of dictionary size and SNR on DBM and DBL methods, using synthetic scalable signals. Average RMSE are given with respect to the SNR for different numbers of parameters and dictionary entries. Average RMSE ($M = 10\,000$ test signals) for the DBM **(a, c)** and DBL **(b, d)** methods. The upper row **(a, b)** shows the results for $P = 5$ parameters, and the lower row **(c, d)** for $P = 7$ parameters. The dashed lines represent the average RMSE in the absence of noise on the test signals. The calibration value K (DBL method) is set to 50, except for $N = 243$ where $K = 20$.

lower than those obtained with the DBM method: $50.0 \pm 10.0\,\%$ lower for $P = 5$ and $38.0 \pm 12.3\,\%$ lower for $P = 7$, whatever the number of dictionary entries. N has a lower impact for DBL than for DBM. Between the smallest and the largest dictionary size, the average RMSE decreases by $38.0 \pm 9.5\,\%$ for the DBM method, while it decreases by only $18.3 \pm 5.2\,\%$ for the DBL method. Moreover, the two highest N yield similar average RMSE for the DBL method (difference smaller than $0.5\,\%$), suggesting that an increase in the number of entries would not further improve the average RMSE. Altogether, compared to the DBM method, the DBL method reduces the average RMSE by $13.1 \pm 5.2\,\%$ (respectively $12.3 \pm 4.3\,\%$) while reducing the number of entries by a factor of 32 (respectively 128) for 5 parameters (respectively 7 parameters). Supporting Information, fig. S5 shows the same experiment for additional P and N values . The average RMSE is always lower for DBL than for DBM ($39.2 \pm 15.9\,\%$ lower).

By eliminating the costly dictionary matching operation, DBL can greatly reduce computation time when N increases. For 7 parameters and $N = 78\,125$, inverting 10 000 test signals takes 2.0 ± 0.1 seconds with DBM and 2.2 ± 0.4 seconds with DBL. When N increases to 279 936, the estimation time increases to 28.6 ± 2.6 seconds for DBM while it remains stable at 2.0 ± 0.2 seconds for DBL. In terms of memory, these dictionaries require 66.9 ($N = 78\,125$) and 239.6 Mo ($N = 279\,936$) whereas they required only 4.35 Mo once summarized by a model. A complete comparison of the

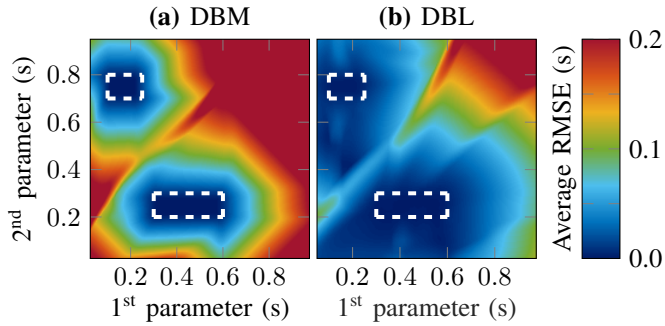

Figure 3: Estimation accuracy outside the limits of the parameter space covered by the dictionary, using synthetic scalable signals: Average RMSE ($M = 2\,000\,000$ test signals) in the parameter space ($P = 2$) obtained **(a)** with the DBM and **(b)** with the DBL method. The white dashed lines delimit the subspace covered by the dictionary. The average RMSE is computed from signals in a 50×50 ms sliding window, moving in 5 ms steps in the parameter space.

performance of the methods in terms of speed and memory is given in Supporting Information, fig. S6. We observe that the number of dictionary entries has no effect on the estimation time or on the memory size once the model is learnt.

4.1.3 Boundary behaviour

The DBL method estimates parameter values using a continuous function that is not limited to the parameter space covered by the dictionary entries. To investigate the behavior of DBM and DBL methods outside the limits of this parameter space, we define a dictionary ($N = 10\,000$) composed of two disjoint patches in the parameter space, generate $M = 2\,000\,000$ test signals and evaluate the average RMSE for each parameter value.

The two methods yield similar estimation accuracy in the subspace covered by the dictionary entries (fig. 3). Outside that subspace, the average RMSE obtained with the DBM method increases with the distance to the subspace. For the DBL method, the average RMSE remains below 100 ms, well beyond the limit of the dictionary subspace.

4.1.4 Confidence index

We investigate the relationship between the CI, available with the DBL method, and the RMSE. We generate $N = 10\,000$ dictionary entries and $M = 10\,000$ test signals. We then add different noise levels to the test signals to obtain a SNR = 20, 30, 40, 60 and 100. A single initial regression model is computed. For each SNR, this model is then updated based on the noise level (denoted by η) which corresponds to the SNR values of the test signals (see section 2.2). We compute the RMSE and CI for the initial model (*i.e.* without accounting for the noise level) and $\text{RMSE}_{\boldsymbol{\eta}}$ and $\text{CI}_{\boldsymbol{\eta}}$ using the updated model. For each SNR value, the experiment is repeated 100 times. Supporting Information, fig. S6, shows that the non-updated CI is proportional to but not equal to the RMSE in the SNR value range. Supporting Information, fig. S7 also shows that the scaling factor between RMSE and non-updated CI depends on the added noise level.

As expected, $\text{RMSE}_{\boldsymbol{\eta}}$ and $\text{CI}_{\boldsymbol{\eta}}$ increase as the SNR decreases (fig. 4a). $\text{RMSE}_{\boldsymbol{\eta}}$ and $\text{CI}_{\boldsymbol{\eta}}$ are

proportional and comparable in the SNR value range (slope: 0.99, $R^2 = 0.95$). Note that $CI_{\boldsymbol{\eta}}$ may slightly under or over-estimate the $RMSE_{\boldsymbol{\eta}}$ (mean difference: 7.8 %). Overall, $CI_{\boldsymbol{\eta}}$ appears to be a good indicator of the $RMSE_{\boldsymbol{\eta}}$. Interestingly, the inclusion of noise in the model slightly improves the estimation accuracy. On average, the $RMSE_{\boldsymbol{\eta}}$ is 4.11 % lower than the RMSE (fig. 4b). In the following, for DBL, RMSE and CI refer to $RMSE_{\boldsymbol{\eta}}$ and $CI_{\boldsymbol{\eta}}$ (updated model).

4.2 Vascular MRF signals

4.2.1 Synthetic vascular MRF signals

We compare the two dictionary-based methods and the CEF method on synthetic vascular MRF signals. The dictionaries (grid and quasi-random sampling) are simulated with a BVf between 0.25 and 30 %, a VSI between 0.5 and 50 μm and a StO_2 between 30 and 95 %. Among the 170 100 combinations, some signals cannot be produced, due to simulation constraints (*e.g.* a very large BVf cannot be produced with distant, small, vessels or small BVf with large vessels). The obtained N values reduce then to $N = 164\,524$ for the grid and to $N = 167\,216$ for quasi-random sampling.

For each method, $M = 100\,000$ test signals (SNR = 100) are generated. To analyze the BVf RMSE, test signals are divided into three parts: small, medium and large vessel sizes. To analyse the VSI RMSE, test signals are divided into three parts: low, medium and large blood volumes.

For all vessel diameters, the RMSE on BVf tends to increase with BVf (fig. 5). The DBM and CEF methods yield similar RMSE for BVf values below 10 %. For medium and large vessels and large BVf values, the CEF method yields the highest errors. The DBL method always yields the lowest error with an RMSE of 2.38 % for CEF, 2.68 % for DBM and 1.30 % for DBL.

For VSI values smaller than 15 μm, the behavior of the RMSE is similar in all three methods for the three BVf ranges. Above 15 μm, the CEF method yields larger errors than the two dictionary-based approaches and the RMSE obtained with CEF is linearly correlated with the VSI value ($R^2 \geq 0.99$). This linear behaviour has already been reported in [32]. DBL yields a 25 % smaller

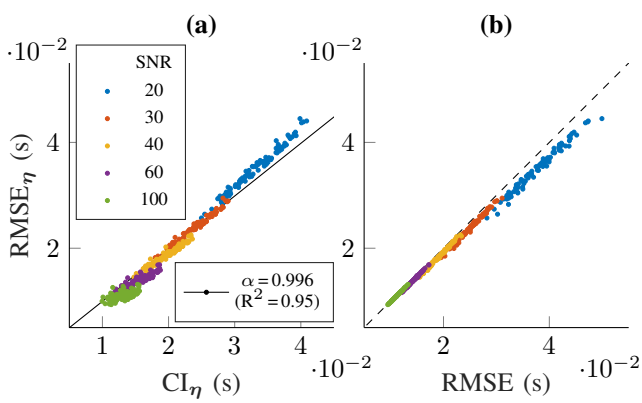

Figure 4: $RMSE_{\boldsymbol{\eta}}$ versus confidence index ($CI_{\boldsymbol{\eta}}$) and RMSE (non-updated model), using synthetic scalable signals. $SNR_{test} = 20$ (blue), 30 (orange), 40 (yellow), 60 (purple) and 100 (green). $M = 10\,000$ test signals. **(a)** The black line represents the proportional regression coefficient α between $RMSE_{\boldsymbol{\eta}}$ and $CI_{\boldsymbol{\eta}}$ for all SNR values. R^2 is the coefficient of determination. **(b)** The dashed black line is the identity function.

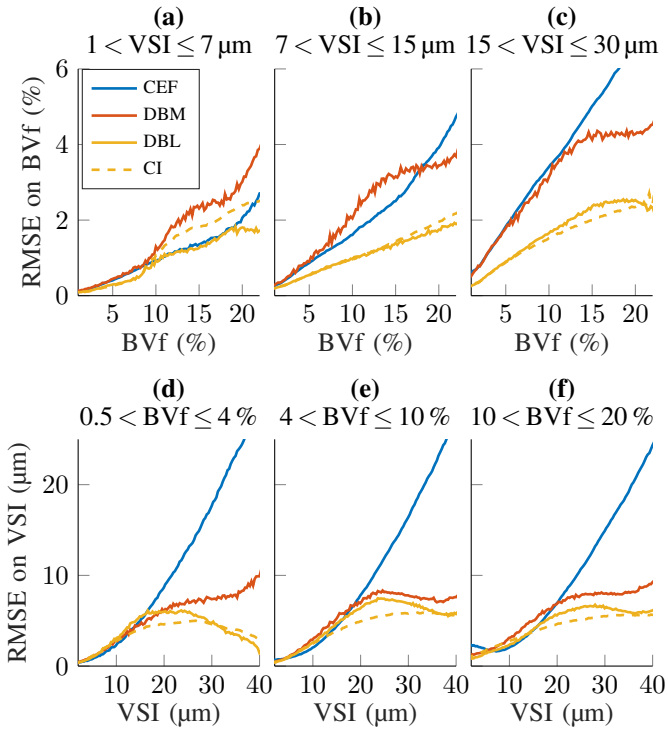

Figure 5: Comparison of the RMSE on BVf and VSI obtained with the two dictionary-based methods (DBM and DBL) and with the closed-form expression fitting (CEF) method, using synthetic vascular MRF signals. Figures **(a, b, c)** show RMSE ($M = 100\,000$ test signals) on BVf for three ranges of VSI, and figures **(d, e, f)** show the RMSE on VSI for three ranges of BVf. The dashed lines represent the average confidence indices (CI) on BVf (first row) and VSI (second row) obtained with DBL. The data are shown after 1-dimensional sliding window filtering (3 % for BVf and 5 μm for VSI). The dictionary dimensions are $P = 3$, $S = 32$, and $N = 164\,524$ for DBM and $N = 167\,216$ for DBL.

RMSE than DBM, on average with an RMSE of 12.46 μm for CEF, 6.11 μm for DBM, and 4.50 μm for DBL. The CI appears again to be a good indicator of the RMSE, with maximum differences between CI and RMSE of 1.07 % for BVf and 2.36 μm for VSI and average differences of 0.25 % and 0.75 μm.

We also compared DBM and DBL on the standard MRF signals proposed by Ma *et al.* [1] (Supporting Information, Synthetic standard MRF signals). Except for the off-resonance parameter with SNR < 20 and with the largest dictionary, DBL always yields smaller average RMSE than DBM (Supporting Information, fig. S1).

4.2.2 Acquired vascular MRF signals

The DBL method is then applied to acquired vascular MRF signals collected from rats bearing 9L and C6 tumors. We quantify BVf, VSI and StO_2 with both dictionary-based methods (DBM and DBL) and using two numbers of dictionary entries. The large dictionary ($N = 170\,100$) is the one used previously in section 4.2.1. The small dictionary ($N = 4\,320$) is simulated with BVf between

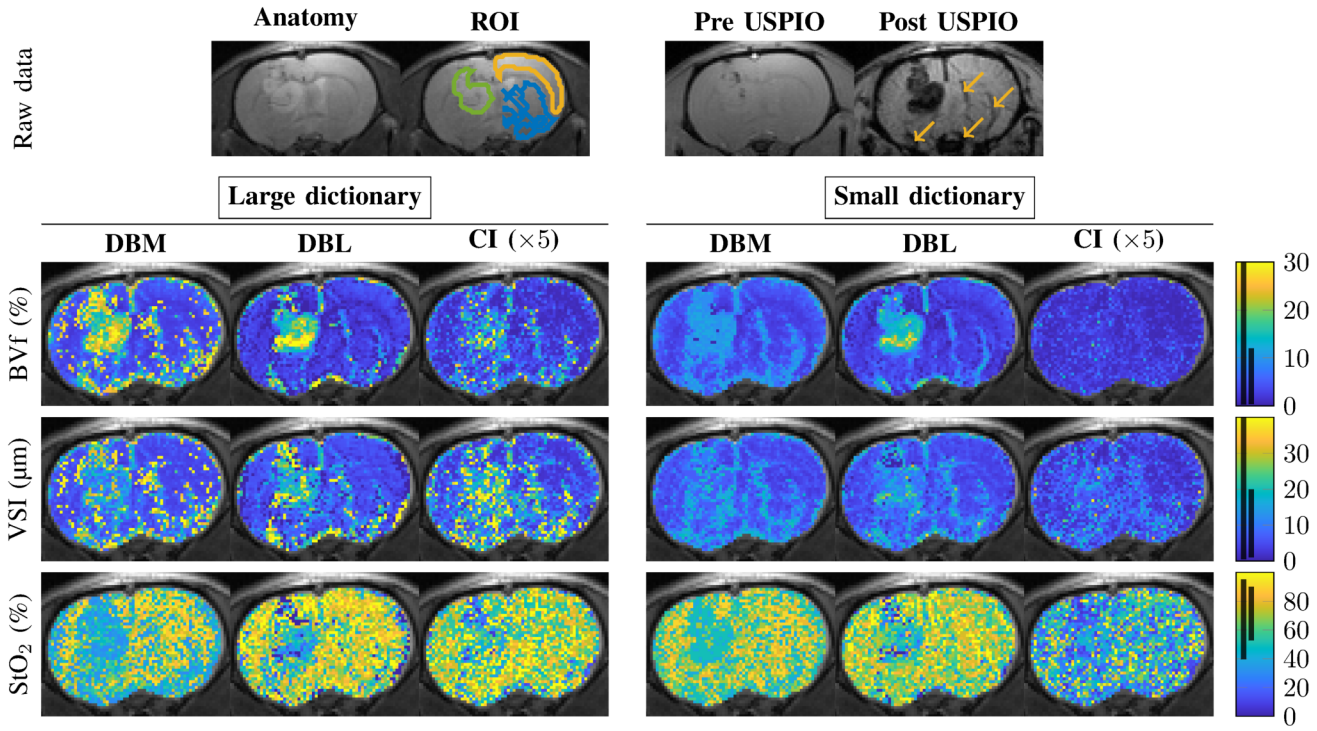

Figure 6: Maps of vascular parameter estimates of a 9L rat tumor model. The first row shows the anatomical image and regions of interest (left) and the MGEFIDSE pre and post USPIO injection (right) for the second echo time (6.3 ms). The tumor, cortex and striatum are respectively delineated with green, yellow and blue lines. The arrows on the post-USPIO injection image indicate large vessels. The estimated maps for BVf, VSI and StO_2 are shown below, using DBM (first and fourth columns) and DBL (second and fifth columns). The third and sixth columns show the DBL confidence index (CI) maps. In the color bars, the black lines represent the parameter ranges covered by the two dictionaries: the short (resp. long) line for the small (resp. large) dictionary. Large dictionary: $N = 164\,524$ for DBM; 70 values for BVf between 0.25 and 30 %, 90 values for VSI between 0.5 and 50 μm and 27 values for StO_2 between 30 and 95 % and $N = 167\,216$ for DBL. Small dictionary: $N = 4\,218$ for DBM; 36 values for BVf between 0.33 and 12 %, 20 values for VSI between 1 and 20 μm and 6 values for StO_2 between 40 and 90 % and $N = 4\,119$ for DBL.

0.33 and 12 %, VSI between 1 and 20 μm and StO_2 between 40 and 90 %.

All methods yield consistent estimates (fig. 6), in which the tumor and the large vessels can be easily depicted on BVf and VSI maps. StO_2 appears constant in healthy tissues. However, for DBM, there are many isolated high values, suggesting noisy maps. In [30], the authors proposed to remove the last 8 signal samples and apply a spatial Gaussian filtering to increase the SNR. Using the DBM method, these additional steps allowed them to produce less noisy maps. Interestingly, with the DBL method, these steps are no longer required. Our approach is therefore more likely to preserve small structure information, which may be otherwise removed by an additional spatial filtering.

For the large dictionary, the mean BVf and VSI obtained in tumor with the DBM and DBL methods are similar but with the DBL method, we can differentiate different subregions within

the lesion. Mean values and standard deviations are $20.20 \pm 6.14\,\%$, $15.90 \pm 7.79\,\mu$m for DBM and $18.22 \pm 9.34\,\%$, $19.91 \pm 11.76\,\mu$m for DBL. For StO_2, the DBL method provides significantly larger values, closer to the expected values for healthy tissue [33]. Values for striatum are $62.67 \pm 21.73\,\%$ for DBM and $74.46 \pm 18.96\,\%$ for DBL. Overall, DBM and DBL yield comparable values.

For the small dictionary, the contrasts are similar. The estimates obtained by the DBM method are limited to the space spanned by the dictionary, while the DBL method yields estimates outside this range and closer to the parameter values obtained with the large dictionary. In the tumor, the mean BVf is $13.66\,\%$ while the maximum dictionary value is $12\,\%$. Supporting Information, fig. S8, shows an evaluation by region of interest (ROI) (8 animals: 4 from each tumor model). The DBL method with a small dictionary produces estimates similar to those obtained with a large dictionary, except for the largest values which are underestimated. This results in a slight reduction in mean ROI values. The average differences between the mean ROI values obtained by the DBL method with the large and small dictionaries are $1.17\,\%$ for BVf, $3.50\,\mu$m for VSI and $1.62\,\%$ for StO_2.

The mean CIs in the tumor are $2.36\,\%$ for BVf, $4.90\,\mu$m for VSI and $12.22\,\%$ for StO_2, while in the cortex the mean CI are $0.81\,\%$ for BVf, $1.96\,\mu$m for VSI and $14.81\,\%$ for StO_2. These results are in agreement with previous results that pointed better estimates for BVf than for VSI and low signal sensitivity to StO_2 [30]. Confidence in the estimates is therefore on average 5 to 6 times lower in the tumor than in the cortex for BVf and VSI but is similar for StO_2. This is an interesting contribution of our DBL method as to our knowledge, it is the first time that such error maps are provided for BVf, VSI, and StO_2.

A second example of a C6 rat tumor model is given in Supporting Information, fig. S9. The results are similar to those observed for the 9L tumor. DBL and DBM produce comparable physiological parameter values. With DBL, however, parameter maps are less noisy and additional CI maps can be produced.

5 Discussion

This work presents a method for estimating vascular MRF parameters based on dictionary learning. It preserves the main advantages of the MRF method, *i.e.* robustness, speed and flexibility, and meets the challenge of producing accurate estimates from a small dictionary, even when the dimension of the parameters is large.

Regarding the design of the dictionary, we observe as [16] that random sampling of the parameter space gives more accurate estimates than grid sampling, when an inverse regression model is used. We further show that quasi-random sampling gives even better estimates. However, when the DBM method is used, the grid remains the most efficient sampling strategy. The appropriate dictionary design depends thus on the chosen inversion approach (grid matching vs statistical learning). The results obtained with the DBL method suggest that the simulation of a few patches (with quasi-random sampling) in the parameter space could save even more time in the construction of the dictionary (section 4.1.3). This result seems specific to the regression method. Indeed, a study using neural networks [14] rather reported increased deviations from the true values at the boundaries of the training dictionary. This is likely due to the vanishing gradient of the activation function in these regions. This patch approach could be implemented by combining a set of parameter values corresponding to healthy tissues and one or more sets corresponding to damaged tissues (*e.g.*, tumorous tissues). If needed, a few additional dictionary entries can be generated to fill gaps

between patches (Supporting Information, fig. S10).

Regarding the quality of the inversion, the regression approach usually involves calibration. In GLLiM, the number of Gaussian distributions K is the only calibration value that needs to be adjusted. This can be done automatically using a standard information criterion such as AIC or BIC [34,35], as illustrated in [20], but this may result in additional learning time. We have observed that our results have little dependence on K. In contrast, neural networks [11–18] used to solve similar inverse problems are very sensitive to their many complex calibration settings: architecture, batch sizes, learning rates, among others. The adjustment of all these calibration settings is usually performed by learning the model for a large number of calibration value combinations, which represents a higher computational cost than determining a single calibration value as in GLLIM. This difference in calibration cost makes the DBL method more flexible in case of change in dictionary design (extension of the parameter range, additional parameters, *etc.*). In addition, GLLiM has the advantage of providing a richer information. Here, we make use of the full posterior parameter distribution provided by GLLiM to derive a CI for each estimate. We observe that this CI matches the RMSE in cases where a ground truth is available (synthetic scalable; section 4.1.4 and synthetic vascular MRF signals; section 4.2.1). Interestingly, this CI reports both signal and model errors (derived from the dictionary) and thus reflects the whole DBL procedure.

Regarding the acquired vascular MRF signals, we observe that the number of dictionary entries can be divided by ≈ 40 using the DBL method and still lead to accurate maps. The maps produced with the DBL method are significantly less noisy than those obtained with the DBM method and some structures, not observed with DBM, appear in the lesions. The additional tissue contrast provided by DBL could therefore contribute to improved tumor characterization [36].

In conclusion, this first evaluation of the DBL method appears promising. It reduces the simulation time and the memory required for dictionary storage, improves parameter accuracy, reduces the estimation time, and provides a first confidence index on parameter estimates. The DBL method could become even more efficient as the number of parameters to be estimated increases, which can happen when considering all the possible contributions of the tissues and the scanner to the signal. In addition, the flexibility of the proposed approach opens the door to further improvements. In particular, future work should include the adaptation and optimization of the dictionary sampling strategy with respect to the targeted range of parameters to estimate. In addition, MRI produces complex-valued data but machine learning methodologies are generally designed for real-valued data. Dealing with complex-valued data could boost the performance of DBL [11]. This requires an adjustment of the inverse regression method also left to future work.

Appendix I - Mri experiments

The study design was approved by the local institutional animal care and use committee. All animal procedures complied with French government guidelines and were performed under permit 380820 and A3851610008 (for experimental and animal care facilities) from the French Ministry of Agriculture (Articles R214–117 to R214–127 published on 7 February 2013). This study complies with the ARRIVE guidelines (Animal Research: Reporting in Vivo Experiments) [37]. Animals 7 weeks old at the start of the experiments (Charles River, France) were housed in groups of 3-4 in Plexiglas cages under standard laboratory condition (12 h light/dark cycle with lights off at 7:00 p.m. and controlled temperature in $22 \pm 2\,^{\circ}$C). Water and standard laboratory chow were provided *ad libitum*. All procedures were performed under anesthesia by isoflurane (IsoFlo, Abbot

Laboratories Ltd, Berkshire, UK). 9LGS cells were implanted in the brain of male Fisher rats. One μl of cell suspension in serum-free RPMI1640 medium containing 10^4 cells were inoculated. MRI was performed 10 days after tumor implantation. C6 cells were implanted in the brain of male Wistar rats. Five μl of cell suspension in serum-free RPMI1640 medium containing 10^5 cells were inoculated. MRI was performed 20 days after tumor implantation. Animals were euthanized by intra-cardiac injection of Pentobarbital 200 mg.kg^{-1} (Dolhethal, Vétoquinol Inc, France).

Appendix II - Closed-form expression fitting (cef) method

Parameters BVf and VSI are estimated from the gradient echo sampling of the free induction decay and spin echo according to [31]. The changes in relaxation rates ΔR_2^* and ΔR_2 induced by the injection of the ultrasmall superparamagnetic iron oxide particles (USPIO) contrast agent are computed using gradient echo (GE) signal intensities and spin echo (SE) signal intensities, respectively. The pre-injection and post-injection relaxation times are obtained by fitting the GE signal intensities to an exponential function. It allows to compute ΔR_2^*. ΔR_2 is directly calculated from the two SE signal intensities. BVf and VSI are computed using:

$$\mathrm{BVf} = \frac{3}{4\pi\ \gamma\ B_0\ \Delta\chi_{\mathrm{USPIO}}}\ \Delta R_2^*\ ,$$

$$\mathrm{VSI} = 0.425\ \left(\frac{\mathrm{ADC}}{\gamma\ B_0\ \Delta\chi_{\mathrm{USPIO}}}\right)^{\frac{1}{2}}\ \left(\frac{\Delta R_2^*}{\Delta R_2}\right)^{\frac{3}{2}}\ ,$$

where $\gamma = 2.6752 \times\ 10^8\,\mathrm{rad.s^{-1}.T^{-1}}$ is the gyromagnetic ratio, $B_0 = 4.7\,\mathrm{T}$ is the magnetic field, $\Delta\chi_{\mathrm{USPIO}} = 3.5\,\mathrm{ppm}$ (SI unit) is the susceptibility difference between blood in the presence and in the absence of USPIO and $\mathrm{ADC} = 800\,\mu\mathrm{m^2.s^{-1}}$.

References

[1] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold, "Magnetic resonance fingerprinting," *Nature*, vol. 495, no. 7440, p. 187, 2013.

[2] B. Bipin Mehta, S. Coppo, D. Frances McGivney, J. Ian Hamilton, Y. Chen, Y. Jiang, D. Ma, N. Seiberlich, V. Gulani, and M. Alan Griswold, "Magnetic resonance fingerprinting: a technical review," *Magnetic resonance in medicine*, vol. 81, no. 1, pp. 25–46, 2019.

[3] T. Christen, N. Pannetier, W. W. Ni, D. Qiu, M. E. Moseley, N. Schuff, and G. Zaharchuk, "MR vascular fingerprinting: a new approach to compute cerebral blood volume, mean vessel radius, and oxygenation maps in the human brain," *Neuroimage*, vol. 89, pp. 262–270, 2014.

[4] D. F. McGivney, E. Pierre, D. Ma, Y. Jiang, H. Saybasili, V. Gulani, and M. A. Griswold, "SVD compression for magnetic resonance fingerprinting in the time domain," *IEEE transactions on medical imaging*, vol. 33, no. 12, pp. 2311–2322, 2014.

[5] S. F. Cauley, K. Setsompop, D. Ma, Y. Jiang, H. Ye, E. Adalsteinsson, M. A. Griswold, and L. L. Wald, "Fast group matching for MR fingerprinting reconstruction," *Magnetic resonance in medicine*, vol. 74, no. 2, pp. 523–528, 2015.

[6] B. Zhao, K. Setsompop, E. Adalsteinsson, B. Gagoski, H. Ye, D. Ma, Y. Jiang, P. Ellen Grant, M. A. Griswold, and L. L. Wald, "Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling," *Magnetic resonance in medicine*, vol. 79, no. 2, pp. 933–942, 2018.

[7] M. Yang, D. Ma, Y. Jiang, J. Hamilton, N. Seiberlich, M. A. Griswold, and D. McGivney, "Low rank approximation methods for MR fingerprinting with large scale dictionaries," *Magnetic resonance in medicine*, vol. 79, no. 4, pp. 2392–2400, 2018.

[8] J. Assländer, M. A. Cloos, F. Knoll, D. K. Sodickson, J. Hennig, and R. Lattanzi, "Low rank alternating direction method of multipliers reconstruction for mr fingerprinting," *Magnetic resonance in medicine*, vol. 79, no. 1, pp. 83–96, 2018.

[9] G. Nataraj, J.-F. Nielsen, C. Scott, and J. A. Fessler, "Dictionary-free MRI perk: Parameter estimation via regression with kernels," *IEEE transactions on medical imaging*, vol. 37, no. 9, pp. 2103–2114, 2018.

[10] B. Zhao, K. Setsompop, H. Ye, S. F. Cauley, and L. L. Wald, "Maximum likelihood reconstruction for magnetic resonance fingerprinting," *IEEE transactions on medical imaging*, vol. 35, no. 8, pp. 1812–1823, 2016.

[11] P. Virtue, X. Y. Stella, and M. Lustig, "Better than real: Complex-valued neural nets for MRI fingerprinting," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3953–3957.

[12] E. Hoppe, G. Körzdörfer, T. Würfl, J. Wetzl, F. Lugauer, J. Pfeuffer, and A. Maier, "Deep learning for magnetic resonance fingerprinting: A new approach for predicting quantitative parameter values from time series," *Stud Health Technol Inform*, vol. 243, pp. 202–206, 2017.

[13] E. Hoppe, F. Thamm, G. Körzdörfer, C. Syben, F. Schirrmacher, M. Nittka, J. Pfeuffer, H. Meyer, and A. Maier, "Rinq fingerprinting: Recurrence-informed quantile networks for magnetic resonance fingerprinting," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 92–100.

[14] O. Cohen, B. Zhu, and M. S. Rosen, "MR fingerprinting deep reconstruction network (drone)," *Magnetic resonance in medicine*, vol. 80, no. 3, pp. 885–894, 2018.

[15] F. Balsiger, A. S. Konar, S. Chikop, V. Chandran, O. Scheidegger, S. Geethanath, and M. Reyes, "Magnetic resonance fingerprinting reconstruction via spatiotemporal convolutional neural networks," in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2018, pp. 39–46.

[16] M. Barbieri, L. Brizi, E. Giampieri, F. Solera, G. Castellani, C. Testa, and D. Remondini, "Circumventing the curse of dimensionality in magnetic resonance fingerprinting through a deep learning approach," *arXiv preprint arXiv:1811.11477*, 2018.

[17] P. Song, Y. C. Eldar, G. Mazor, and M. R. Rodrigues, "Hydra: Hybrid deep magnetic resonance fingerprinting," *Medical physics*, vol. 46, no. 11, pp. 4951–4969, 2019.

[18] M. Golbabaee, D. Chen, P. A. Gómez, M. I. Menzel, and M. E. Davies, "Geometry of deep learning for magnetic resonance fingerprinting," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 7825–7829.

[19] F. Boux, F. Forbes, J. Arbel, and E. Barbier, "Dictionary-free MR fingerprinting parameter estimation via inverse regression," in *26th Annual Meeting ISMRM, Paris*, 2018, p. 4259.

[20] A. Deleforge, F. Forbes, and R. Horaud, "High-dimensional regression with Gaussian mixtures and partially-latent response variables," *Statistics and Computing*, vol. 25, no. 5, pp. 893–911, 2015.

[21] K.-C. Li, "Sliced Inverse Regression for dimension reduction," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 316–327, 1991.

[22] R. D. Cook and L. Forzani, "Partial least squares prediction in high-dimensional regression," *Ann. Statist.*, vol. 47, no. 2, pp. 884–908, 04 2019.

[23] H. D. Nguyen, F. Chamroukhi, and F. Forbes, "Approximation results regarding the multiple-output Gaussian gated mixture of linear experts model," *Neurocomputing*, 2019.

[24] S. Ingrassia, S. C. Minotti, and G. Vittadini, "Local statistical modeling via a cluster-weighted approach with elliptical distributions," *Journal of Classification*, vol. 29, pp. 363–401, 2012.

[25] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*. Siam, 1992, vol. 63.

[26] A. B. Owen, "Randomly permuted (t, m, s)-nets and (t, s)-sequences," in *Monte Carlo and quasi-Monte Carlo methods in scientific computing*. Springer, 1995, pp. 299–317.

[27] P. Bratley and B. L. Fox, "Algorithm 659: Implementing sobol's quasirandom sequence generator," *ACM Transactions on Mathematical Software (TOMS)*, vol. 14, no. 1, pp. 88–100, 1988.

[28] J. Matoušek, "On the l_2-discrepancy for anchored boxes," *Journal of Complexity*, vol. 14, no. 4, pp. 527–556, 1998.

[29] N. A. Pannetier, C. S. Debacker, F. Mauconduit, T. Christen, and E. L. Barbier, "A simulation tool for dynamic contrast enhanced MRI," *PloS one*, vol. 8, no. 3, p. e57636, 2013.

[30] B. Lemasson, N. Pannetier, N. Coquery, L. S. Boisserand, N. Collomb, N. Schuff, M. Moseley, G. Zaharchuk, E. Barbier, and T. Christen, "MR vascular fingerprinting in stroke and brain tumors models," *Scientific reports*, vol. 6, p. 37071, 2016.

[31] I. Tropres, S. Grimault, A. Vaeth, E. Grillon, C. Julien, J.-F. Payen, L. Lamalle, and M. Décorps, "Vessel size imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 45, no. 3, pp. 397–408, 2001.

[32] I. Troprès, N. Pannetier, S. Grand, B. Lemasson, A. Moisan, M. Péoc'h, C. Rémy, and E. L. Barbier, "Imaging the microvessel caliber and density: principles and applications of microvascular MRI," *Magnetic resonance in medicine*, vol. 73, no. 1, pp. 325–341, 2015.

[33] B. Lemasson, T. Christen, R. Serduc, C. Maisin, A. Bouchet, G. Le Duc, C. Rémy, and E. L. Barbier, "Evaluation of the relationship between MR estimates of blood oxygen saturation and hypoxia: effect of an antiangiogenic treatment on a gliosarcoma model," *Radiology*, vol. 265, no. 3, pp. 743–752, 2012.

[34] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[35] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[36] A. Arnaud, F. Forbes, N. Coquery, N. Collomb, B. Lemasson, and E. L. Barbier, "Fully automatic lesion localization and characterization: Application to brain tumors using multiparametric quantitative MRI data," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1678–1689, 2018.

[37] C. Kilkenny, W. Browne, I. C. Cuthill, M. Emerson, and D. G. Altman, "Animal research: reporting in vivo experiments: the arrive guidelines," *British journal of pharmacology*, vol. 160, no. 7, pp. 1577–1579, 2010.