A comparison of Bayesian and non-linear regression methods for robust estimation of pharmacokinetics in DCE-MRI and how it affects cancer diagnosis

Nikolaos Dikaios[a)], David Atkinson [b)], Chiara Tudisca [b)], Pierpaolo Purpura[b)], Martin Forster[c,d)], Hashim Ahmed[e)], Timothy Beale[c)], Mark Emberton[e)], and Shonit Punwani[b,c)]

[a)] Centre for Vision, Speech and Signal Processing, University of Surrey, UK

[b)] Centre for Medical Imaging, University College London, UK

[c)] Department of Head and Neck Oncology, University College London Hospital, UK.

[d)] Cancer Institute, University College London, UK

[e)] Department of Urology, University College London, London, UK. NW1 2PG.

Author to whom correspondence should be addressed;

Dr Nikolaos Dikaios

Centre for Vision, Speech and Signal Processing

388 Stag Hill

Guildford

GU2 7XH

electronic mail: nd301@hotmail.co.uk.

ABSTRACT

The aim of this work is to compare Bayesian Inference methods with commonly used non-linear regression (NR) algorithms for estimating pharmacokinetics in Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI). The algorithms are compared in terms of accuracy, and reproducibility under different initialization settings. Further it is investigated how a more robust estimation of pharmacokinetics affects cancer diagnosis. The derived pharmacokinetics from the Bayesian inference algorithm were validated against NR algorithms (i.e. Levenberg-Marquardt, simplex) in terms of accuracy on a digital DCE phantom and in terms of goodness-of-fit (Kolmogorov-Smirnov test) on ROI-based concentration time courses from two different patient cohorts. The first cohort consisted of 76 men, 20 of whom had significant peripheral zone prostate cancer (any cancer-core-length (CCL) with Gleason>3+3 or any-grade with CCL>=4mm) following transperineal template prostate mapping biopsy. The second cohort consisted of 9 healthy volunteers and 24 patients with head and neck squamous cell carcinoma. The diagnostic ability of the derived pharmacokinetics was assessed with receiver operating characteristic area under curve (ROC AUC) analysis. The Bayesian inference algorithm accurately recovered the ground-truth pharmacokinetics for the digital DCE phantom consistently improving the Structural Similarity Index (SSIM) across the 50 different initializations compared to NR. For optimized initialization, Bayesian inference did not improve significantly the fitting accuracy on both patient cohorts, and it only significantly improved the v_e ROC AUC on the HN population from ROC AUC=0.56 for the simplex to ROC AUC=0.76. For both cohorts, the values and the diagnostic ability of pharmacokinetic parameters estimated with Bayesian Inference weren't affected by their initialization. To conclude, the Bayesian inference led to a more accurate and reproducible quantification of pharmacokinetic parameters in DCE-MRI, improving their ROC-AUC and decreasing their dependence on initialization settings.

Keywords: DCE analysis, Bayesian Inference algorithms, Prostate cancer, Head and Neck

Acknowledgements

This work was undertaken at UCLH/UCL, which receives funding from the Department of Health's NIHR Comprehensive Biomedical Research Centre funding scheme. This work was supported by UK EPSRC; Grant numbers: EP/I018700/1, EP/H046410/1 and the joint CR-UK & EPSRC funded King's College London and UCL Comprehensive Cancer Imaging Centre, in association with the MRC and DoH (England).

1. Introduction

Dynamic contrast enhanced (DCE) magnetic resonance imaging (MRI) is influenced by the micro-vascular characteristics of tissue, such as blood flow/volume, surface area/permeability of vessel walls, and micro-vascular density. These characteristics are associated with the expression of potent cytokines (such as the vascular endothelial growth factor) that support the development of tumor vessels. This makes DCE-MRI a valuable diagnostic tool in oncology. The purpose of this study is to investigate whether accurate quantification of pharmacokinetic parameters using the proposed Bayesian Inference algorithm can improve cancer diagnosis compared to non-linear regression fitting algorithms.

Quantification of pharmacokinetic parameters is affected by field inhomogeneities, gradients, SNR of the reconstructed images, and spatiotemporal resolution [1]. Besides limitations in acquisition, quantification of pharmacokinetic parameters will depend on the selection of pharmacokinetic model, the accurate estimation of the arterial input function, the estimation of the native T1 of the tissue [2] and the selection of fitting algorithm. Heyes et al [[3], [4]] studied the variation within- and between workstations in the derivation of pharmacokinetic parameters and reported a 25.1%–74.1% within-subject coefficient of variation. The conclusion of these studies is that unless the contrast agent material, the definition of AIF, the image SNR, and the fitting process are standardized DCE MRI related parameters will not be reproducible. Pharmacokinetic models such as the extended Toft model [5] that describe the enhancement process are often used to derive quantitative parameters and are increasingly used in diagnostic models [6] including computer aided diagnostic (CAD) software [[7], [8]]. Accurate quantification that will be reproducible between different clinical sites is necessary for the widespread of DCE based CAD software. This work will investigate how the optimization process itself can affect the quantification and the diagnostic ability of the quantified parameters..

Quantitative DCE parameters are usually extracted by fitting the estimated concentration to the measured concentration time course, using algorithms such as non-linear

least squares or the simplex algorithm. These fitting algorithms are prone to hit local minima [9] resulting in fitting errors and fitted parameters that depend on their initialization. To the best of our knowledge there are no guidelines on how to initialize the pharmacokinetics, and its clinical site uses its own initialization settings. Consequently there is a clear need to develop robust fitting strategies that will not be affected by the initialization of the pharmacokinetics.

To overcome these issues, Bayesian inference algorithms were suggested [[10], [11], [12]]. Bayesian inference algorithms can model the noise of the measured concentration of the contrast agent and have a theoretical guarantee to converge if run long enough [13]. This work suggests a Bayesian inference algorithm similar to the ones proposed by other groups [[10], [11], [12]] and evaluates its robustness and diagnostic value against the Levenberg–Marquardt and the simplex algorithms on two separate cohorts of patients:

i) a cohort of 76 men, 20 of whom had significant prostate cancer in the peripheral zone

ii) a cohort of 9 healthy volunteers and 24 patients with squamous cell carcinoma.

The proposed Bayesian inference algorithm is described in the theory section. The robustness value is assessed based on goodness-of-fit, and how robust the algorithm is when using different initialization settings of the estimated pharmacokinetic parameters. Receiver operating characteristic (ROC) analysis is performed on the derived pharmacokinetics to assess their ability to classify significant cancer.

2. Theory

2.1 Pharmacokinetic Modeling

A pharmacokinetic analysis was performed by fitting the extended Toft [5] (Eq. 1) modelled concentration $C(\mathrm{t})$ (mmol/L) to the concentration time course $C_{TIC}(\mathrm{t})$ (Eq. 2).

$$C(\mathrm{t}) = v_p \cdot C_a(\mathrm{t}) + K_{trans} \cdot \int_0^{\mathrm{t}} C_a(\tau - t_0) \otimes \mathrm{e}^{\left(-\frac{K_{trans}}{v_e}(t-\tau)\right)} \mathrm{d}\tau \qquad (1)$$

Where $C_a(\mathrm{t})$ is the arterial input function (mmol/L), $v_p$is the blood plasma volume fraction, $K_{trans}$is the transfer constant between plasma and interstitial space (min^{-1}),$v_e$is the interstitial space volume and $t_0$is the arrival time of the bolus at the tissue (secs). Population arterial input function was used [15].

The concentration time course was calculated from the image signal intensities $S(\mathrm{t})$ using the approximation Repetition time$\ll$T1

$$C_{TIC}(\mathrm{t}) = \frac{1}{r_1 \cdot T10}\left(\frac{S(\mathrm{t}) - S_0}{S_0}\right) \quad (2)$$

Where r_1 is the in-vivo relaxivity (4.51 L mmol^{-1} sec^{-1}), $T10$ is the native T1 of the tissue before contrast agent injection, calculated from a multiple flip angle dataset (sec), and S_0 is the average of the acquired images before the injection of the contrast agent.

2.2 Non linear regression algorithms

Pharmacokinetic models are fitted with two commonly used non-linear regression algorithms i.e. the Levenberg-Marquardt and the simplex algorithm. Levenberg-Marquardt is a least squares curve fitting algorithm that is a blend between the Gauss–Newton and the gradient descent method. The update rule of the pharmacokinetics parameters is: $k^{i+1}=k^i-(\mathrm{H}+\lambda\mathrm{I})^{-1}\nabla L(k^i)$, where H is the Hessian matrix at k^i, λ is a regularization parameter and L is the likelihood function to be minimized i.e. $L(k^i)=\sum_t(C_{TIC}(\mathrm{t}) - C(\mathrm{t}))^2$. When the likelihood is decreased λ is also reduced, but if the likelihood is increased λ will also be increased to reduce the influence of gradient descent. Contrary to other gradient based methods Levenberg-Marquardt is not performing a line minimization (where the direction of gradient descent is decided prior to step size estimation) hence requires less likelihood evaluations reducing the computational cost.

The simplex algorithm is also an iterative procedure but unlike the Levenberg-Marquardt does not require derivative information. The algorithm will create a “random” simplex of n+1 points, where n is the number dimensions (number of pharmacokinetic parameters to be

estimated). The simplex moves iteratively by reflection, expansion or contraction steps trying to find the pharmacokinetic parameters that minimize the likelihood function. In this work we used a constrained variation of the simplex algorithm [[16],[17]] and an $\ell 1$-norm in the likelihood function to improve robustness [18]. Simplex algorithm is particularly advantageous in cases where the gradient of the likelihood functions is hard to calculate.

2.3 Bayesian inference algorithm

In the proposed Bayesian Inference algorithm the measured concentration $C_{TIC}(\mathrm{t})$ is modelled using additive Gaussian noise σ and the pharmacokinetic parameters, k= {v_p, K_{trans}, v_e, t_0} for the extended Tofts model or k= { γ, K_{trans}, v_e, t_0} for the Orton model

$$C_{TIC}(\mathrm{t}) \sim \mathbf{normal}(C(\mathrm{t}), \sigma) \tag{3}$$

The suggested Bayesian inference algorithm similar to [[10], [11], [12]] maximizes the posterior probability distribution function $p(k,\sigma|C_{TIC})$ as a function of k and σ

$$\hat{k}, \hat{\sigma} = argmax_{k,\sigma}\, p(k, \sigma | C_{TIC}) \tag{4}$$

According to the Bayes theorem $p(k,\sigma|C_{TIC})$ is given by,

$$p(k, \sigma | C_{TIC}) = p(k, \sigma) \cdot p(C_{TIC} | k, \sigma) \int_{k^*, \sigma^*} p(k^*, \sigma^*) \cdot p(C_{TIC} | k^*, \sigma^*) \tag{5}$$

Where $p(C_{TIC}|k,\sigma)$ is the likelihood function of C_{TIC} given the pharmacokinetic parameters k,

$$p(C_{TIC} | k, \sigma) = (2\pi\sigma^2)^{-1} \exp\left(-\frac{1}{2\sigma^2} \| C_{TIC}(\boldsymbol{r}, t) - C(\boldsymbol{r}, t) \|_2^2\right) \tag{7}$$

and $p(k,\sigma)$ is the product of the prior probability distribution functions of k and σ, $p(\mathrm{k}, \sigma)$. Prior probability distribution functions reflect our prior knowledge about the k and σ parameters. We assume the subsequent prior distributions for every pharmacokinetic parameter

- v_p follows a Beta distribution, $v_p \sim \boldsymbol{Beta}(1,19)$ [19] reflecting an a priori expected value of 0.05.
- K_{trans} was parameterized as suggested by Schmid et al [10] as e^{θ} where θ follows a Gaussian distribution $\theta \sim \boldsymbol{Normal}(0,1)$
- v_e follows a Beta distribution, $v_e \sim$ ***Beta*** (2,1.5) reflecting an a priori expected value of 0.57
- t_0 follows a random distribution.

– σ^2 follows an uninformative Inverse Gamma distribution **IG**(10^{-4}, 10^{-4})

The integral $\int_{k^*,\sigma^*} p(k^*,\sigma^*)\cdot p(y|k^*,\sigma^*)$ is estimated with the Metropolis–Hastings algorithm.

3. Materials and methods

3.1 Generate Simulated DCE data

The DCE simulation used is similar to the one published from our group in Dikaios et al (2014) [20]. A normal volunteer underwent a fast gradient echo DCE-MRI protocol (flip angle α=10^o, repetition time TR=2.3 msecs). A T1-weighted abdominal image was acquired without contrast injection. The first time-frame was manually segmented into: liver, bowel, right and left heart, aorta, portal vein. Such segmentation was used as a map to simulate contrast enhancement using the extended Tofts model or the dual input function Orton model for the liver. Ground truth parametric maps i.e. native T10 (range 382-1932 msecs), v_p (range 0-1), v_e (range 0-1), and K_{trans} (range 0-1.38 min^{-1}) were used to simulate fifty DCE images with temporal resolution 3 secs using the spoiled gradient echo model.

$$S(t) = \rho \frac{\sin(\alpha)\cdot\left(1-\exp\left(-\frac{TR}{T1(t;k)}\right)\right)}{1-\cos(\alpha)\cdot\exp\left(-\frac{TR}{T1(t;k)}\right)} \qquad (8)$$

Where ρ is the proton density image, and was calculated analytically using Eq. 8 from the T1-weighted abdominal image without contrast injection and the graund truth T10 maps.
DCE images were transformed to (k, t)-space with fast Fourier transformation where noise was added. The noise of complex valued (k, t)-space MR data can be reasonably modelled by an additive white Gaussian distribution on both real and imaginary components (independent and identically distributed random variables). Simulated DCE data were generated for 2 different noise levels , one corresponding to the average SNR before contrast injection of prostate T1w images (SNR~9.2, noise level=2500) and a separate one corresponding to the average SNR before contrast injection of neck T1w images (SNR~15.1, noise level=800). The SNRs were calculated as described in Dikaios et al [21].

3.2 Patient populations

Institutional review board (IRB) approval for the study was obtained. The requirement for consent was waived for use of images acquired in routine clinical practice (prostate peripheral zone population) and obtained from all patients undergoing imaging as part of a separate clinical trial (head and neck population).

Prostate population

The prostate population consisted of men with clinically suspected prostate cancer (elevated prostate specific antigen (PSA) ± abnormal digital rectal examination ± family history of prostate cancer ± urinary symptoms,) undergoing prostatic multiparametric MRI (mp-MRI: T2 weighted, diffusion weighted and DCE imaging) prior to template-prostate-mapping (TPM) biopsies as part of standard of care at our institution. In total 76 men (mean age 63 years, range 45-79) with a mean prostate specific antigen of 7.8 ng/ml (range 1.2-20 ng/ml) and a mean prostate gland volume of 48.2 ml (range 23-137 ml) were included from 06/2007 to 03/2011. Twenty of the 76 men had histologically verified clinically significant peripheral zone prostate cancer.

Imaging was performed using a 1.5T magnet (Avanto, Siemens, Erlangen, Germany) with a pelvic phased array coil. The contrast media was Dotarem with an application dose 0.2 mL/Kgr. Prior to imaging, 0.2 mg/kg (maximum 20 mg) of spasmolytic (Buscopan; Boehringer Ingelheim, Ingelheim, Germany) was administered intravenously to reduce peristalsis. DCE-MRI was performed with a T1 weighted volumetric FLASH sequence with TR/TE 5.61/2.5 ms, flip angle 15^{o}, 384×384 matrix dimensions, field of view 269 mm, slice thickness 3 mm, 26 reconstructed slices, temporal resolution of 16 seconds, and number of time points 35.
For the purpose of this study and to match with the target performance of mp-MRI as defined by recent consensus [24]; histopathologists identified all locations with clinically significant cancer

based on volume assessment (0.2 ml) estimated by the cancer core length (CCL)>= 4 mm and/or the presence of Gleason pattern 4 disease [25]. Small volume (<0.2 ml) and low grade (<=Gleason 3+3) tumour was identified as clinically insignificant cancer.

An experienced radiologist (with 10 years of mp-MRI experience, reporting 500 mp-MRI prostate scans/year) and using the TPM biopsy histology as a guide, carefully matched the histopathology template to the mp-MRI; and contoured a region of interest (ROI) on early contrast enhanced T1 weighted images at the single largest histologically confirmed significant cancer site. For patients where the entire prostate was benign or contained only insignificant cancer, the radiologist contoured a 1-cm^2 ROI at a confirmed benign location within the PZ.

Head and Neck population

Twenty-four consecutive patients (mean age 60 years, standard deviation 9 years, range 44 to 80 years) satisfying inclusion criteria of histologically confirmed head and neck SCC with cervical nodal metastatic disease at pre-therapy staging, and 9 normal volunteers (mean age 48 years, standard deviation 16 years, range 20 to 75 years) were recruited between March 2010 and May 2012. All patients underwent contrast enhanced neck computed tomography (CECT), anatomical MRI and neck ultrasound as part of routine pre-treatment staging; and were consented for additional DCE MRI of the neck for research purposes.

All MRI studies were acquired using a 1.5T Siemens Avanto (Siemens, Erlangen, Germany) magnet with the manufacturer's carotid coils. The contrast media was Dotarem with an application dose 0.2 mL/Kgr. DCE-MRI was performed with a T1 weighted volumetric FLASH sequence with TR/TE 2.3/1.0 ms, flip angle 10^o, 256×256 matrix dimensions, field of view 269 mm, slice thickness 4 mm, temporal resolution of 3 seconds, and number of time points 50.

The reference standard was established by experienced head and neck radiologists (with 8 years and 24 years of head and neck experience respectively) through review of all CT and anatomical MRI, and performance of ultrasound evaluation of the neck in all patients. Cervical nodes were

assessed as per the Union for International Cancer Control (UICC): Tumor nodal-metastasis (TNM) classification of malignant tumours [26]. Equivocal nodes were sampled at the time of ultrasound by fine needle aspiration (FNA) and classified by in-room cytology.

3.3 Optimization details of the fitting algorithms

Fitting algorithms were implemented with in-house–developed software in MATLAB (The Mathworks Inc, Natick, MA). The pharmacokinetic parameters for the simplex, the Levenberg-Marquardt and the Bayesian inference algorithm were initialized as $v_p^0 = 0.05$, $K_{trans}^0 = 0.4\ min^{-1}$, $v_e^0 = 0.5$ for both the simplex and the Levenberg-Marquardt. The constraints of the pharmacokinetic parameters were: $v_p \in [0,1]$, $K_{trans} \in [0,2.7]\ min^{-1}$, $v_e \in [0,1]$, $t_0 \in [injection\ time, injection\ time + 40\ secs]$. Onset time was initialized with the time point the contrast agent was administered.

The simplex, the Levenberg-Marquardt and the Bayesian inference algorithms were run using multiple initialisations of the pharmacokinetic parameters. In addition to the aforementioned initialization, 49 different initialisations were also generated (50 initializations in total) using uniform distributions supported within intervals as described by the following formulas: $v_p^0 = unif\,(0, 0.2)$, $K_{trans}^0 = unif\,(0.3, 1.0)$, $v_e^0 = unif\,(0.3, 0.6)$.

For the proposed Bayesian inference algorithm the total number of iterations was 500, burn-in iterations were 300, thinning equal to 5, and tune iteration (number of iterations for tuning) was 67.

3.4 Statistical analysis

Statistical analysis was performed using SPSS (SPSS Base 20.0 for Windows. SPSS Inc., Chicago IL). The same statistical analysis was performed for both the head and neck and PZ prostate population.

A Mann–Whitney U test (MWU sig) was performed to compare the median values of

the pharmacokinetic parameters between normal and cancer ROIs. The goodness-of-fits of the simplex, Levenberg-Marquardt and the Bayesian inference algorithms were assessed with the Kolmogorov-Smirnov (KS) test statistic.

Separate univariate logistic regression models were built for the pharmacokinetic parameters derived using the simplex and the Bayesian inference algorithms. The ability of individual pharmacokinetic parameters to classify cancer was assessed by receiver operator characteristic (ROC) area under curve (AUC) analysis.

Leave-one-out analysis [21] was used for internal validation of predictive models. One case (out of the total patient population) was excluded, and a model generated from the remainder of the cases. The model was then tested on the excluded case and a predictive probability calculated. The process was repeated for all cases, excluding successive cases in turn allowing calculation of a predictive probability per case. An ROC (LOO ROC) was then created using the derived predictive probabilities. ROC curves were compared using the significance test suggested by Hanley and McNeil [27].

4. Results

4.1 Simulated DCE data

Table 1 demonstrates the similarity in terms of Structural SIMilarity (SSIM) index of the estimated pharmacokinetic maps estimated with the simplex, the Levenberg Marquardt and the Bayesian algorithm to the ground truth pharmacokinetic maps. Results are shown for two different noise realizations, one corresponding to the SNR of prostate T1w images (~9.2) before contrast injection and one corresponding to the SNR of neck T1w images (~15) before contrast injection. The pharmacokinetic maps estimated with the Bayesian algorithm have substantially higher SSIM and are less affected from the different initializations of the pharmacokinetic parameters (lower interquartile range across the 50 different initializations). The simplex algorithm has similar performance to the Levenberg-Marquardt (LM), with marginally higher

SSIM. Fig. 1 provides a visual comparison between the pharmacokinetics maps estimated with the simplex and the Bayesian algorithm.

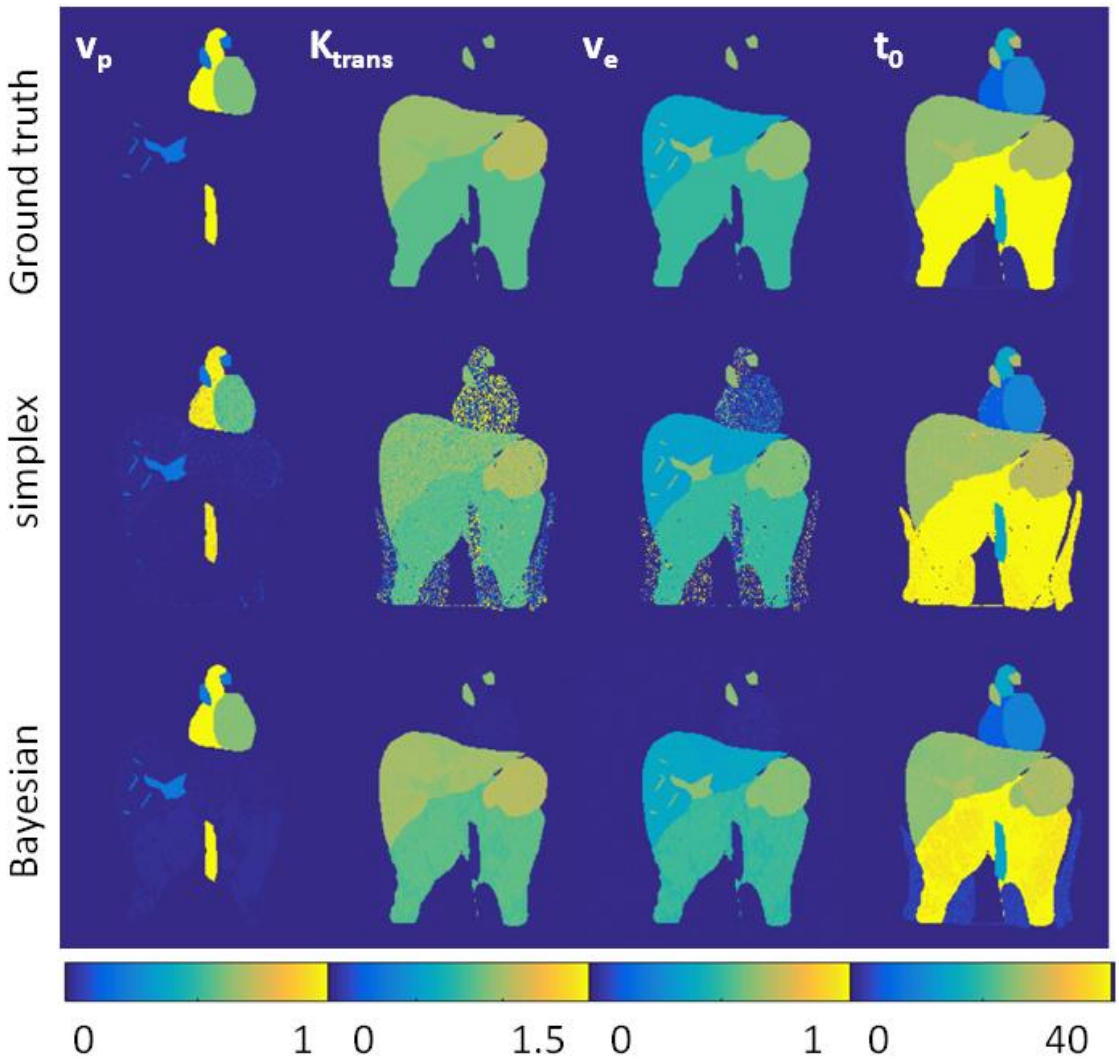

Fig. 1

Parametric maps (v_p, K_{trans}, v_e, t_0) estimated with pixel-by-pixel fitting of the simulated DCE images with SNR=9.2 using the simplex and the Bayesian inference algorithms. Ground truth pharmacokinetics maps are shown at the top row.

Table 1

SIMilarity (SSIM) index between the parametric maps (v_p, K_{trans}, v_e, t0) estimated with pixel-by-pixel fitting (using the simplex, the Levenberg Marquardt and the Bayesian inference algorithms) and the ground truth parametric maps of simulated DCE data. Results are shown for different

noise realizations with SNR=9.2 (prostate T1w images) and SNR=15 (neck T1w images). Median and Interquartile range (iQR) of SSIM were calculated across the 50 different initializations for each method.

Median (iQR)	v_p	K_{trans}	v_e	t_0
		SNR=9.2		
simplex	0.90 (0.17)	0.69 (0.13)	0.81 (0.14)	0.70 (0.12)
LM	0.89 (0.17)	0.69 (0.14)	0.80 (0.15)	0.68 (0.14)
Bayesian	0.92 (0.02)	0.81 (0.01)	0.84 (0.02)	0.76 (0.01)
		SNR=15		
simplex	0.95 (0.11)	0.81 (0.10)	0.87 (0.08)	0.79 (0.09)
LM	0.95 (0.11)	0.81 (0.10)	0.86 (0.08)	0.79 (0.09)
Bayesian	0.98 (<0.01)	0.91 (<0.01)	0.93 (0.01)	0.86 (0.01)

4.2 Prostate population

Multiple initialisations – Robustness of evaluated algorithms

The simplex, the Levenberg-Marquardt and the Bayesian inference algorithms were all run with the same 50 different initializations, Fig. 2 shows the KS test statistic (across the 76 mean ROI profiles of the PZ prostate population) for each initialization. The interquartile range of the medians was 0.019 for the simplex algorithm, 0.018 for the Levenberg-Marquardt algorithm and 0.002 for the Bayesian inference algorithm. Simplex algorithm had consistently better goodness-of-fit than the Levenberg-Marquardt; hence hereafter the Levenberg-Marquardt was excluded from the comparison.

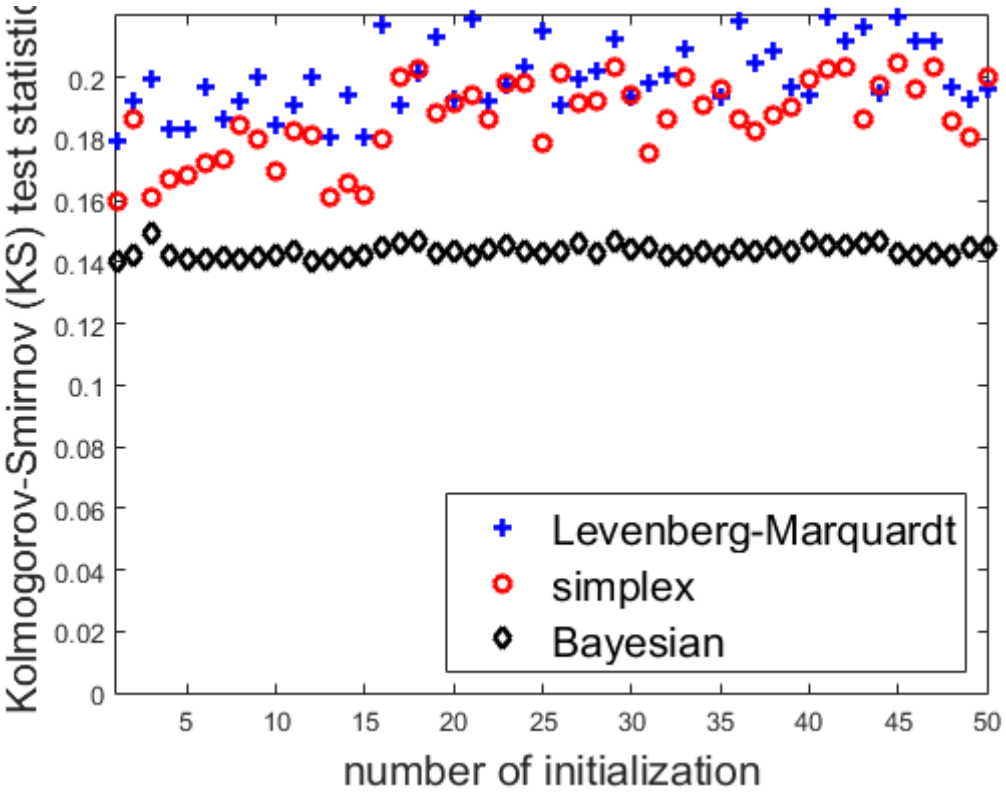

Fig. 2

Plot of the median KS statistic test (median KS statistic test across the 76 mean ROI PZ prostate profiles) across the 50 different initializations for the Levenberg-Marquardt, the simplex and the Bayesian inference algorithms. The interquartile range (iQR) of the median KS statistic test is 0.019 for the Levenberg-Marquardt, 0.018 for the simplex and 0.002 for the Bayesian inference algorithm.

Univariate ROC analysis

Table 2 shows the ROC analysis of the pharmacokinetic parameters estimated with the simplex and the Bayesian inference algorithms using the optimum pharmacokinetic initialization in terms of goodness-of-fit. K_{trans} was the best classifier of PZ prostate cancer for both the simplex and the Bayesian inference algorithm. According to the score test only K_{trans} estimated with the Bayesian inference algorithm could significantly discriminate PZ prostate cancer (p=0.02) (Table 2). However following a significance test between ROC curves, the AUC of K_{trans} estimated with the Bayesian inference algorithm (shown in Table 2) was not significantly better.

The simplex and the Bayesian inference algorithms were run with different initializations as described in section 3.3 and the ROC AUC were estimated per pharmacokinetic parameter for each initialization. The median (interquartile range) ROC AUC across the 50 different initializations were v_p:0.55 (0.05), K_{trans}:0.57 (0.14), and v_e:0.56 (0.05) for the simplex algorithm and v_p:0.63 (0.02), K_{trans}:0.67 (0.02), and v_e:0.56 (0.01) for the Bayesian algorithm. The median ROC AUC values for v_p and K_{trans} between the simplex and the Bayesian were significantly different.

Table 2

Score test and univariate ROC analysis of the pharmacokinetic parameters derived with the simplex and the Bayesian inference algorithms (using the optimum pharmacokinetic initialization in terms of goodness-of-fit) performed on the whole PZ population and following LOO analysis.

		score (p-value)	ROC AUC (CI)	LOO ROC AUC (CI)
simplex	v_p	0.21 (0.65)	0.61 (0.47-0.76)	0.22 (0.11-0.33)
	K_{trans}	3.22 (0.07)	0.64 (0.50-0.78)	0.57 (0.41-0.72)
	v_e	0.69 (0.41)	0.54 (0.40-0.68)	0.41 (0.28-0.54)
Bayesian	v_p	2.31 (0.13)	0.58 (0.43-0.74)	0.48 (0.32-0.65)
	K_{trans}	5.46 (0.02)	0.67 (0.54-0.81)	0.63 (0.50-0.77)
	v_e	0.75 (0.39)	0.56 (0.41-0.71)	0.44 (0.30-0.58)

Comparison of pharmacokinetic parameters between PZ prostate cancer/benign ROIs

Parametric maps of a PZ prostate cancer patient estimated with the simplex and the Bayesian inference algorithms are illustrated in Fig. 3. The modelled concentration C(t) is fitted to the mean concentration profile along the PZ prostate cancer ROI $C_{TIC}(t)$ (Fig. 3). In Fig. 3, while v_p values estimated from the cancer ROI profile are almost zero for the simplex algorithm, following pixel-by-pixel fitting the cancer area in the v_p seems to be slightly higher than zero. Pharmacokinetic parameters estimated by fitting mean ROI profiles will not necessarily correlate with pharmacokinetic parameters estimated by pixel-by-pixel fitting. Taking the mean

of an ROI and propagating it in time will generate a "smooth" profile, resulting in an approximated time-intensity curve. Ideally pixel-by-pixel fitting needs to be performed, but because it is more computationally demanding many clinical papers resort to mean ROI profile fitting.

Following MWU test, none of the pharmacokinetic parameters estimated with the Bayesian inference algorithm were significantly different from the ones estimated with the simplex algorithm for either the benign or the cancer ROIs (Fig. 4).

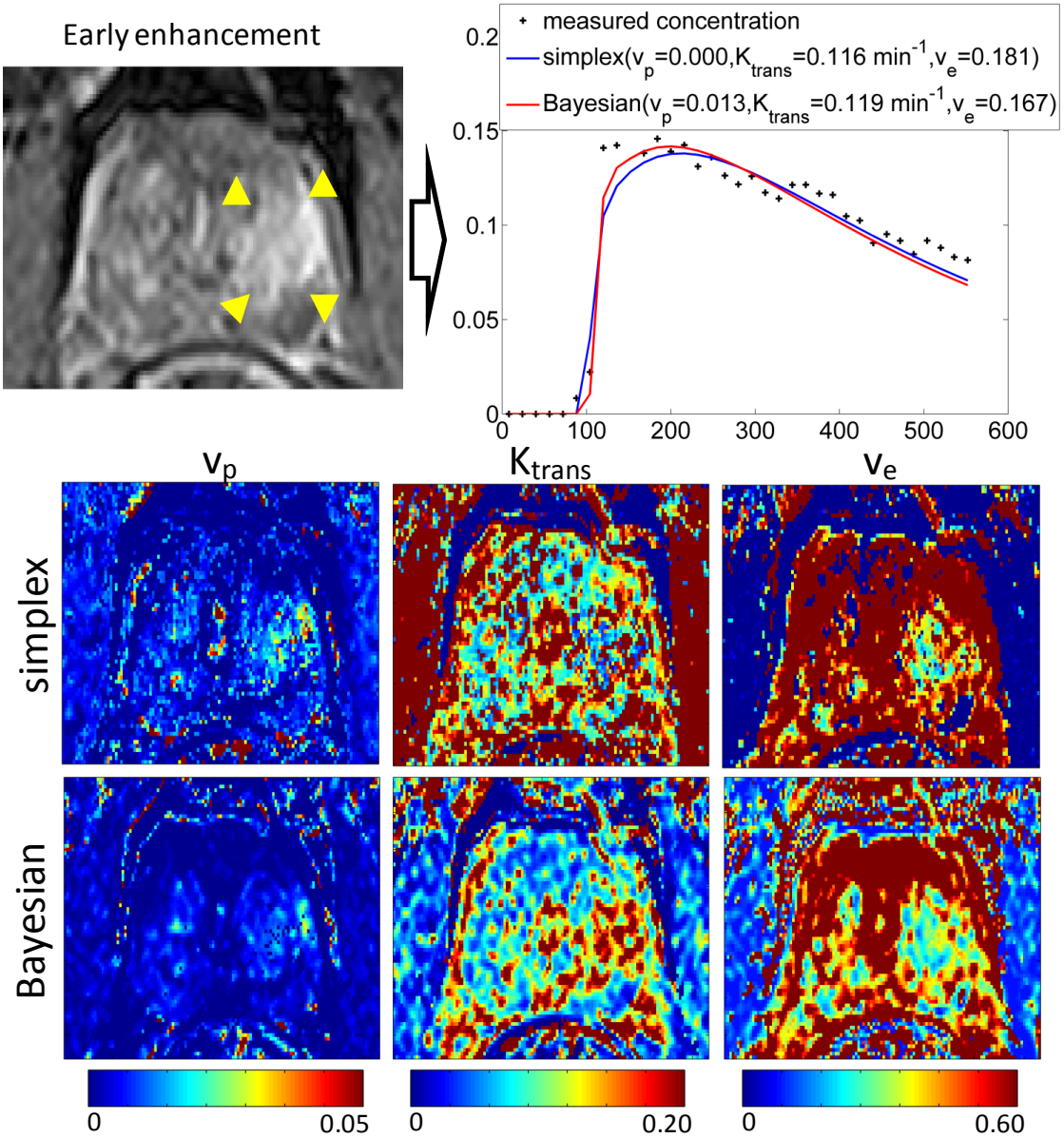

Fig. 3

Pharmacokinetic maps (v_p, K_{trans}, v_e) estimated with pixel-by-pixel fitting using the simplex and the Bayesian inference algorithms for a PZ prostate cancer patient. A plot of the mean ROI concentration profile $C_{TIC}(t)$ and the fitted to curve using the simplex and the Bayesian inference algorithms is also shown.

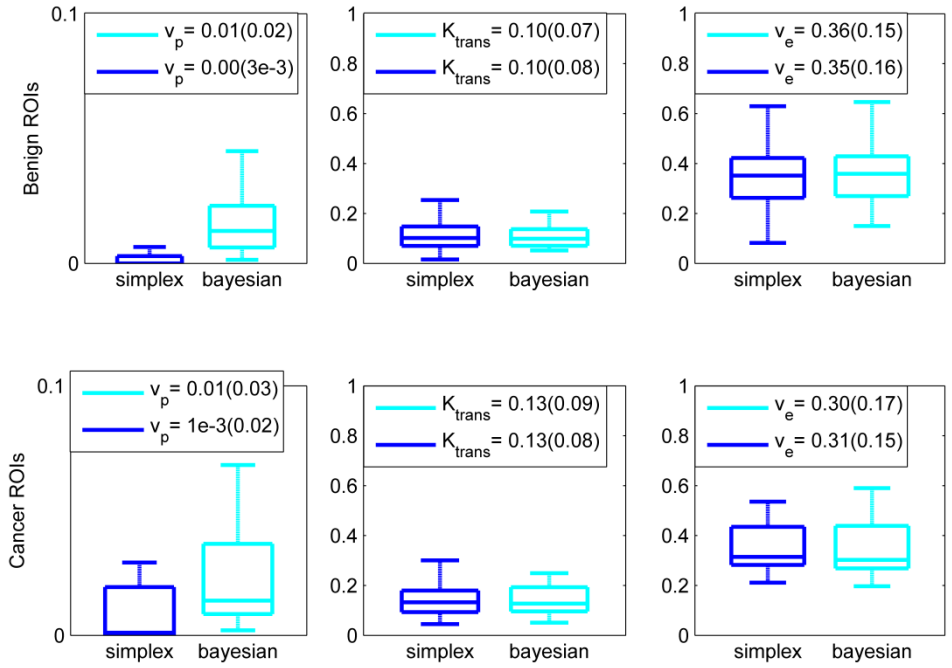

Fig. 4

Boxplot diagram of the pharmacokinetic parameters derived with the simplex and the proposed Bayesian inference algorithm, performed separately for the normal and cancer PZ prostate ROIs. The terms in brackets refer to the median value (interquartile range) of the estimated pharmacokinetic parameters.

4.3 Head and Neck population

Multiple initialisations – Robustness of evaluated algorithms

The simplex, the Levenberg-Marquardt and the Bayesian inference algorithms were all run with the same 50 different initializations, Fig. 5 shows the median KS statistic test (across the 33 mean ROI profiles of the head and neck patients and volunteers) for each initialization. The interquartile range of the medians was 0.0083 for the simplex algorithm, 0.010 for the Levenberg-

Marquardt algorithm and 0.0021 for the Bayesian inference algorithm. Simplex algorithm had consistently better goodness-of-fit than the Levenberg-Marquardt; hence from hereafter the Levenberg-Marquardt was excluded from the comparison.

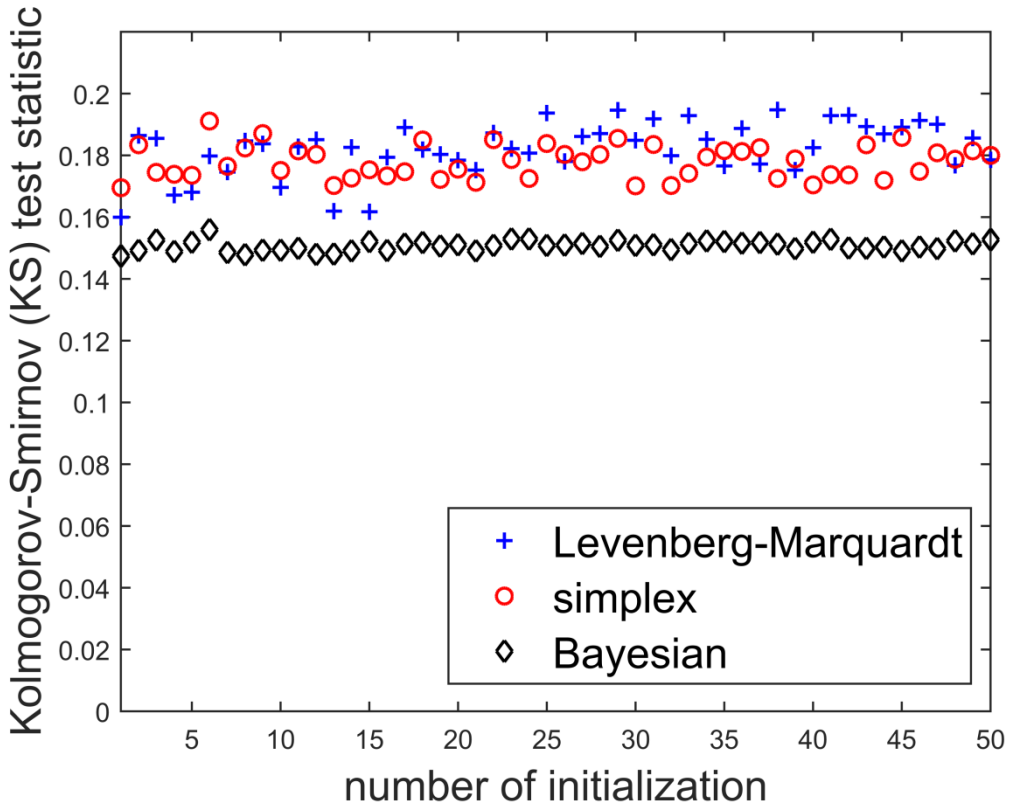

Fig. 5

Plot of the median KS statistic test (median KS statistic test across the 33 mean ROI head and neck profiles) across the 50 different initializations for the simplex and the Bayesian inference algorithms. The interquartile range (iQR) of the median KS statistic test is 0.010 for the Levenberg-Marquardt, 0.0083 for the simplex and 0.0021 for the Bayesian inference algorithm.

Univariate ROC analysis

Table 3 shows the ROC analysis of the pharmacokinetic parameters estimated with the simplex and the Bayesian inference algorithms using the optimum pharmacokinetic initialization in terms of goodness-of-fit. K_{trans} was the best classifier of head and neck metastatic patients for both the simplex and the Bayesian inference algorithms. According to the score test, for the

simplex algorithm only K_{trans} could significantly classify metastatic patients, whereas for the Bayesian inference both K_{trans} and v_e were significant classifiers (table 3).

Following a significance test between ROC curves, the AUC (on the original population or following LOO analysis) of K_{trans}, estimated with the Bayesian inference algorithm, was not significantly better. Significant difference was only found for the v_e AUC between the simplex and the Bayesian inference algorithms (table 3).

The simplex and the Bayesian inference algorithms were run with different initializations as described in section 3.3 and the ROC AUC were estimated per pharmacokinetic parameter for each initialization. ROC AUC were estimated per pharmacokinetic parameter for each initialization. The median (interquartile range) ROC AUC across the 50 different initializations were v_p: 0.54 (0.14), K_{trans}:0.76 (0.13), and v_e:0.56 (0.15) for the simplex algorithm and v_p:0.59 (0.03), K_{trans}:0.81 (0.01), and v_e:0.79 (0.02) for the Bayesian algorithm. The median ROC AUC values for v_e between the simplex and the Bayesian were significantly different.

Table 3

Score test and univariate ROC analysis of the pharmacokinetic parameters derived with the simplex and the Bayesian inference algorithms (using the optimum pharmacokinetic initialization in terms of goodness-of-fit) performed on the whole head and neck patient population and following LOO analysis. Asterisk (*) denotes the cases where the pharmacokinetic parameter estimated with the Bayesian inference algorithm is significantly different from the corresponding one derived with the simplex algorithm.

		score (p-value)	ROC AUC (CI)	LOO ROC AUC (CI)
simplex	v_p	0.12 (0.73)	0.56 (0.33-0.79)	0.30 (0.13-0.48)
	K_{trans}	5.43 (0.02)	0.74 (0.58-0.90)	0.66 (0.43-0.89)
	v_e^*	0.54 (0.49)	0.56 (0.34-0.77)	0.31 (0.15-0.50)
Bayesian	v_p	1.05 (0.31)	0.58 (0.35-0.80)	0.51 (0.31-0.72)
	K_{trans}	6.37 (0.01)	0.80 (0.64-0.94)	0.75 (0.57-0.92)
	v_e^*	4.76 (0.03)	0.76 (0.60-0.93)	0.70 (0.52-0.89)

Comparison of pharmacokinetic parameters between metastatic/benign ROIs

Parametric maps of a head and neck metastatic patient estimated with the simplex and the Bayesian inference algorithms are illustrated in Fig. 6. Fitting the estimated concentration C(t) to the mean ROI concentration profile along the head and neck metastatic nodes C_{TIC} (t) is also shown in Fig. 6.

Following MWU test all the pharmacokinetic parameters estimated with the Bayesian inference algorithm were significantly different from the ones estimated with the simplex algorithm for both the benign and the cancer ROIs (Fig. 7).

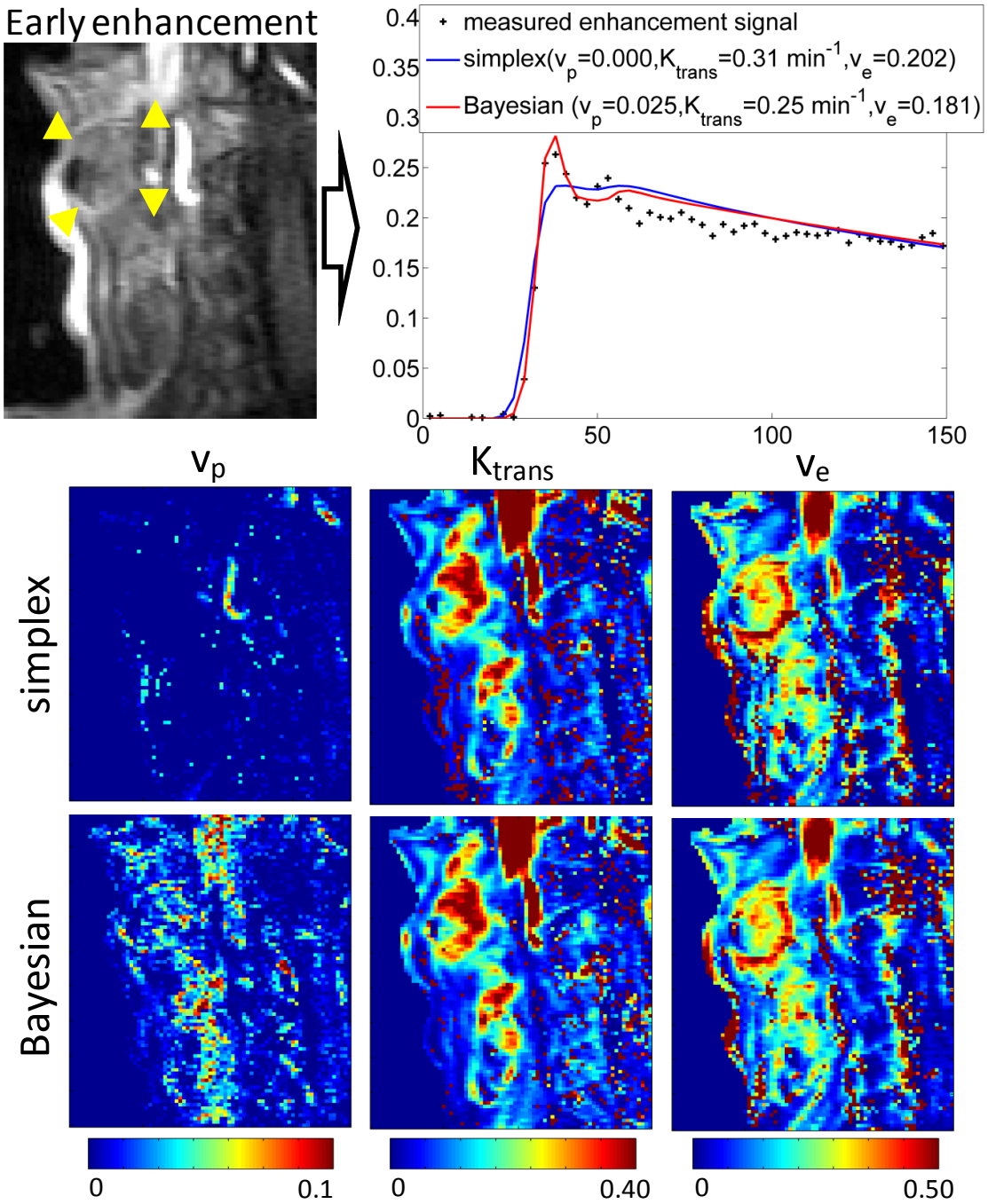

Fig. 6

Parametric pharmacokinetic maps (v_p, K_{trans}, v_e) estimated with pixel-by-pixel fitting using the simplex and the Bayesian inference algorithms for a head and neck patient with a metastasis. A plot of the mean ROI concentration profile $C_{TIC}(t)$ and the fitted curve using the simplex and the Bayesian inference algorithms is also shown.

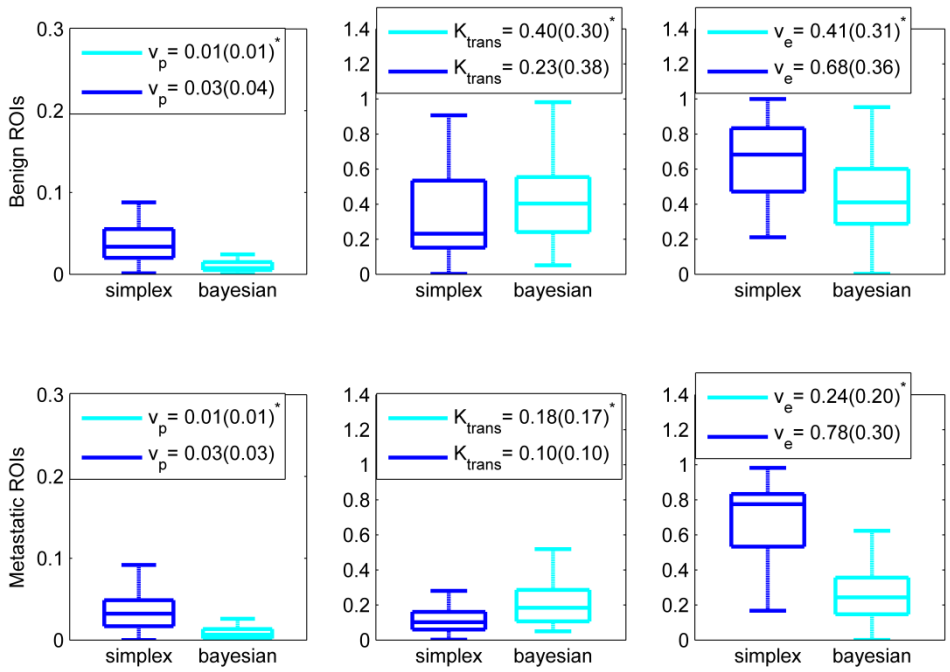

Fig. 7

Box-plot diagram of the pharmacokinetic parameters derived with the simplex and the proposed Bayesian inference algorithms, performed separately for the benign and metastatic neck node ROIs. The terms in brackets refer to the median value (interquartile range) of the estimated pharmacokinetic parameters. Asterisk (*) denotes significant difference ($p<0.05$) between the simplex and Bayesian inference algorithms.

5. Discussion

This works aims to investigate the diagnostic benefits of using Bayesian Inference algorithms for the derivation of pharmacokinetic parameters in DCE-MRI. The proposed Bayesian Inference algorithm is compared against non-linear regression algorithms (i.e. Levenberg-Marquardt and simplex) in terms of accuracy, reproducibility under different initialization settings and ability to classify cancer.

The simplex algorithm had consistently marginally higher SSIM with the ground truth kinetics of the simulated DCE phantom and better goodness-of-fit for the ROI-based TIC of both

populations than the Levenberg–Marquardt, which could be attributed to its convergence properties [17]. Unlike the Levenberg–Marquardt, the simplex algorithm does not use gradients, which provides some resilience to noise and local minima.

When running the proposed Bayesian inference algorithms for different initializations we found that

i. The SSIM with the ground truth pharmacokinetic maps for the Bayesian inference algorithm was consistently higher than for the non-linear regression algorithms for all initializations.

ii. The goodness-of-fit (KS statistic test) for the Bayesian inference algorithm was almost constant and consistently lower than the non-linear regression algorithms for all initializations and for both populations.

iii. The ROC AUC of the pharmacokinetic parameters estimated with the Bayesian inference algorithms have an interquartile range across the different initializations up to 0.03, whereas for the simplex algorithm the interquartile range is up to 0.14 (PZ prostate population) and 0.15 (head and neck population).

Pharmacokinetic parameters estimated with the proposed Bayesian inference algorithm had higher classification ability for both PZ prostate and head and neck cancer. Pharmacokinetic parameters estimated with the simplex algorithm that could not significantly classify disease, when estimated with the proposed Bayesian inference algorithm were significant classifiers of PZ prostate cancer (i.e. K_{trans}) and metastatic head and neck cancer (i.e. v_e). However the ROC AUC improvement achieved with the Bayesian inference algorithm was not significant for the PZ prostate cancer. For the head and neck metastasis only the ROC AUC improvement for v_e was significant.

Bayesian inference algorithms have been proposed before in the literature [[10], [11], [12]] to estimate unbiased quantitative pharmacokinetic parameters. The proposed scheme is similar to the one suggested by Schmid et al [10], the main difference is on the estimation of the

onset time. The accuracy of the estimated pharmacokinetic parameters will depend on the arrival time of the contrast agent to the tissue (onset time) [28]. Schmidt et al [7] calculated the onset time as the minimum time t*, for which the contrast concentration significantly exceeds zero minus $C(t^*)/\partial_t C(t^*)$. For the simulated DCE phantom with SNR=9.2, the SSIM index of the onset time calculated with the method of Schmidt et al [10] is 0.5754, whereas for the proposed Bayesian algorithm the respective SSIM is 0.76 (Table 1). This affected the estimation of the pharmacokinetic parameters, but if the same onset time was used the Bayesian method suggested by Schmid et al [10] has similar performance with the one proposed in this work. This is expected since both use the Metropolis–Hastings Markov chain Monte Carlo (MCMC) method and similar prior information. Their only difference is that we parameterized the posterior probability distribution function $p(k,\sigma|C_{TIC})$ with v_e to optimize EES volume directly instead of calculating it via k_{ep} ($v_e=K_{trans}/k_{ep}$) [7].

6. Limitations

For the PZ prostate population, we were reliant upon visual matching of the Barzell zone histology on TPM with the ROIs on the mp-MRI. Therefore, results may be influenced by mis-registration errors. Although no biopsy is free from sampling error [29] we used TPM to address as much of the systematic error inherent to transrectal ultrasound (TRUS) guided biopsy as possible [[29], [30]]. For the head and neck population there was a relatively small sample size. We took great care to be certain about positive and negative disease status within individual nodes by recruiting patients with N2/3 disease confirmed by CT, MRI and US ± FNA.

7. Conclusions

DCE MRI pharmacokinetic parameters are increasingly used in clinical practice; their diagnostic ability will depend on their accurate and reproducible quantification. The proposed Bayesian inference algorithm has been shown in this work to improve the diagnostic ability compared to

the simplex algorithm and was robust when different initializations of the pharmacokinetic parameters were used. These assets of the algorithm are essential to train and validate robust CAD software based on DCE-MRI that could be used between different sites. The performance of the Bayesian inference algorithm was consistent on two different populations, acquired with different settings.

References

[1] Yuan J, Chow SK, Yeung DK, Ahuja AT, King AD. Quantitative evaluation of dual-flip-angle T1 mapping on DCE-MRI kinetic parameter estimation in head and neck. Quant Imaging Med Surg. 2012; 2: 245-53.

[2] Heye T, Boll DT, Reiner CS, Bashir MR, Dale BM, Merkle EM. Impact of precontrast T10 relaxation times on dynamic contrast-enhanced MRI pharmacokinetic parameters: T10 mapping versus a fixed T10 reference value. J Magn Reson Imaging. 2014; 39(5):1136-45.

[3] Heye T, Davenport MS, Horvath JJ, Feuerlein S, Breault SR, Bashir MR, Merkle EM, Boll DT. Reproducibility of dynamic contrast-enhanced MR imaging. Part I. Perfusion characteristics in the female pelvis by using multiple computer-aided diagnosis perfusion analysis solutions. Radiology. 2013; 266(3):801-11.

[4] Heye T, Merkle EM, Reiner CS, Davenport MS, Horvath JJ, Feuerlein S, Breault SR, Gall P, Bashir MR, Dale BM, Kiraly AP, Boll DT. Reproducibility of dynamic contrast-enhanced MR imaging. Part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. Radiology. 2013; 266(3):812-21.

[5] Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. J Magn Reson Imaging 1997; 7(1):91-101.

[6] Dikaios N, Alkalbani J, Abd-Alazeez M, Sidhu HS, Kirkham A, Ahmed HU, Emberton M, Freeman A, Halligan S, Taylor S, Atkinson D, Punwani S. Zone-specific logistic regression models improve classification of prostate cancer on multi-parametric MRI. Eur Radiol. 2014; 25(9):2727-37.

[7] Sung YS, Kwon HJ, Park BW, Cho G, Lee CK, Cho KS, Kim JK. Prostate cancer detection on dynamic contrast-enhanced MRI: computer-aided diagnosis versus single perfusion parameter maps. AJR Am J Roentgenol. 2011; 197(5):1122-9.

[8] Kim MH, Kim JK, Lee Y, Park BW, Lee CK, Kim N, Cho G, Choi HJ, Cho KS. Diagnosis of lymph node metastasis in uterine cervical cancer: usefulness of computer-aided diagnosis with

comprehensive evaluation of MR images and clinical findings. Acta Radiol. 2011; 52:1175-83.
[9] McKinnon, K.I.M. Convergence of the Nelder–Mead simplex method to a non-stationary point. SIAM J Optimization 1999; 9: 148–158.
[10] Schmid V, Witcher B, Padhani A, Taylor N and Yang G-Z. Bayesian methods for pharmacokinetic models in dynamic constrast-enhanced magnetic resonance imaging. IEEE Trans. Med. Imaging 2006; 25 1627–36.
[11] Bhushan M, Schnabel JA, Risser L, Heinrich MP, Brady JM, Jenkinson M. Motion correction and parameter estimation in dceMRI sequences: application to colorectal cancer. Med Image Comput Comput Assist Interv 2011; 14:476-83.
[12] Orton MR, Collins DJ, Walker-Samuel S, d'Arcy JA, Hawkes DJ, Atkinson D, Leach MO. Bayesian estimation of pharmacokinetic parameters for DCE-MRI with a robust treatment of enhancement onset time. Phys Med Biol. 2007; 52(9):2393-408.
[13] Gelman A and Shirley K. Inference from simulations and monitoring convergence. In {\em Handbook of Markov Chain Monte Carlo}, ed.\ S. Brooks, A. Gelman, G. Jones, and X. L. Meng. CRC Press, 2011.
[14] Orton MR, Miyazaki K, Koh DM, Collins DJ, Hawkes DJ, Atkinson D, Leach MO. Optimizing functional parameter accuracy for breath-hold DCE-MRI of liver tumours. Phys Med Biol. 2009; 54: 2197-215.
[15] Parker GJ, Roberts C, Macdonald A, Buonaccorsi GA, Cheung S, Buckley DL, Jackson A, Watson Y, Davies K, Jayson GC. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. Magnetic Resonance in Medicine 2006; 56: 993–1000.
[16] Errico JD, 2005. Bound Constrained Optimization, Computer Program. MATLAB and Simulink Centre, The MathWorks, Inc..
[17] Lagarias JC, Reeds JA, Wright MH, and Wright PE. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM Journal of Optimization 1998; 9:112–147.
[18] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C. Cambridge, New York, 1994.
[19] Schmid VJ, Whitcher B, Padhani AR, Taylor NJ, Yang GZ. A Bayesian hierarchical model for the analysis of a longitudinal dynamic contrast-enhanced MRI oncology study. Magn Reson Med. 2008; 61(1):163-74.
[20] Dikaios N, Arridge S, Hamy V, Punwani S, Atkinson D. Direct parametric reconstruction from undersampled (k, t)-space data in dynamic contrast enhanced MRI. Med Image Anal 2014; 18(7): 989-1001.
[21] Dikaios N, Punwani S, Hamy V, Purpura P, Rice S, Forster M, Mendes R, Taylor S, Atkinson

D. Noise estimation from averaged diffusion weighted images: Can unbiased quantitative decay parameters assist cancer evaluation? Magn Reson Med. 2014; 71: 2105-17. [22] Onik G, Barzell W. Transperineal 3D mapping biopsy of the prostate: an essential tool in selecting patients for focal prostate cancer therapy. Urol Oncol 2008; 26:506-10.

[23] Taira AV, Merrick GS, Galbreath RW, Andreini H, Taubenslag W, Curtis R, Butler WM, Adamovich E, Wallner KE. Performance of transperineal template-guided mapping biopsy in detecting prostate cancer in the initial and repeat biopsy setting. Prostate Cancer Prostatic Dis. 2010; 13:71-7.

[24] Dickinson L, Ahmed HU, Allen C, Barentsz JO, Carey B, Futterer JJ, Heijmink SW, Hoskin PJ, Kirkham A, Padhani AR, Persad R, Puech P, Punwani S, Sohaib AS, Tombal B, Villers A, van der Meulen J, Emberton M. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. Eur Urol 2011; 59(4):477-94.

[25] Ahmed HU, Hu Y, Carter T, Arumainayagam N, Lecornet E, Freeman A, Hawkes D, Barratt DC, Emberton M. Characterizing clinically significant prostate cancer using template prostate mapping biopsy. J Urol. 2011; 186:458-64.

[26] Sobin LH. TNM classification of malignant tumours. 7 ed. Chichester, West Sussex, UK; Hoboken, NJ: John Wiley and Sons, 2009.

[27] Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology, 1982; 143, 29-36.

[28] Calamante F, Gadian D and Connelly A. Delay and dispersion effects in dynamic susceptibility contrast MRI: simulations using singular value decomposition. Magn. Reson. Med. 2000; 44: 466–73.

[29] Ahmed HU, Emberton M, Kepner G, Kepner J. A biomedical engineering approach to mitigate the errors of prostate biopsy. Nat Rev Urol. 2012; 9:227-31.

[30] Hoeks CM, Barentsz JO, Hambrock T, Yakar D, Somford DM, Heijmink SW, Scheenen TW, Vos PC, Huisman H, van Oort IM, Witjes JA, Heerschap A, Fütterer JJ. Prostate cancer: multiparametric MR imaging for detection, localization, and staging. Radiology 2011; 261:46-66.

[31] Roy C, Foudi F, Charton J, Jung M, Lang H, Saussine C, Jacqmin D. Comparative sensitivities of functional MRI sequences in detection of local recurrence of prostate carcinoma after radical prostatectomy or external-beam radiotherapy. AJR Am J Roentgenol. 2013; 200: 361-8.

[32] Turkbey B, Pinto PA, Mani H, Bernardo M, Pang Y, McKinney YL, Khurana K, Ravizzini GC, Albert PS, Merino MJ, Choyke PL. Prostate cancer: value of multiparametric MR imaging

at 3 T for detection--histopathologic correlation. Radiology. 2010; 255(1):89-99.