

# Estimating Kinetic Parameter Maps from Dynamic Contrast-Enhanced MRI using Spatial Prior Knowledge

B. Michael Kelm, *Member, IEEE*, Bjoern H. Menze, Oliver Nix, Christian M. Zechmann, and Fred A. Hamprecht

**Abstract**—Dynamic contrast-enhanced magnetic resonance (DCE-MR) imaging can be used to study microvascular structure *in vivo* by monitoring the abundance of an injected diffusible contrast agent over time. The resulting spatially resolved intensity-time curves are usually interpreted in terms of kinetic parameters obtained by fitting a pharmacokinetic model to the observed data. Least squares estimates of the highly nonlinear model parameters, however, can exhibit high variance and can be severely biased. As a remedy, we bring to bear spatial prior knowledge by means of a generalized Gaussian Markov random field (GGMRF). By using information from neighboring voxels and computing the maximum a posteriori solution for entire parameter maps at once, both bias and variance of the parameter estimates can be reduced thus leading to smaller root mean square error (RMSE). Since the number of variables gets very big for common image resolutions, sparse solvers have to be employed. To this end, we propose a generalized iterated conditional modes (ICM) algorithm operating on blocks instead of sites which is shown to converge considerably faster than the conventional ICM algorithm. Results on simulated DCE-MR images show a clear reduction of RMSE and variance as well as, in some cases, reduced estimation bias. The mean residual bias (MRB) is reduced on the simulated data as well as for all 37 patients of a prostate DCE-MRI dataset. Using the proposed algorithm, average computation times only increase by a factor of 1.18 (871 ms per voxel) for a Gaussian prior and 1.51 (1.12 s per voxel) for an edge-preserving prior compared to the single voxel approach (740 ms per voxel).

**Index Terms**—dynamic contrast-enhanced imaging, Markov random field, nonlinear least squares, block iterated conditional modes, kinetic parameter maps

## I. INTRODUCTION

**D**YNAMIC contrast-enhanced MR imaging is used to track the diffusion of a paramagnetic contrast medium such as gadopentate dimeglumine (Gd-DTPA) and to study tissue perfusion and vascular permeability *in vivo* [1]–[3]. By recording a sequence of T1-weighted MR images at intervals

of a few seconds the uptake and washout of the administered contrast medium (CM) can be observed in the imaged tissue, leading to characteristic intensity-time curves. Since perfusion and permeability are usually changed in tumors, dynamic contrast-enhanced MR imaging (DCE-MRI) can be a valuable tool for clinical diagnostics [1], [4]–[8].

A variety of postprocessing strategies for DCE-MRI have been proposed, ranging from simple descriptive statistics [9] over unsupervised and supervised learning approaches [10], [11] to methods using physiologic and pharmacokinetic models which describe the expected signal enhancement dynamics [12]–[17]. The latter require the estimation of model parameters which is usually done using a nonlinear least squares (NLS) approach. Because of signal noise and the small number of sampling points, these parameter estimates may exhibit large variance as well as considerable bias [18]–[21]. Furthermore, since the NLS objective is not convex and can have multiple local optima, an iterative NLS solver may converge to erroneous solutions or fail to converge altogether [18]–[21].

In [21], Ahearn *et al.* show by means of a systematic Monte-Carlo study using an instance of the generalized kinetic model (GKM) [14] that certain areas in parameter space tend to generate poor fits and propose to use a multiple-starting-points strategy. Schmid *et al.* [20] propose to use Bayesian prior knowledge and obtain parameter estimates by means of sampling methods. They show that Bayesian prior knowledge can help reduce problems due to local optima and reduce the variance of the parameter estimates. Improvements are also achieved by appropriate parameter constraints, which can be regarded as Bayesian priors [18] and help to exclude areas of the parameter space that might contain local optima. Orton *et al.* [18] also observe severe estimation bias which they attribute to the existence of multiple local and biased global optima in the common approach. As a remedy, they propose to marginalize the onset time (lag time  $t_0$ ) which they identify as a critical parameter in their model.

In all approaches mentioned so far, intensity-time curves have been processed voxel by voxel. Thus, arbitrarily exchanging voxels would not change the evaluation results. Only recently, the application of *spatial prior knowledge* in DCE-MRI has been proposed [19], [22]. In addition to a pharmacokinetic model it is assumed that the characteristics of the tissue within homogeneous regions only vary gradually from voxel to voxel and, hence, that the parameter maps that best describe the physiologic properties of the tissue should

Manuscript received October 1, 2006; revised December 31, 2008; accepted March 16, 2009. This work was supported by the Deutsche Forschungsgemeinschaft under Grant DFG-HA-4364.

B. M. Kelm and B. H. Menze were with the Interdisciplinary Center for Scientific Computing, University of Heidelberg, Germany while conducting this work. B. M. Kelm is now with Siemens Corporate Technology, Erlangen, Germany. B. H. Menze is now with the Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. O. Nix and C. M. Zechmann are with the German Cancer Research Center, Heidelberg, Germany. F. A. Hamprecht is with the Heidelberg Collaboratory for Image Processing, Heidelberg, Germany (email: fred.hamprecht@iwr.uni-heidelberg.de).

Copyright (c) 2009 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

exhibit some spatial smoothness.

In the present paper the approach from [22] is refined and studied in more detail. Kinetic parameter maps are modeled as a generalized Gaussian random field (GGMRF) and the recorded DCE-MRI data are regarded as noisy observations of a nonlinear transformation of the hidden parameter maps. Hence, the parameter estimates at each voxel are supported by the DCE-MR data of a local neighborhood which reduces estimation error and which helps to alleviate problems due to local optima of the NLS objective. Furthermore, we present a block variant of the iterated conditional modes algorithm (Block-ICM) that can be used to tackle the huge but sparse optimization problem and which converges considerably faster than the conventional ICM algorithm.

Our approach is different from previous approaches such as [18], [20], [21] which perform estimation of pharmacokinetic parameters for each voxel individually. Such single-voxel approaches can only employ *single-voxel prior knowledge*, *i.e.* prior knowledge that concerns parameters at a single voxel only. To the best of our knowledge, apart from our previous work [22], only Schmid et al. [19] have examined using spatial prior knowledge for the processing of DCE-MRI before. However, Schmid et al. have chosen a fully Bayesian approach which requires the specification of prior distributions for *all* unknown parameters as well as additional hyper-parameters. Also, [19] relies on Markov chain Monte Carlo (MCMC) simulations for parameter estimation. As compared to that, our approach makes fewer assumptions on the underlying distributions and, like the conventional NLS approach, uses an iterative optimization method. In the present work, we also provide a detailed bias-variance analysis which was missing in both, [22] and [19].

The paper is organized as follows. In section II we briefly describe the employed pharmacokinetic model. Section III reviews the conventional approach to parameter estimation and associated problems and introduces the GGMRF prior followed by a description of the Block-ICM algorithm. Data simulation and acquisition parameters as well as algorithmic hyperparameters for the conducted experimental evaluation are described in section IV. We then present results for simulated DCE-MRI data as well as patient DCE-MRI data from a prostate study in section V. After a general discussion in section VI we conclude with a summary in section VII.

## II. DYNAMIC ENHANCEMENT MODEL

Various dynamic enhancement (DE) models that attempt to capture the pharmacokinetic behavior of the imaged tissue in DCE-MRI have been proposed, *e.g.* [12], [13], [15], [16]. The Generalized Kinetic Model (GKM) described by Tofts *et al.* [14] is a standard two-compartment model that unifies many of the previously defined pharmacokinetic models. By using a particular arterial input function (AIF), *e.g.* biexponential [17], [23], sum of Gaussians [24] or model-free [12], various model functions for the intensity-time curves can be derived from the GKM.

One major objective in pharmacokinetic modeling is the absolute quantification of physiological parameters in order to

make results comparable across patients and institutions. With recent models such as [12], [13], [24] this goal is achieved increasingly well, however, at the cost of increased complexity. More complex models are also more demanding and require, for example, higher temporal resolution (*e.g.* [12], [24]) or the estimation of more parameters (*e.g.* [13]). Furthermore, the accuracy of any model is ultimately limited by the complexity and variation of the underlying pathophysiology [1], [14]. For clinical purposes, the choice of the “right” model is still a topic of debate. Simple descriptive statistics, for example, have been found to yield competitive performance in the characterization of prostate cancer [5], [9]. The present work does not attempt to address the question of which model best describes the physiological processes. Instead we propose an approach that helps reduce problems with parameter estimation using any such DE model by exploiting spatial information.

For the present work the two-compartment model by Brix *et al.* [16] has been used, an instance of the GKM [14]. Unlike the improved model presented in [12], this model allows for a lower temporal resolution ( $\Delta T > 10$  s) and hence it allows to acquire DCE-MRI at a higher spatial resolution. This is particularly important for prostate screening where the whole organ is to be mapped [1], [4], [5]. Note that, although the proposed approach could be applied to any other parametric DE model as well, its benefits would have to be assessed anew. For convenience, we briefly review the model derived in [16] using the standardized notation of [14]. According to the GKM, the Gd-DTPA concentration in the tissue  $C_t(t)$  responds as a first-order dynamic system to concentration changes in the arterial blood plasma  $C_p(t)$  (the arterial input function). Hence,

$$\frac{dC_t}{dt}(t) + k_{ep}C_t(t) = K^{\text{trans}} C_p(t), \quad (1)$$

where  $K^{\text{trans}}$  is the volume transfer constant between plasma and extravascular extracellular space (EES),  $k_{ep} = K^{\text{trans}}/v_e$  is the rate constant between EES and plasma and the fractional EES volume  $v_e = V_e/V_t$  is defined as the ratio between total EES volume  $V_e$  and tissue volume  $V_t$  [3], [14].

For the model by Brix *et al.* [16] a particular AIF  $C_p(t)$  is used. It is assumed that the CM is administered at a constant rate  $K_{\text{in}}$  over a time-span  $\tau$  and is eliminated from the plasma compartment with first-order dynamics (elimination rate  $k_{\text{el}}$ ):

$$\frac{dC_p}{dt}(t) + k_{\text{el}}C_p(t) = \frac{K_{\text{in}}}{V_p} (h(t) - h(t - \tau)) \quad (2)$$

where  $V_p$  is the volume of the plasma compartment and  $h(t)$  the Heaviside step function. Using the AIF in Eqn. (2), an explicit expression for the tissue Gd-DTPA concentration can be derived (cf. appendix). In particular,

$$C_t(t) = K^{\text{trans}} \frac{K_{\text{in}}}{V_p} \cdot \frac{g_{k_{\text{el}}}(t) - g_{k_{ep}}(t)}{k_{ep} - k_{\text{el}}} \quad (3)$$

with

$$g_k(t) = \frac{\exp(kt') - 1}{k \exp(kt)} \quad (4)$$

where  $t' \equiv t$  for  $0 \leq t < \tau$  and  $t' \equiv \tau$  for  $t \geq \tau$ . Furthermore, an affine relationship between tissue Gd-DTPA

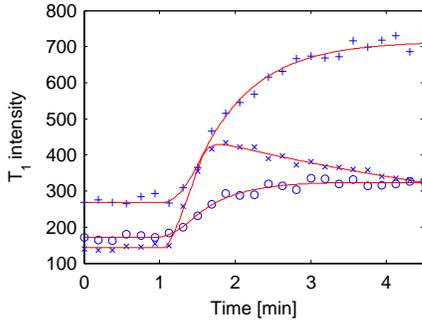


Fig. 1. Three examples for the employed pharmacokinetic model from [16] fitted to measured T1 intensity-time curves from a prostate data set. After a few baseline scans the contrast medium is administered and starts to accumulate in the tissue. The accumulation rate, the amplitude and the wash-out behavior characterize the microvascular structure of the imaged tissue.

concentration  $C_i(t)$  and enhanced signal  $S_\theta(t)$  can be assumed under suitable MR sequence parameters [16]. Thus the model for the T1-intensity-time curves is

$$S_\theta(t) = \begin{cases} S_0 & t \leq t_0 \\ S_0(1 + F C_i(t - t_0)) & t > t_0 \end{cases} \quad (5)$$

where  $S_0$  is the signal intensity obtained without CM and  $t_0$  is the lag time of the CM (also arrival or onset time). The constant  $F$  depends on several tissue properties and sequence parameters [16]. Since the factors  $F$  and  $K^{\text{trans}}K_{\text{in}}/V_p$  from Eqn. (3) cannot be distinguished based on the measured intensity-time curves only, these are summarized in the enhancement amplitude

$$A = F K^{\text{trans}}K_{\text{in}}/V_p. \quad (6)$$

In summary, five parameters of the described enhancement model have to be estimated given an observed intensity-time curve. For convenience these are summarized in the parameter vector  $\theta = (S_0, A, k_{\text{el}}, k_{\text{ep}}, t_0)$  in the following. The duration  $\tau$  of the CM injection is known and fixed to 30 s. Both  $S_0$  and  $A$  describe intensity values without unit. If not indicated otherwise, the unit for time is one frame ( $\Delta T = 11.25$  s) which means that the values for  $\tau$  and  $t_0$  have to be multiplied by  $\Delta T$  whereas the exchange rates  $k_{\text{el}}$  and  $k_{\text{ep}}$  have to be divided by  $\Delta T$  in order to obtain entities with units. Some examples of the model function fitted to patient data are shown in Fig. 1.

### III. PARAMETER ESTIMATION

#### A. Nonlinear Least Squares

Given the observed signal intensities  $\mathbf{y}_s = (y_i^s)^T$  for a certain voxel  $s$  at  $n$  discrete time points  $t_i$  the standard nonlinear regression model is

$$y_i^s = S_{\theta_s}(t_i) + \epsilon_i, \quad i = 1, 2, \dots, n \quad (7)$$

where  $\epsilon_i$  are independent, identically distributed normal random variables with mean 0 and variance  $\sigma^2$ . The vector  $\theta_s$  summarizes the  $p = 5$  unknown parameters with true value  $\theta_s^*$ .

Then, the  $n$ -vector  $\mathbf{f}(\theta_s) = (S_{\theta_s}(t_i))^T$  describes a  $p$ -dimensional surface in the sample space  $\mathcal{L} \subset \mathbb{R}^n$  which is

called the *solution locus* [25]. Its tangent space is spanned by the columns of the  $n \times p$  Jacobian matrix  $\mathbf{V}(\theta_s)$  with elements  $V_{ij} = \partial f_i / \partial \theta_j$ . Under this geometrical interpretation, the *maximum likelihood* estimate  $\hat{\theta}_s$  parametrizes the point on the solution locus  $\mathbf{f}(\theta_s)$  closest to the observation  $\mathbf{y}_s$ . It is obtained as solution to a nonlinear least squares (NLS) problem, i.e. by minimizing the sum of squares of residuals

$$J(\theta_s) = \|\epsilon(\theta_s)\|^2 = \|\mathbf{y}_s - \mathbf{f}(\theta_s)\|^2. \quad (8)$$

For large  $n$  the maximum likelihood estimate  $\hat{\theta}_s$  is consistent (unbiased), efficient (minimum variance) and it is normally distributed with covariance  $\Sigma = \sigma^2 (\mathbf{V}^T \mathbf{V})^{-1}$  [26]. The usual estimator of  $\sigma^2$  is

$$\hat{\sigma}^2 = J(\hat{\theta}_s) / (n - p). \quad (9)$$

#### B. Problems in NLS fitting of DE models

It is known that the maximum likelihood estimator can yield biased parameter estimates for small  $n$  [27], [28]. An approximate expression for the *parameter bias*, the expected difference between estimated and true parameter, can be derived as

$$\langle \hat{\theta}_s - \theta_s^* \rangle \approx \mathbf{b} = (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T \mathbf{d}, \quad (10)$$

with the  $n$ -vector  $\mathbf{d} = -\frac{\sigma^2}{2} (\text{trace}[(\mathbf{V}^T \mathbf{V})^{-1} \mathbf{H}_i])^T$  where the  $\mathbf{H}_i$  are  $p \times p$  Hessian matrices with elements  $H_{ijk} = \partial^2 f_i / \partial \theta_j \partial \theta_k$  [27], [28]. The bias not only vanishes for large  $n$  and for linear models for which  $H_{ijk} = 0$ , but also if  $\mathbf{d}$  is orthogonal to the tangent space  $\mathbf{V}$ . In a more detailed analysis, Bates and Watts [25], distinguish the orthogonal *intrinsic curvature* and the tangential *parameter-effects curvature* of the solution locus [26, Ch.4]. They show that the parameter bias only depends on the parameter-effects curvature which can, in principle, be annihilated by a suitable reparametrization  $\theta(\vartheta)$  of the model function. Unfortunately, finding such a reparametrization is usually difficult [29], [30], [25, Discussion by Dr. Beale].

In nonlinear least squares also the residuals can be biased in the sense that the expectation of the residuals is nonzero [27], in particular

$$\langle \epsilon(\hat{\theta}_s) \rangle \approx (\mathbf{I} - \mathbf{V}(\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T) \mathbf{d}. \quad (11)$$

Unlike the parameter bias, the *residual bias* only depends on the intrinsic curvature and can therefore not be reduced by reparametrization [26].

Apart from the parameter bias, determining the minimizer of the possibly non-convex objective function  $J(\theta_s)$  can also be difficult [18], [21]. Since for the employed pharmacokinetic model,  $J(\theta_s)$  in fact exhibits multiple local optima and long, narrow and curved valleys, an iterative optimization routine may stop before reaching the real minimum. The existence of such valleys is an indication that for some configurations at least two of the parameters have a similar effect on the objective function (ill-conditioning). More details on these general problems can be found in [26, Ch.3].

In order to give some insight into the non-convexity of the NLS objective function for the DE model used in the

present work, Fig. 2 shows a contour plot of  $f(A, t_0) = \min_{S_0, k_{el}, k_{ep}} J(\boldsymbol{\theta}_s)$  which visualizes the smallest attainable objective function for a certain amplitude  $A$  and lag time  $t_0$ .

Depending on the initialization, the iterative optimizer converges to different points on the energy landscape, both different from the actual optimum which is only reached using a very close initial guess. Certainly, prior knowledge tells that negative lag times are impossible which should be considered by using the parametric constraint  $t_0 > 0$ .

Different kinds of problems are demonstrated with the projection  $f(k_{el}, k_{ep}) = \min_{S_0, A, t_0} J(\boldsymbol{\theta}_s)$  in Fig. 3. In this example two local optima exist for  $k_{ep}$  and  $k_{el}$  which are connected through a steep curved valley (ill-conditioning). The model functions in this valley all look very alike. For example, the model functions in the right part of Fig. 3 that correspond to the points marked with circle and star in the left part are hardly different which makes parameter estimation difficult. Even worse, the two optima marked by a cross and a star seem to yield *identical* curves. In fact, an analysis of Eqn. (3) (or Eqn. (25)) shows that the two parameters  $k_{el}$  and  $k_{ep}$  can always be exchanged, yielding an identifiability problem. The latter can be avoided by assuming that the elimination rate  $k_{el}$  is always smaller than the exchange rate  $k_{ep}$ , however, the ill-conditioning problem remains.

In summary, the encountered statistical problems are:

- estimation variance: ill-conditioning and identifiability.
- estimation bias: high parameter-effects curvature along with small sample sizes.
- existence of non-global optima.

All these problems depend on the model as well as on the sampling points  $t_i$ . Choosing both optimally is the aim of *experimental design* [26]. But even if a model was derived in the optimal way, some problems may persist.

Although we have demonstrated problems only for the particular DE model proposed in [16], they are not unusual for this class of models. Using the model by Tofts *et al.* [15], for example, Orton *et al.* [18] encounter similar problems in their work and present an example with two local optima. Further evidence is provided by Ahearn *et al.* [21] as well as Schmid *et al.* [19]. Finally, more complex (multi-compartment) models with higher-order dynamics cannot be expected to lead to more well-behaved objective functions since models with exponentially decaying signal components are known to be problematic (cf. [26, Ch.8]). Simply increasing temporal resolution does not always help, either. An identifiability problem, for example, could not be resolved by increasing the number of sampling points (cf. example in Fig. 3).

A complementary approach is needed. As seen above, prior knowledge is very valuable and should be used to avoid problematic regions of the parameter space. In addition to this *single-voxel prior knowledge*, which can be applied for every voxel independently, we propose to incorporate *spatial prior knowledge* by using a spatial smoothness prior in the form of a generalized Gaussian Markov random field (GGMRF) [31]. Then, the model fit in each voxel is additionally supported by data from its local neighborhood.

### C. GGMRF: A Generalized Gaussian Markov Random Field Prior

The generalized Gaussian Markov random field [31] is a Markov random field [32], [33] with particular compatibility functions (the logarithms of which are known as potentials). Every voxel in the region of interest is represented by a site  $s \in S$  which is associated with the vector-valued random variable  $\boldsymbol{\theta}_s$ . Like in the single-voxel case, the observation likelihood is Gaussian. Imposing the spatial GGMRF prior on the parameter maps  $\boldsymbol{\theta}$  yields a joint distribution over  $\mathbf{y}$  and  $\boldsymbol{\theta}$  in the form of the Gibbs distribution:

$$\Pr(\boldsymbol{\theta}, \mathbf{y}) = \frac{1}{Z} \prod_{s \sim t} \Psi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_t) \prod_s \Phi(\boldsymbol{\theta}_s, \mathbf{y}_s), \quad (12)$$

where  $\mathbf{y}$  and  $\boldsymbol{\theta}$  are vector variables obtained by stacking the site vector variables  $\mathbf{y}_s$  and  $\boldsymbol{\theta}_s$ .  $Z$  is the global normalizer (partition function) and  $s \sim t$  denotes pairs of neighboring sites according to the employed neighborhood system. The compatibility functions  $\Phi(\boldsymbol{\theta}_s, \mathbf{y}_s)$  and  $\Psi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_t)$  are defined by the potentials

$$\log \Phi(\boldsymbol{\theta}_s, \mathbf{y}_s) = -\frac{1}{2\sigma^2} \|\mathbf{y}_s - \mathbf{f}(\boldsymbol{\theta}_s)\|^2 \quad (13)$$

$$\log \Psi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_t) = -\frac{\alpha}{2} \|\mathbf{W}(\boldsymbol{\theta}_s - \boldsymbol{\theta}_t)\|_p^p \quad (14)$$

with the spatial coupling factor  $\alpha \geq 0$  and the  $p$ -norm  $\|\cdot\|_p$  ( $1 < p \leq 2$ ) [31]. Note that  $p = 1$  is excluded since the L1-norm is not differentiable at the origin.  $\mathbf{W}$  is a diagonal weighting matrix which accounts for the different scales and variability of the parameters in  $\boldsymbol{\theta}_s$  and which can be used to adjust the smoothness of individual parameter maps.

The application of a GGMRF allows to vary continuously between a smoothing Gaussian MRF prior ( $p = 2$ ) and an edge-preserving MRF ( $p \rightarrow 1$ ) with properties comparable to a weighted median filter [31]. Furthermore, the GGMRF potential defined by (14) is convex and, as opposed to robust alternatives such as the Huber potential [34], [35], it does not have an extra threshold parameter. Connections between robust statistics and edge-preservation of different priors have been explored in [36].

### D. Block Iterated Conditional Modes

Given an observed DCE-MRI sequence  $\mathbf{y} = (\mathbf{y}_s)$ , the maximum a posteriori (MAP) estimate  $\hat{\boldsymbol{\theta}}$  is then found by minimizing

$$I(\boldsymbol{\theta}) = \sum_{s \in S} \|\mathbf{y}_s - \mathbf{f}(\boldsymbol{\theta}_s)\|^2 + \sigma^2 \alpha \sum_{s \sim t} \|\mathbf{W}(\boldsymbol{\theta}_s - \boldsymbol{\theta}_t)\|_p^p \quad (15)$$

When comparing this objective function to Eqn. (8) it becomes evident that a trade-off between the *data term* (first sum) and the *prior term* (second sum) is sought this way.

Minimizing Eqn. (15) is a challenging optimization problem since, for example, a DCE-MR image of  $100 \times 100$  voxels would already yield a parameter vector  $\boldsymbol{\theta}$  with  $5 \cdot 10^4$  entries and thus an optimization problem in a 50,000-dimensional space. However, the problem is sparse in the sense that most of the  $\boldsymbol{\theta}_s$  are not directly coupled. The MRF framework provides

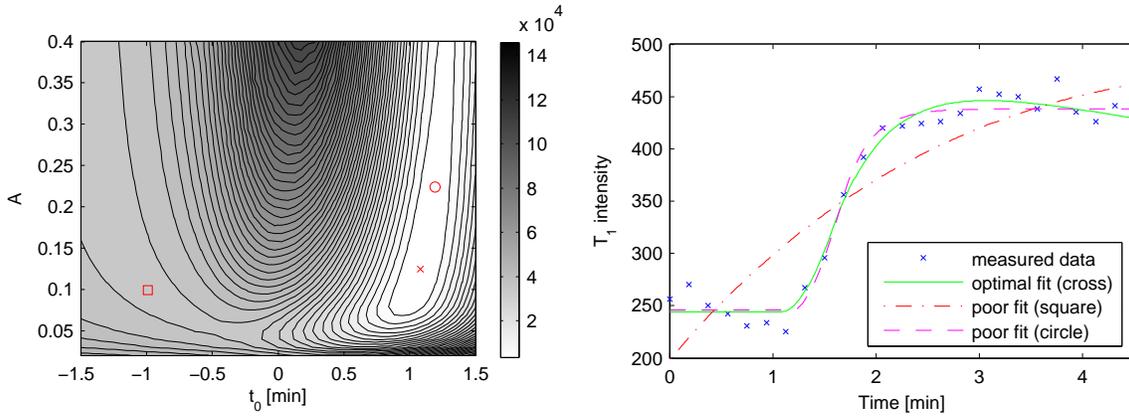


Fig. 2. *Left*: Projected NLS objective  $f(A, t_0) = \min_{S_0, k_{el}, k_{ep}} J(\theta)$  for an example from the prostate data set. When initialized with  $\theta_0 = (S_0, A, k_{el}, k_{ep}, t_0) = (100, 0.3, 0.01 \text{ min}^{-1}, 2 \text{ min}^{-1}, 1 \text{ min})$  the iterative optimization routine converges to the point marked with a square (“Change in the residual smaller than the specified tolerance”, TolFun) and when initialized with  $\theta_0 = (250, 0.3, 0.01 \text{ min}^{-1}, 5 \text{ min}^{-1}, 1 \text{ min})$  it converges to the point marked with a circle (“Change in X smaller than the specified tolerance”, TolX). The cross marks the point of the true minimum  $\hat{\theta} = (244.174, 0.124, 0.0958 \text{ min}^{-1}, 1.807 \text{ min}^{-1}, 1.082 \text{ min})$ . *Right*: Original data along with the three model fits that correspond to the points marked in the left figure.

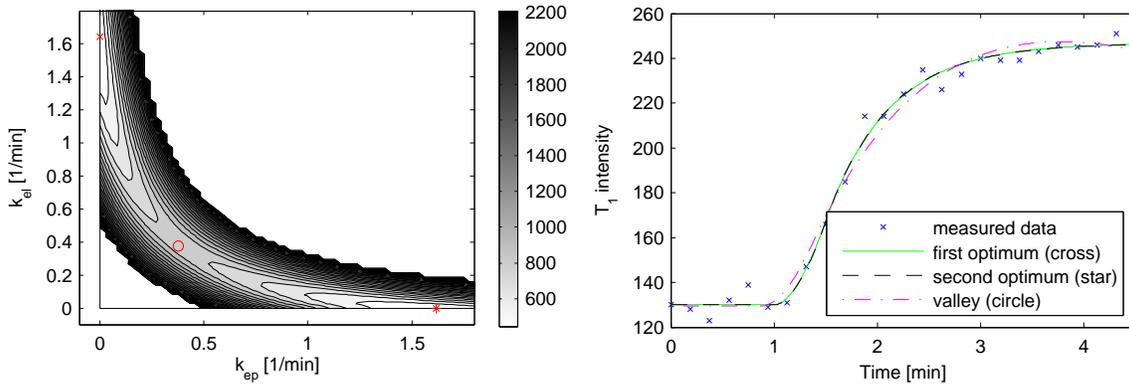


Fig. 3. *Left*: Projected NLS objective  $f(k_{el}, k_{ep}) = \min_{S_0, A, t_0} J(\theta)$  s.t.  $k_{ep} > 0$  and  $k_{el} > 0$  for another example from the prostate data set. *Right*: The corresponding data fits reveal that the two optima (cross and star) yield the same model curve which reflects that the employed model is not identifiable.

special algorithms that can exploit this sparsity such as the ICM (iterated conditional modes) algorithm [32].

Here, we use a generalized ICM algorithm which will be shown to converge faster than the standard ICM approach. As the algorithm considers collections of sites instead of single sites at each step, we call this approach *Block-ICM*. Given an arbitrary subset of sites  $\tilde{S} \subseteq S$ , it follows from the Hammersley-Clifford theorem [33] that the posterior distribution  $\Pr(\theta | \mathbf{y}) = \Pr(\theta_S | \mathbf{y})$  can be factored as

$$\Pr(\theta_S | \mathbf{y}) = \Pr(\theta_{\tilde{S}} | \theta_{\partial\tilde{S}}, \mathbf{y}) \Pr(\theta_{S \setminus \tilde{S}} | \mathbf{y}) \quad (16)$$

where  $\partial\tilde{S} = \{t : t \in S \setminus \tilde{S}, t \sim s, s \in \tilde{S}\}$  is the border of  $\tilde{S}$ . Increasing the first factor  $\Pr(\theta_{\tilde{S}} | \theta_{\partial\tilde{S}}, \mathbf{y})$  with respect to  $\theta_{\tilde{S}}$  certainly cannot decrease  $\Pr(\theta_S | \mathbf{y})$  since the second factor  $\Pr(\theta_{S \setminus \tilde{S}} | \mathbf{y})$  does not depend on any of the variables in  $\theta_{\tilde{S}}$ . Hence, the MAP problem (15) can be solved iteratively by solving a series of smaller MAP problems over subsets of sites

$$\theta_{\tilde{S}}^{(k+1)} = \underset{\theta_{\tilde{S}}}{\operatorname{argmin}} \left[ \sum_{s \in \tilde{S}} \|\mathbf{y}_s - \mathbf{f}(\theta_s)\|^2 \right]$$

$$\begin{aligned} & + \sigma^2 \alpha \sum_{\substack{s \sim t \\ s, t \in \tilde{S}}} \|\mathbf{W}(\theta_s - \theta_t)\|_p^p \\ & + \sigma^2 \alpha \sum_{\substack{s \sim t \\ s \in \tilde{S} \\ t \in \partial\tilde{S}}} \left\| \mathbf{W}(\theta_s - \theta_t^{(k)}) \right\|_p^p \end{aligned} \quad (17)$$

The Block-ICM algorithm can also be viewed as an iterative coordinate descent approach in which the potentially intersecting subsets  $\tilde{S}^{(k)}$  redefine generalized coordinates  $\theta_{\tilde{S}}^{(k)}$  for every descent step. Also, it suffices to find a realization  $\theta_{\tilde{S}}^{(k+1)}$  which decreases the objective (17) instead of finding the exact minimum in every descent step. The proposed procedure still converges to a local minimum.

Shape, size and update sequence of the subsets  $\tilde{S}$  are design parameters of the Block-ICM algorithm and should be chosen so as to trade off the problem size in each step against the number of sweeps required for convergence. If, for example, each of the subsets  $\tilde{S}^{(k)}$  contains only one site  $s$ , the standard ICM algorithm is recovered [32] which is known to often converge rather slowly. If, on the other hand, only one (sub)set  $\tilde{S} \equiv S$  is chosen, the entire MAP problem (15) is obtained.

Hence, small subsets of sites should be chosen depending on the size of the local neighborhood and the strength of the mutual influence. Because of the locality of this influence, the size of the subsets does not have to be increased with growing lattices, yielding an algorithm which is linear in the number of sites. The Block-ICM approach is very similar to domain decomposition methods [37] which are often used in combination with multigrid methods for solving partial differential equations or associated variational problems [38].

#### IV. EXPERIMENTAL SETUP

##### A. Data Sets

1) *Patient Data*: Patient data from an ongoing prostate study have been collected at the German Cancer Research Center (Heidelberg, Germany). The data were acquired on a clinical 1.5T scanner (Magnetom Symphony; Siemens Healthcare, Erlangen, Germany) with a disposable endorectal coil (MRInnervu; Medrad Inc., Indianola, PA, USA).

For the DCE-MR data, 10 transverse slices have been defined (FOV 200 mm  $\times$  125 mm, slice thickness 3 mm, 1 mm gap) from which 25 dynamic image data sets were measured with a temporal resolution of  $\Delta T = 11.25$  s using an optimized 2D FLASH sequence (TR/TE=125 ms/3.11 ms, flip angle 90°, matrix size 128  $\times$  60, sample percentage 75%). Subsequent interpolation finally yielded image slices with 256  $\times$  160 voxels. After 33.75 s a total dose of 0.1 mmol Gadolinium-DTPA per kg body weight was administered intravenously by constant rate infusion within  $\tau = 30$  s. From the 37 patients which have been available for the present study, regions of about 100  $\times$  100 voxels have been selected in slices that contained suspicious tissue as determined from T1-MRI and MR spectroscopic imaging.

2) *Simulation Studies*: A Monte-Carlo study has been conducted using two sets of simulated DCE-MR images allowing for a detailed analysis of estimation errors and the influence of hyper parameters.

In analogy to the “wedding cake” example used in [36], ground truth maps with 60  $\times$  60 voxels based on three sets of parameters have been created for the first set of simulations (cf. Fig. 5). For the innermost square  $\theta_1 = (200, 0.50, 0.001, 0.7, 5.2)$ , for the middle part  $\theta_2 = (100, 0.25, 0.050, 0.3, 4.6)$  and for the border  $\theta_3 = (150, 0.76, 0.008, 1.2, 5.5)$  has been used. Sequences of  $n = 25$  DCE-MR images at intervals of  $\Delta T = 11.25$  s have been generated based on the model function in Eqn. (5).  $R = 5$  repetitions of the data have been stored after adding independent Gaussian noise with mean zero and standard deviation  $\sigma = 20$  which reflects an upper noise level encountered in the patient data.

In order to obtain a more realistic spatial distribution, a second set of data has been simulated based on a detail from a real prostate DCE-MR image with particularly low noise (“real detail”, cf. Fig. 7). The single voxel fit to the original data has been used as “ground truth” and  $R = 100$  repetitions of simulated data have been generated by adding Gaussian noise with standard deviation  $\sigma = 10$ .

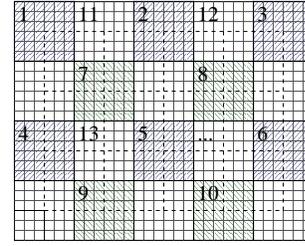


Fig. 4. Scheme of blocks and update schedule used for the Block-ICM algorithm. In every odd sweep, square blocks (6  $\times$  6 sites in the figure) are visited following the pattern indicated by the numbering. The even sweeps are performed in the same way but shifted by half a block (dashed squares).

##### B. Optimization Details

Parameter estimates have been calculated with the single voxel as well as two versions of the GGMRF approach, the GGMRF-L2 with an L2-norm ( $p = 2$ ) and the GGMRF-L1 with  $p = 1.01$ . All calculations have been performed on a dual core 2.4GHz Intel PC with 2GB of main memory. The algorithms have been implemented with Matlab R2006b using interior trust region methods from the Matlab optimization toolbox [39].

In all approaches, box constraints have been enforced:  $S_0 > 0$ ,  $A > 0$ ,  $k_{ep} > k_{el} > 0$  and  $3 < t_0/\Delta T < n$ . The same initialization scheme has been used for all experiments and voxels. The mean of the first three DCE-MR images has been used as the starting value for the unenhanced T1-intensity  $S_0$ . The other parameters have been initialized with  $A = 0.3$ ,  $k_{el} = 0.003/\Delta T$ ,  $k_{ep} = 0.4/\Delta T$  and  $t_0 = 5.5\Delta T$ , respectively.

For the single voxel approach the Matlab function “lsqnonlin” with analytically derived Jacobian has been used. A maximum number of 500 iterations per voxel was allowed for (“MaxIter”) and the termination tolerances on the function value (“TolFun”) as well as on the parameters (“TolX”) were set to  $10^{-9}$ . Usually, the optimizer converged within less than 100 iterations.

The GGMRF approaches have been used based on a 2D regular lattice with first-order neighborhood system (four nearest neighbors) [40]. For the Block-ICM algorithm, the whole lattice was subdivided into two sets of square blocks such that the second set had a horizontal and vertical displacement of three voxels (the dashed squares in Fig. 4). In every odd sweep, the blocks in the first set were visited in a doubly-quincunx pattern as indicated by the numbering in Fig. 4. In every even sweep, the same procedure was performed on the second, shifted set of blocks. A total of 14 sweeps were performed. Each block was optimized using the “lsqnonlin” function with Jacobian for the GGMRF-L2 and the “fmincon” function with supplied gradient for the GGMRF-L1 prior. Since the subproblems in the Block-ICM approach still exhibit some sparsity, the sparse arithmetic capabilities of Matlab and the optimization toolbox have been exploited in addition. To prevent premature convergence to a local minimum, the maximum number of allowed optimization iteration has been confined in the first few sweeps. In particular, the schedule (5, 5, 10, 10, 30, 30, 100, 100, 100, 100, 500, 500, 500, 500) has been used. The termination tolerances were again set to  $10^{-9}$ .

### C. Hyperparameters

The GGMRF model defines three hyperparameters. The standard deviation  $\sigma$  of the signal noise, the parameter weighting matrix  $\mathbf{W}$  and the spatial coupling factor  $\alpha$ . All these parameters have been determined once for the prostate and the simulated data respectively and have subsequently been used for all examples.

Equation (15) reveals that the signal noise  $\sigma$  and the spatial coupling factor  $\alpha$  could in fact be summarized in a common factor which again could be absorbed into  $\mathbf{W}$ . Nevertheless, it is convenient to distinguish the three hyperparameters since all of them have a different meaning.

For the prostate data,  $\sigma$  has been determined from the single voxel solution of a representative example using Eqn. (9). For the simulated data, the true standard deviation of the simulated Gaussian noise ( $\sigma = 20$ ) has been used.

The diagonal of the parameter weighting matrix  $\mathbf{W}$  has been determined based on the normalizing assumption that equal weight is given to each parameter, *i.e.*

$$W_{ii} = \left( \left\langle (\theta_i^s - \theta_i^t)^2 \right\rangle_{\text{Pr}(\theta | \mathbf{y})} \right)^{-\frac{1}{2}} \quad i = 1 \dots p \quad (18)$$

where the expectation is taken with respect to the posterior parameter distribution. Assuming that this expectation is the same for all pairs  $s$  and  $t$  and can be replaced by a spatial average (stationarity and ergodicity, [41]),  $\mathbf{W}$  can be estimated from the posterior parameter map  $\hat{\theta}$  obtained with the single voxel approach.

In particular, we analyzed robust mean values of the squared parameter differences  $(\theta_i^s - \theta_i^t)^2$  from a single voxel solution. Based on this analysis, we have chosen  $W_{S_0} = 0.01$ ,  $W_A = 2$ ,  $W_{k_{e1}} = 20$ ,  $W_{k_{ep}} = 1$  and  $W_{t_0} = 1$  for the prostate data as well as for the simulated data.

The influence of the spatial coupling factor  $\alpha$  has been examined on the simulated data. Based on these results and some preliminary experiments with patient data, a suitable value for  $\alpha$  has been chosen for the prostate data set. The respective values of  $\alpha$  employed for each of the experiments are specified in the following.

## V. RESULTS

### A. Simulation Studies

The exchange rate  $k_{ep}$  is an important parameter for the identification of tumor tissue and has therefore been preferred for the evaluation. In Fig. 5, three estimates for an example of the simulated “wedding cake” data along with the ground truth  $k_{ep}$ -map are shown. The single voxel estimate exhibits many voxels for which a poor solution was found. The  $k_{ep}$ -map appears rather noisy and speckled, in particular in the middle ring. The corresponding estimates using a GGMRF-L2 prior with  $\alpha = 1$  and a GGMRF-L1 prior with  $\alpha = 1$  show visible improvements. Slight tendencies for edge blurring, *i.e.* oversmoothing, can be observed when using the GGMRF-L2 prior. The GGMRF-L1 solution ( $\alpha = 1$ ) yields the best reconstruction of the ground truth.

The influence of the spatial coupling factor  $\alpha$  on the absolute bias, the standard deviation and the root mean square error

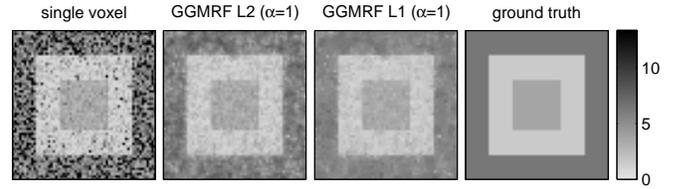


Fig. 5. Comparison of  $k_{ep}$  [ $\text{min}^{-1}$ ] estimates on the simulated “wedding cake” data. In some parts, the single voxel approach produces very noisy and speckled parameter maps. Using a GGMRF prior with L2 ( $p = 2$ ,  $\alpha = 1$ ) or L1 norm ( $p = 1.01$ ,  $\alpha = 1$ ) improves the results considerably.

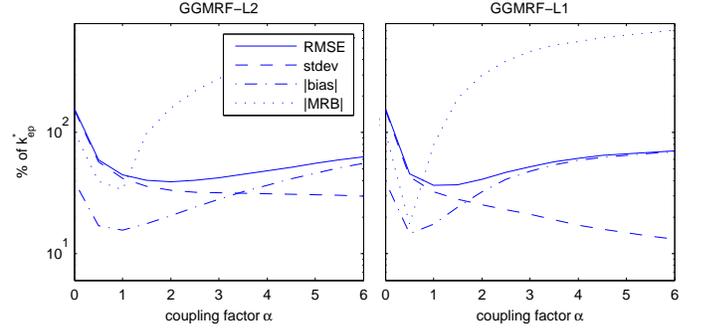


Fig. 6. Root mean square error (RMSE), absolute bias and standard deviation of the  $k_{ep}$  estimates for the “wedding cake” data (Fig. 5) and absolute mean residual bias (MRB) using GGMRF-L2 (left) and GGMRF-L1 (right). As compared to the single voxel approach (coupling factor  $\alpha = 0$ ) both spatial priors can reduce bias as well as standard deviation of the fit to the data. While the standard deviation keeps decreasing with increasing  $\alpha$ , the bias reascends after an initial reduction, reflecting oversmoothing at the sharp edges. The MRB behaves very similar to the bias and is minimum for about the same value of  $\alpha$ . In the limit  $\alpha \rightarrow \infty$  the bias, stdev and RMSE are 154.32%, 1.85% and 154.33% respectively. Thus, the RMSE reascends to a value similar to the single voxel solution but trading standard deviation against bias.

(RMSE) is depicted in Fig. 6. These have been calculated, in steps of  $\Delta\alpha = 0.5$ , from the middle part of the estimated  $k_{ep}$ -maps according to

$$\text{bias} = \frac{100\%}{k_{ep}^*} \left( \left\langle \hat{k}_{ep} \right\rangle_N - k_{ep}^* \right) \quad (19)$$

$$\text{stdev} = \frac{100\%}{k_{ep}^*} \sqrt{\left\langle \left( \hat{k}_{ep} - \left\langle \hat{k}_{ep} \right\rangle_N \right)^2 \right\rangle_N} \quad (20)$$

$$\text{RMSE} = \frac{100\%}{k_{ep}^*} \sqrt{\left\langle \left( \hat{k}_{ep} - k_{ep}^* \right)^2 \right\rangle_N} \quad (21)$$

The expectations are taken with respect to the empirical distribution such that  $\langle \cdot \rangle_N = \frac{1}{N} \sum_{j=1}^N \cdot$  where  $N$  is the number of examples used for the estimate. Note that bias and standard deviation provide a decomposition of the RMSE in the sense that  $\text{RMSE}^2 = \text{bias}^2 + \text{stdev}^2$ . In addition, the *mean residual bias* has been estimated as

$$\text{MRB} = \left\langle \frac{1}{n} \sum_i \epsilon_i(\hat{\theta}) \right\rangle_N \quad (22)$$

with  $\epsilon$  as defined in Eqn. (8). Figure 6 shows  $\text{MRB}(\alpha)/\text{MRB}(0)$ , *i.e.* the mean residual bias of the GGMRF approaches relative to the SV approach.

In Fig. 6, the standard deviation improves more and more with increasing  $\alpha$  for both GGMRF priors. The bias initially

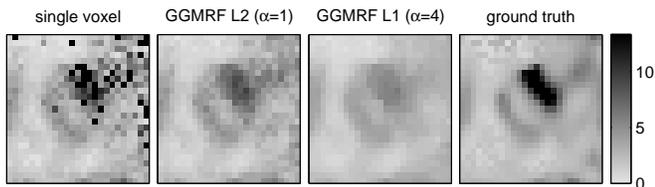


Fig. 7. Comparison of  $k_{ep}$  [ $\text{min}^{-1}$ ] estimates on the simulated “real detail” data. Again, the single voxel approach produces noticeable speckles. Although, using a GGMRF prior with L2 or L1 norm clearly underestimates the  $k_{ep}$  values in the dark elongated enhancement area, it reveals the original structure of the  $k_{ep}$  ground truth map which is obscured in the single voxel solution.

reduces but then, beyond a certain value of  $\alpha$ , reascends again. For the GGMRF-L2 this critical value is about  $\alpha = 1$  while for the GGMRF-L1 it is about  $\alpha = 0.5$ . Limit values for  $\alpha \rightarrow \infty$  are obtained from the optimum constant parameter maps and are a bias of 154.32%, a standard deviation of 1.85% and a RMSE of 154.33%.

Figure 7 shows the ground truth  $k_{ep}$ -map and three estimates for an example of the simulated “real detail” data. In the single voxel estimate the original spatial structure is hard to recognize while it is clearly revealed in the GGMRF solutions. However, the  $k_{ep}$  values in the elongated enhancement area are all underestimated using the spatial priors. The bias-variance decomposition presented in Fig. 8 also shows the increased bias for the GGMRF approaches while low and homogeneous standard deviations are observed in all areas of the estimated  $k_{ep}$ -maps. For the single voxel estimates, huge bias as well as standard deviation is obtained in the elongated enhancement area (first row Fig. 8). The histograms of estimates from a voxel within this area in Fig. 9 reveal that this is caused by severe outliers. Unlike the GGMRF estimates which are symmetrically Gaussian-like distributed, the single voxel estimates are heavily skewed. Ignoring outliers, the single voxel estimates are less biased than the GGMRF estimates but still tend to underestimate the ground truth value of  $k_{ep} = 12.0 \text{ min}^{-1}$ .

## B. Patient Data

In Fig. 10,  $k_{ep}$ -maps from several patients are shown. They allow for a qualitative comparison of the voxel-wise (left column) and the GGMRF approaches (middle and right column). The depicted range of values for  $k_{ep}$  was set to the interval between  $-0.05 \text{ min}^{-1}$  (white) and  $29.3 \text{ min}^{-1}$  (black) for all images. Furthermore, the outline of the prostate gland is indicated by the red contour. For all patients the noise is visibly reduced in the bright areas (low  $k_{ep}$ ) when using a spatial prior. Although no radical changes in the  $k_{ep}$ -maps are observed, using a spatial prior gives a much clearer picture. In the upper left corner of patient P30, for example, the SV solution shows an area which could easily be mistaken for an area of increased  $k_{ep}$  values, an indication for tumor. Using a spatial prior resolves this issue. Several similar but less extreme examples can be found in the other patients in Fig. 10.

Average mean residual biases (MRB) for the the SV and GGMRF approaches as well as standard deviations are listed in

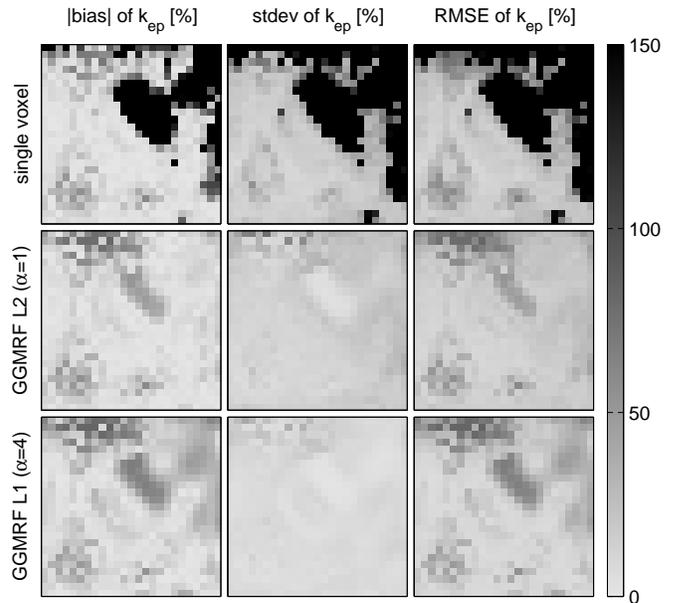


Fig. 8. Absolute bias, standard deviation (stdev) and root mean square error (RMSE) of  $k_{ep}$  estimates obtained from  $R = 100$  repetitions of the simulated “real detail” data. While both spatial priors cause considerable bias in the elongated enhancement area, the single voxel approach exhibits bias and standard deviation that exceed the displayed range by far. For one of these voxels, Fig. 9 shows a histogram of the  $k_{ep}$  estimates allowing for a more detailed analysis.

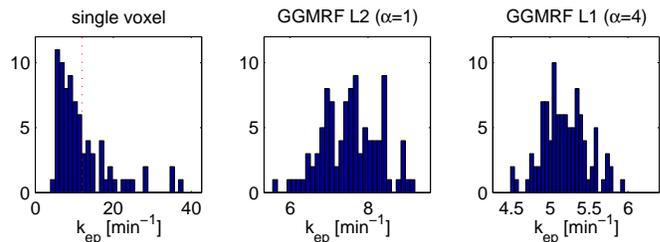


Fig. 9. Histograms of  $k_{ep}$ -estimates for a voxel from the elongated enhancement area of the simulated “real detail” data (cf. Fig. 8) with ground truth value  $k_{ep} = 12.0 \text{ min}^{-1}$  (vertical line). From the single voxel histogram 22 outliers with  $k_{ep} > 50 \text{ min}^{-1}$  have been excluded. Ignoring outliers, the single voxel estimates are less biased than the GGMRF estimates albeit a tendency for systematic underestimation is observed.

Table I. Only voxels within the prostate gland have been considered. For each of the 37 patients, the GGMRF approaches yielded a significantly smaller (absolute) MRB than the SV approach.

Also, computation times have been recorded for the SV approach ( $t_{SV}$ ) as well as the GGMRF approaches ( $t_{L1}$ ,  $t_{L2}$ ). The SV approach took about 740 ms per voxel on average; total computation times varied between 0.61 h and 2.81 h ( $1.79 \text{ h} \pm 0.49 \text{ h}$  on average). The GGMRF approach with L2 prior took about 871 ms per voxel which is only a factor of about 1.18 slower than the SV approach. In total, computations took between 0.76 h and 3.41 h ( $2.08 \text{ h} \pm 0.57 \text{ h}$ ). Notably, the GGMRF-L2 approach ran faster than the SV approach in 5 cases. The factor  $t_{L2}/t_{SV}$  ranged between 0.65 (P7) and 1.44 (P19) which means that computations took at most 1.44 times longer than with the SV approach. Using the GGMRF-L1 prior, an average computation time of about 1.12 s per

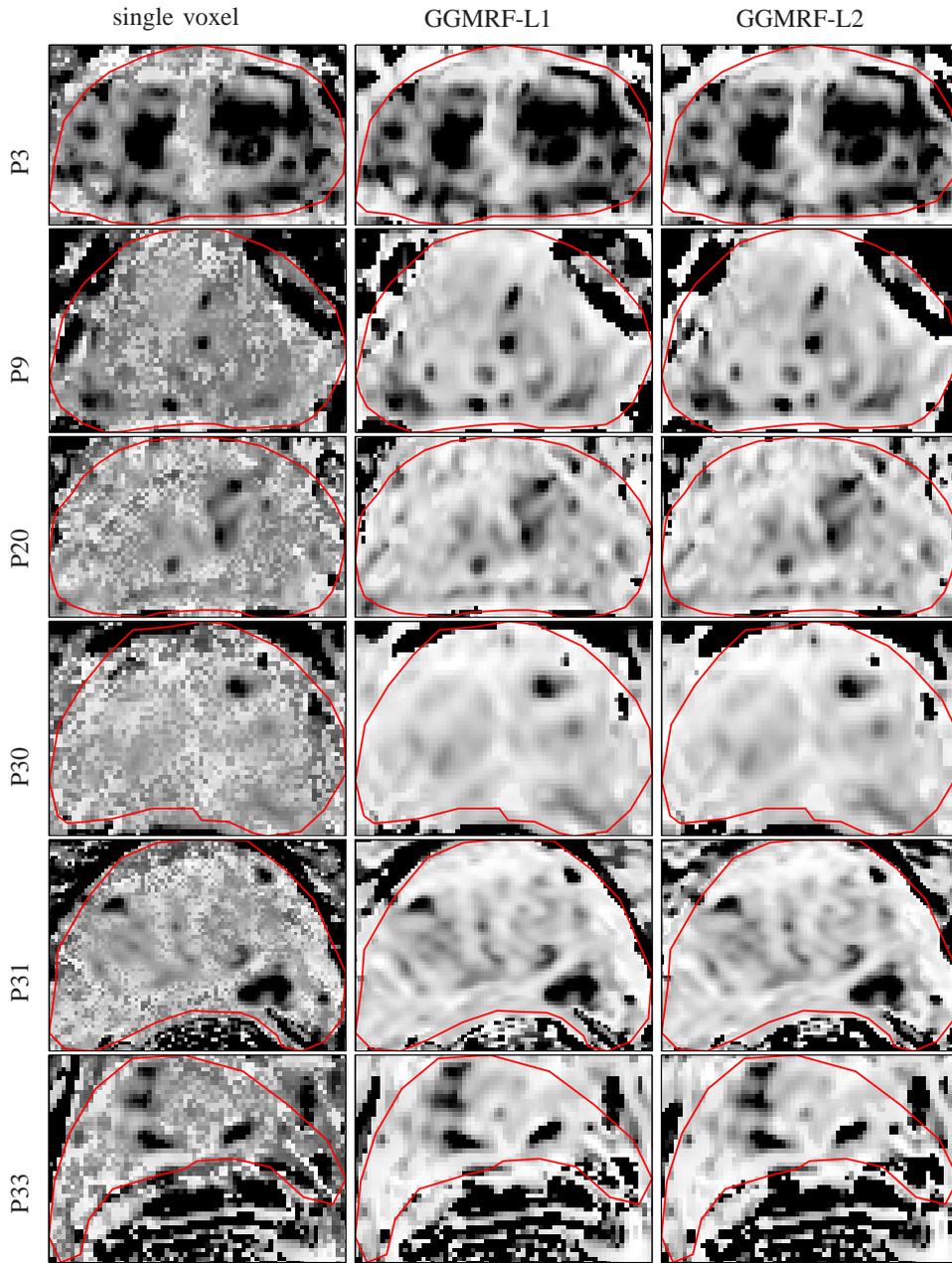


Fig. 10. Comparison of  $k_{ep}$ -maps estimated with the single voxel, the GGMRF-L1 ( $\alpha = 4$ ) and the GGMRF-L2 ( $\alpha = 1$ ) approaches. The parameter estimates obtained with the single voxel approach appear, like in the simulation study, very noisy and speckled. Both GGMRF approaches (L1 and L2) can reduce estimation noise and allow for an easier interpretation of the  $k_{ep}$ -map. Structures that are present in the single voxel solution are preserved without blurring by both of the GGMRF priors.

TABLE I  
MEAN RESIDUAL BIASES (MRB) AND COMPUTATION TIMES FOR  
SINGLE-VOXEL AND GGMRF ESTIMATES BASED ON 37 PATIENTS.

	average	standard deviation
$MRB_{SV}$	-0.2492	0.1287
$MRB_{L2}$	$7.24 \times 10^{-4}$	$8.74 \times 10^{-4}$
$MRB_{L1}$	$-9.60 \times 10^{-3}$	$1.67 \times 10^{-2}$
$t_{SV}$	739.63 ms	82.21 ms
$t_{L2}$	871.17 ms	152.73 ms
$t_{L1}$	1119.81 ms	261.98 ms

voxel was obtained, a factor of about 1.51. Total computation times ranged from 0.82 h to 5.13 h ( $2.76 \text{ h} \pm 1.07 \text{ h}$ ). In several cases computations took more than twice (up to 2.62 times) as long as with the SV approach. An explanation for this behavior is deferred to the later discussion.

A more detailed analysis of the resulting model fits was performed based on patient P1. Figure 11 shows all but the  $S_0$  parameter map of this patient for the three compared approaches. The  $S_0$  maps have been omitted since no differences between the three versions could be observed. The  $k_{ep}$ -maps for patient P1 show the same improvements as those of the other patients shown in Fig. 10. Also, the SV parameter maps

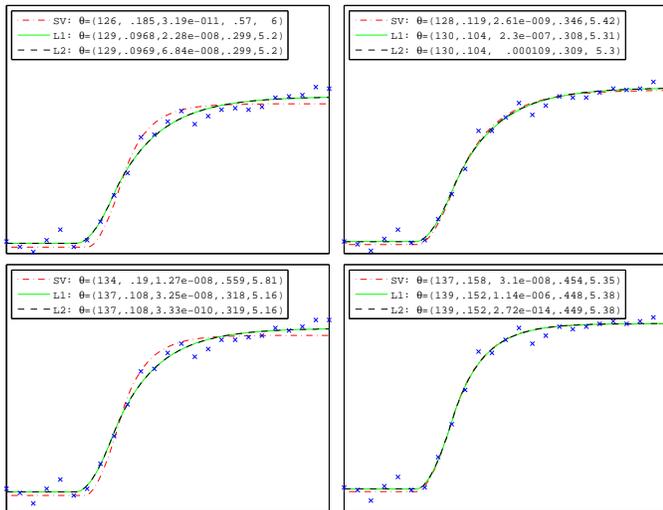


Fig. 12. Model fits using the single voxel, the GGMRF-L1 and the GGMRF-L2 approaches for the voxels within the red square in Fig. 11. The GGMRF approaches yield indistinguishable curves in all voxels. Reasonable but different results are obtained using the single voxel approach. Except for the lower left voxel, the parameter estimates for  $k_{ep}$ ,  $A$  and also  $t_0$  differ significantly between the single voxel and the GGMRF solutions. In the upper left voxel, both amplitude  $A$  and exchange rate  $k_{ep}$  even differ by a factor of two. The upper right voxel demonstrates that clearly different parametrizations may yield very similar curves (very much like for the examples shown in Fig. 2 and Fig. 3).

for  $A$ ,  $k_{el}$  and  $t_0$  exhibit a lot of noise which is removed using the spatial priors. Apart from that, an interaction between the parameter maps is visible in the SV solution, meaning that speckles or groups of speckles are visible in multiple maps, for example in the  $A$  and  $k_{el}$  maps.

Figure 12 shows the intensity-time curves from the four adjacent voxels marked with a square in Fig. 11 together with the parameter estimates and the corresponding model curves. Four voxels have been chosen that are located at an edge of the  $k_{ep}$  and  $A$  parameter maps and thus show curve fits in different regions of the parameter space. The results obtained with the GGMRF approaches are virtually identical in terms of the fitted curve as well as in the estimated parameters. The SV solutions look reasonable and only appear suboptimal when contrasted with the GGMRF solutions which seem to have less correlated residuals. This is particularly true for the upper left voxel where the parameter estimates for the amplitude  $A$  and the exchange rate  $k_{ep}$  differ by a factor of about two between the SV and the GGMRF solutions. Also the estimates for the lag time  $t_0$  are clearly different. In the lower right voxel, the SV curve is hardly different from the GGMRF curves and also the parameter estimates are quite similar for all three solutions. Thus, the four adjacent voxels lying in different areas of the parameters space are not equally difficult to fit.

Finally, Fig. 13 provides convergence results that show the influence of using different block sizes in the proposed Block-ICM algorithm. The special case of  $1 \times 1$  blocks results in the conventional single-site ICM algorithm which clearly converges much slower than Block-ICM using bigger blocks. Hardly any difference in convergence speed is observed between the block sizes of  $4 \times 4$ ,  $6 \times 6$  and  $9 \times 9$ .

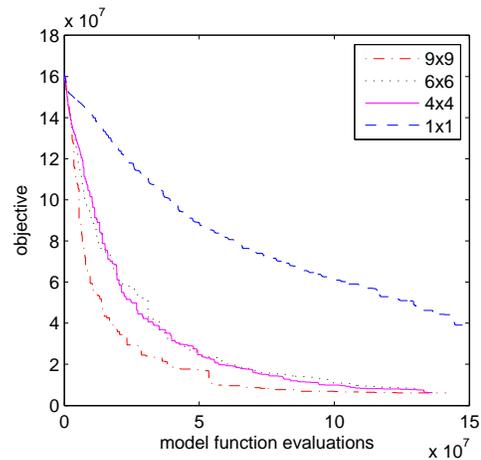


Fig. 13. Convergence behavior of ICM ( $1 \times 1$ ) and Block-ICM for different choices of block sizes. Block-ICM clearly outperforms the common ICM algorithm ( $1 \times 1$ ). Very similar performance is obtained for the block sizes  $4 \times 4$ ,  $6 \times 6$  and  $9 \times 9$ .

## VI. DISCUSSION

### A. Variance, Bias and Multiple Optima

Three problems in the estimation of DE model parameters have been pointed out in section III-B: high estimation variance, bias and the problem of multiple optima. Using a spatial prior has a different effect on each of these problems:

a) *Variance*: The results depicted in Fig. 6 show that the standard deviation of the estimates continuously reduces with increasing spatial coupling. This reduction is reflected in parameter maps with less noise which can be observed when using spatial prior knowledge on the prostate data, for example in Fig. 11. One way of explaining the reduction of estimation variance when using a spatial prior is the usage of an increased amount of data for the estimation at each voxel. Another way of explaining the gain is by noting that ill-conditioning, which usually leads to high estimation variance, only occurs in particular regions of the parameter space. Therefore, ill-conditioning is addressed best if the true parameters of neighboring voxels differ slightly.

b) *Bias*: The results on the simulated “real detail” data showed that considerable bias can occur when using a spatial prior on very noisy data (cf. Fig. 8). Surprisingly, the simulation studies also showed that in certain regions of the parameter space and for small values of the coupling factor  $\alpha$ , the parameter bias can actually be reduced (Fig. 8, Fig. 6). This is explained by observing that the parameter bias (cf. Eqn. (10)) reduces with an increasing number of samples. In an extreme form of spatial smoothness ( $\alpha \rightarrow \infty$ ), two neighboring parameters  $\theta_s$  and  $\theta_t$  would have to be equal. Then, the measurement at voxel  $t$  could be regarded as a repetition of the measurement at voxel  $s$ . Using Eqn. (10) and Eqn. (11) it is not difficult to show that this would halve the parameter bias as well as the mean residual bias. In the GGMRF approach, not only the direct neighbors but a whole local neighborhood around every voxel supports the local fit. The size of the influencing neighborhood can continuously be increased using the spatial coupling factor  $\alpha$ .

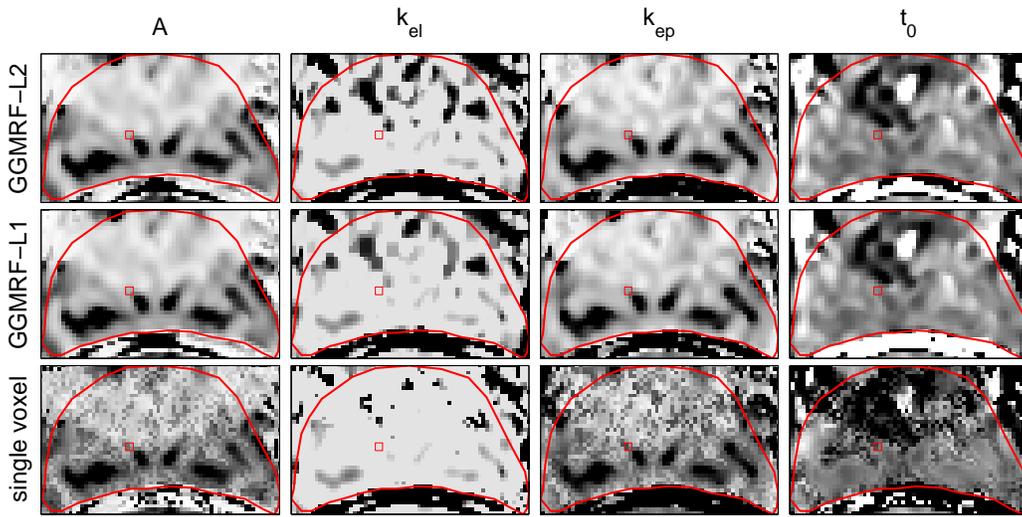


Fig. 11. Comparison of parameter maps from patient P1 obtained with the single voxel approach (last row) and using a GGMRF-L2/L1 prior (first/second row). Like for the simulations, results improve visibly when using the GGMRF priors.

*c) Multiple Optima:* Despite the convexity of the GGMRF priors, the problem of multiple (local) optima is not addressed in a principled way using the proposed approach. Even for very large spatial coupling factors the objective function Eqn. (15) is not guaranteed to be convex. This can be seen by noting that for  $\alpha \rightarrow \infty$  an equivalent problem is obtained by omitting the spatial terms in Eqn. (15) and introducing the constraint of all  $\theta_s$  being equal. The resulting sum of nonconvex functions does not yield a convex function in general. Empirically, however, the problem of multiple optima is alleviated when using the spatial prior. The GGMRF-L2 prior, for example, yields the same results with and without the constraint  $k_{el} > k_{ep}$  whereas the SV solution arbitrarily exchanges  $k_{ep}$  and  $k_{el}$  values which leads to notable speckles in the parameter maps [22]. This is also an example where the spatial prior can replace single-voxel prior knowledge. Furthermore, we noticed that the GGMRF-L1 prior is less effective in resolving such ambiguities since it is designed to allow for sudden spatial parameter changes, *i.e.* edges. Certainly, using a spatial prior together with the Block-ICM algorithm yields a different trajectory through parameter space as compared to optimizing the SV objective. Thus, running into a different optimum is not surprising.

### B. Spatial Smoothness

The basic assumption in using the GGMRF prior is the spatial smoothness of the true parameter maps. While this is certainly valid within homogeneous tissue regions, the assumption is violated at tissue borders where sudden changes may occur. An edge-preserving smoothness prior such as the GGMRF-L1 is designed to handle this by limiting the influence of voxels across edges [36]. Still, cases are conceivable in which an individual voxel is erroneously smoothed away with the GGMRF-L1 prior. Hence, using a spatial prior commonly leads to a loss in resolution on the one hand. On the other hand, signals with lower SNR can be processed which allows for acquiring MR data at higher physical resolutions. Clearly,

there is a trade-off to consider among MR imaging parameters and the strength of the spatial smoothness prior (*i.e.*  $\alpha$ ). A detailed experimental analysis of this trade-off, however, is beyond the scope of the present paper.

Based on the results presented above, we recommend using a spatial prior with small  $\alpha$  in any case. The obtained estimates are usually improved for the following four reasons:

First, since a weak spatial *prior* is applied it can only exert influence if the single-voxel evidence is weak. As an example, consider unenhancing tissue surrounding the prostate which yields a pure noise signal. For a small amplitude  $A$ , such a signal admits any value for  $k_{el}$ ,  $k_{ep}$  and  $t_0$  and a spatial prior would therefore not impose constraints on these parameters in the neighboring voxels. The other way round, if the single-voxel evidence for a particular parameter set is strong, *i.e.* its likelihood is much higher than that of any other, a weak spatial prior would hardly change this estimate.

Second, as compared to linear least squares, for which similar spatial priors have been studied [35], [36], nonlinear regression potentially profits much more. Due to the existence of local optima in the latter case, the parameter estimate  $\hat{\theta}(\alpha)$  is not necessarily a continuous function of  $\alpha$ . Therefore, only slight shifts in the likelihood of the parameter configurations, as induced by a weak spatial prior (small  $\alpha$ ), may yield very different solutions. This explains why the GGMRF prior helps remove speckles as shown in figures 10, 7 and 11.

Third, the simulation studies show that even parameter bias can be reduced for small coupling factors. This is true until the bias produced at edges outweighs the gain obtained in homogeneous regions. Interestingly, the MRB, which can be calculated without access to ground truth information, reascends at approximately the same value of  $\alpha$  as the parameter bias does.

Fourth, the reduction of the MRB for all patients in the prostate data set (cf. Table I) indicates that the same effects are obtained as in the simulation study.

The question remains as to which  $p$ -norm should be chosen for the GGMRF prior. On the simulated “wedding cake” data,

best results were obtained with  $p = 1.01$ . However, since the edge-preserving L1 prior is particularly suitable for data with sharp edges it is not clear whether the GGMRF-L1 prior is also best on natural data. For smooth ground truth images one would expect the GGMRF-L2 prior to perform better. On the patient data (Fig. 11, Fig. 10) as well as the simulated “real detail” data (Fig. 7) hardly any difference between the L1 and L2 norms is observed. However, considering that the GGMRF-L2 approach can be calculated more efficiently, preferences might be given to the L2 norm.

### C. Block-ICM

Despite its favorable properties, using the GGMRF prior certainly makes parameter estimation computationally more demanding. Without a specialized optimization strategy such as ICM which can exploit the inherent sparseness of the MAP problem, the GGMRF prior would not be applicable. Block-ICM can speed up convergence significantly and does not seem to be very sensitive to the choice of block size (as long as it does not reduce to conventional ICM).

Overall, the computational cost of using the GGMRF prior is only moderately higher than computing the SV solution as shown in Table I. This can be explained by two aspects. First, using a spatial prior changes the energy landscape in a way that allows the optimizer to converge more rapidly. The demonstrated reduction in standard deviation of the parameter estimates indicates that the Hessian at the optimum becomes more positive definite<sup>1</sup> which again has a positive influence on the speed of convergence. Second, using a small coupling factor  $\alpha$ , the effort of minimizing the data term dominates the spatial compensation.

For large  $\alpha$  we have observed much slower convergence and sometimes block-shaped artefacts in the estimated parameter maps. Thus, the Block-ICM approach is not recommended if the spatial smoothing term dominates. In that case, one should better resort to more complex, related methods based on domain decomposition and multigrid [38]. For small  $\alpha$  as we recommend for the estimation of kinetic parameter maps, however, the proposed Block-ICM algorithm provides a simple approach to tackle the high-dimensional optimization problem efficiently.

## VII. CONCLUSION

The application of spatial prior knowledge in the form of a generalized Gaussian Markov random field prior has been proposed to improve the estimation of kinetic parameter maps from DCE-MRI. The nonlinear regression problem that needs to be solved to determine pharmacokinetic parameters exhibits severe difficulties. Parameter bias, estimation variance and the existence of non-global optima have been exemplified using an instance of the generalized kinetic model. It was shown that using a GGMRF prior, either with L2 or L1 norm, can help reduce parameter bias as well as the variance and alleviate problems due to local optima. When using only a weak

<sup>1</sup>with respect to the usual partial ordering:  $A \succ B \Leftrightarrow A - B$  is positive definite.

spatial prior, improved parameter estimates can be obtained without blurring the resulting parameter maps. The proposed Block-ICM procedure provides means to tackle the resulting, very high-dimensional optimization problem efficiently. Using the GGMRF prior only resulted in a moderate increase in computation time. Future work should investigate ways of estimating the hyperparameters  $\alpha$ ,  $\mathbf{W}$ , and  $\sigma^2$  automatically from the examined DCE-MRI data. Also the question of which  $p$ -norm to use requires further consideration.

## APPENDIX

### AN ANALYTICAL GKM SOLUTION FOR THE FIRST-ORDER ELIMINATION MODEL

Using the Generalized Kinetic Model from Eqn. (1) and the arterial input function from Eqn. (2) an explicit solution to the system of ordinary linear differential equations is conveniently derived using Laplace transforms [42]. In the frequency domain, Eqns. (1) and (2) read

$$s \hat{C}_t(s) + k_{ep} \hat{C}_t(s) = K^{\text{trans}} \hat{C}_p(s) \quad (23)$$

$$s \hat{C}_p(s) + k_{el} \hat{C}_p(s) = \frac{K_{in}}{V_p} \left( \frac{1}{s} - \frac{e^{-\tau s}}{s} \right) \quad (24)$$

where  $s$  is the complex frequency and  $\hat{f}(s) = \mathcal{L}\{f(t)\}$  the Laplace transform of the function  $f(t)$ . Thus, the solution in the frequency domain is

$$\hat{C}_t(s) = K^{\text{trans}} \frac{K_{in}}{V_p} \frac{1 - e^{-\tau s}}{s(s + k_{ep})(s + k_{el})}. \quad (25)$$

Using the partial fraction expansion

$$\frac{1}{s(s + k_{ep})(s + k_{el})} = \frac{u}{s + k_{ep}} - \frac{v}{s + k_{el}} + \frac{v - u}{s} \quad (26)$$

where  $u^{-1} = k_{ep}(k_{ep} - k_{el})$  and  $v^{-1} = k_{el}(k_{ep} - k_{el})$ , and the time shift theorem of the Laplace transform [42]

$$\mathcal{L}^{-1} \left\{ e^{-\tau s} \hat{f}(s) \right\} = h(t - \tau) f(t - \tau), \quad (27)$$

the inverse transform of (25) and sought solution becomes

$$C_t(t) = K^{\text{trans}} \frac{K_{in}}{V_p} \begin{cases} C_t^0(t) & \text{for } 0 < t \leq \tau \\ C_t^\tau(t) & \text{for } t > \tau \end{cases} \quad (28)$$

where  $C_t^0(t) = u e^{-k_{ep}t} - v e^{-k_{el}t} + v - u$  and  $C_t^\tau(t) = u (e^{-k_{ep}t} - e^{-k_{ep}(t-\tau)}) - v (e^{-k_{el}t} - e^{-k_{el}(t-\tau)})$  which is equal to Eqn. (3).

## ACKNOWLEDGEMENT

The authors would like to thank Moritz Diehl and Johannes Schlöder for their helpful suggestions and discussions on nonlinear programming and high-dimensional optimization. Thanks to three anonymous reviewers whose constructive comments helped improve the manuscript significantly. Furthermore we would like to thank Peter Bachert and Wolfgang Schlegel from the German Cancer Research Center for their support.

## REFERENCES

- [1] F. Kiessling, M. Jugold, E. C. Woenne, and G. Brix, "Non-invasive assessment of vessel morphology and function in tumors by magnetic resonance imaging." *Eur Radiol*, vol. 17, no. 8, pp. 2136–2148, Aug 2007.
- [2] M. O. Leach, K. M. Brindle, J. L. Evelhoch, J. R. Griffiths, M. R. Horsman, A. Jackson, G. C. Jayson, I. R. Judson, M. V. Knopp, R. J. Maxwell, D. McIntyre, A. R. Padhani, P. Price, R. Rathbone, G. J. Rustin, P. S. Tofts, G. M. Tozer, W. Vennart, J. C. Waterton, S. R. Williams, and P. Workman, "The assessment of antiangiogenic and antivasular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations." *Br J Cancer*, vol. 92, no. 9, pp. 1599–1610, May 2005.
- [3] G. J. M. Parker and A. R. Padhani, *T<sub>1</sub>-w DCE-MRI: T<sub>1</sub>-weighted dynamic contrast-enhanced MRI*. Wiley, 2003, ch. 10, pp. 341–364.
- [4] C. Zechmann, K. Baudendistel, K. Aftab, L. Trojan, M.-S. Michel, H.-U. Kauczor, and S. Delorme, "Dynamic contrast-enhanced T1-weighted MRI combined with MR spectroscopic imaging in patients with prostate cancer - initial experience." in *Proc Intl Soc Mag Reson Med*, vol. 13, 2005.
- [5] F. Kiessling, M. Lichy, R. Grobholz, M. Heilmann, N. Farhan, M. S. Michel, L. Trojan, J. Ederle, U. Abel, H.-U. Kauczor, W. Semmler, and S. Delorme, "Simple models improve the discrimination of prostate cancers from the peripheral gland by T1-weighted dynamic MRI." *Eur Radiol*, vol. 14, no. 10, pp. 1793–801, Oct 2004.
- [6] H.-P. Schlemmer, J. Merkle, R. Grobholz, T. Jaeger, M. S. Michel, A. Werner, J. Rabe, and G. van Kaick, "Can pre-operative contrast-enhanced dynamic MR imaging for prostate cancer predict microvessel density in prostatectomy specimens?" *Eur Radiol*, vol. 14, no. 2, pp. 309–17, Feb 2004.
- [7] G. Brix, M. Henze, M. Knopp, R. Lucht, J. Doll, H. Junkermann, H. Hawighorst, and U. Haberkorn, "Comparison of pharmacokinetic MRI and [18F] fluorodeoxyglucose PET in the diagnosis of breast cancer: initial experience." *Eur Radiol*, vol. 11, no. 10, pp. 2058–70, 2001.
- [8] H. Hawighorst, M. Libicher, M. Knopp, T. Moehler, G. Kauffmann, and G. Kaick, "Evaluation of angiogenesis and perfusion of bone marrow lesions: role of semiquantitative and quantitative dynamic MRI." *J Magn Reson Imaging*, vol. 10, no. 3, pp. 286–94, Sep 1999.
- [9] H. J. Huisman, M. R. Engelbrecht, and J. O. Barentsz, "Accurate estimation of pharmacokinetic contrast-enhanced dynamic MRI parameters of the prostate." *J Magn Reson Imaging*, vol. 13, no. 4, pp. 607–614, Apr 2001.
- [10] T. Twellmann, O. Lichte, and T. W. Nattkemper, "An adaptive tissue characterization network for model-free visualization of dynamic contrast-enhanced magnetic resonance image data." *IEEE Trans. Med. Imag.*, vol. 24, no. 10, pp. 1256–66, Oct 2005.
- [11] T. W. Nattkemper and A. Wismüller, "Tumor feature visualization with unsupervised learning." *Med Image Anal*, vol. 9, no. 4, pp. 344–51, Aug 2005.
- [12] G. Brix, F. Kiessling, R. Lucht, S. Darai, K. Wasser, S. Delorme, and J. Griebel, "Microcirculation and microvasculature in breast tumors: pharmacokinetic analysis of dynamic MR image series." *Magn Reson Med*, vol. 52, no. 2, pp. 420–429, Aug 2004.
- [13] R. E. Port, M. V. Knopp, U. Hoffmann, S. Milker-Zabel, and G. Brix, "Multicompartment analysis of gadolinium chelate kinetics: blood-tissue exchange in mammary tumors as monitored by dynamic MR imaging." *J Magn Reson Imaging*, vol. 10, no. 3, pp. 233–241, Sep 1999.
- [14] P. Tofts, G. Brix, D. Buckley, J. Evelhoch, E. Henderson, M. Knopp, H. Larsson, T. Lee, N. Mayr, G. Parker, R. Port, J. Taylor, and R. Weiskoff, "Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusible tracer: standardized quantities and symbols." *J Magn Reson Imaging*, vol. 10, no. 3, pp. 223–32, Sep 1999.
- [15] P. S. Tofts, "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging." *J Magn Reson Imaging*, vol. 7, no. 1, pp. 91–101, 1997.
- [16] G. Brix, W. Semmler, R. Port, L. Schad, G. Layer, and W. Lorenz, "Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging." *J Comput Assist Tomogr*, vol. 15, no. 4, pp. 621–8, 1991.
- [17] P. S. Tofts and A. G. Kermode, "Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts." *Magn Reson Med*, vol. 17, no. 2, pp. 357–367, Feb 1991.
- [18] M. R. Orton, D. J. Collins, S. Walker-Samuel, J. A. d'Arcy, D. J. Hawkes, D. Atkinson, and M. O. Leach, "Bayesian estimation of pharmacokinetic parameters for DCE-MRI with a robust treatment of enhancement onset time." *Phys Med Biol*, vol. 52, no. 9, pp. 2393–2408, May 2007.
- [19] V. J. Schmid, B. Whitcher, A. R. Padhani, N. J. Taylor, and G.-Z. Yang, "Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging." *IEEE Trans. Med. Imag.*, vol. 25, no. 12, pp. 1627–1636, Dec 2006.
- [20] V. J. Schmid, B. J. Whitcher, G.-Z. Yang, N. J. Taylor, and A. R. Padhani, "Statistical analysis of pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging." *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, vol. 8, no. Pt 2, pp. 886–893, 2005.
- [21] T. S. Ahearn, R. T. Staff, T. W. Redpath, and S. I. K. Semple, "The use of the Levenberg-Marquardt curve-fitting algorithm in pharmacokinetic modelling of DCE-MRI data." *Phys Med Biol*, vol. 50, no. 9, pp. N85–N92, May 2005.
- [22] B. M. Kelm, N. Müller, B. H. Menze, and F. A. Hamprecht, "Bayesian estimation of smooth parameter maps for dynamic contrast-enhanced MR images with block-ICM," in *Proc. Computer Vision and Pattern Recognition Workshop (MMBIA)*. IEEE Computer Society, June 2006, pp. 96–103.
- [23] S. Walker-Samuel, M. O. Leach, and D. J. Collins, "Evaluation of response to treatment using DCE-MRI: the relationship between initial area under the gadolinium curve (IAUGC) and quantitative pharmacokinetic analysis." *Phys Med Biol*, vol. 51, no. 14, pp. 3593–3602, Jul 2006.
- [24] G. J. M. Parker, C. Roberts, A. Macdonald, G. A. Buonaccorsi, S. Cheung, D. L. Buckley, A. Jackson, Y. Watson, K. Davies, and G. C. Jayson, "Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI." *Magn Reson Med*, vol. 56, no. 5, pp. 993–1000, Nov 2006.
- [25] D. M. Bates and D. G. Watts, "Relative curvature measures of nonlinearity." *J Roy Statist Soc Ser B*, vol. 42, no. 1, pp. 1–25, 1980.
- [26] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. Wiley, 1989.
- [27] R. D. Cook, C.-L. Tsai, and B. C. Wei, "Bias in nonlinear regression." *Biometrika*, vol. 73, no. 3, pp. 615–623, 1986.
- [28] M. J. Box, "Bias in nonlinear estimation." *J Roy Statist Soc Ser B*, vol. 33, no. 2, pp. 171–201, 1971.
- [29] M. Clyde and K. Chaloner, "Constrained design strategies for improving normal approximations in nonlinear regression problems," *J Stat Planning and Inference*, vol. 104, pp. 175–196, 2002.
- [30] P. Hougaard, "The asymptotic distribution of nonlinear regression parameter estimates: Improving the approximation," *Int Stat Review*, vol. 56, no. 3, pp. 221–233, 1988.
- [31] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation." *IEEE Trans. Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.
- [32] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)." *J Roy Statist Soc Ser B*, vol. 36, pp. 192–236, 1974.
- [33] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, 2nd ed. Springer, 2003.
- [34] P. J. Huber, *Robust Statistics*. New York: John Wiley and Sons Inc., 1981.
- [35] P. Charbonnier, B.-F. Laure, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging." *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298–311, 1997.
- [36] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision." *Int J Comp Vision*, vol. 19, no. 1, pp. 57–92, 1996.
- [37] A. Toselli and O. Widlund, *Domain Decomposition Methods - Algorithms and Theory*, ser. Springer Series in Computational Mathematics. Springer, Hambrug, 2005, vol. 34.
- [38] U. Trottenberg, C. W. Oosterlee, and A. Schuller, *Multigrid*. Academic Press, 2000.
- [39] T. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds." *SIAM J on Optim*, vol. 6, pp. 418–445, 1996.
- [40] N. Balram and J. M. F. Moura, "Noncausal Gauss-Markov random fields: Parameter structure and estimation." *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1333–1355, July 1993.
- [41] J.-P. Chilès and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*. John Wiley, 1999.
- [42] O. Föllinger, *Laplace- und Fourier-Transformation*, 6th ed., ser. Studienreihe Elektrotechnik. Hüthig, 1993.