Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

Bayesian pharmacokinetic modeling of dynamic contrast-enhanced magnetic resonance imaging: validation and application

To cite this article: Andreas Mittermeier *et al* 2019 *Phys. Med. Biol.* **64** 18NT02

View the article online for updates and enhancements.

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

NOTE

Bayesian pharmacokinetic modeling of dynamic contrast-enhanced magnetic resonance imaging: validation and application

Andreas Mittermeier[1,3] [iD], Birgit Ertl-Wagner[1,2], Jens Ricke[1], Olaf Dietrich[1] [iD] and Michael Ingrisch[1] [iD]

[1] Department of Radiology, Ludwig-Maximilians-University Hospital Munich, Munich, Germany
[2] Department of Medical Imaging, The Hospital for Sick Children, University of Toronto, Toronto, Canada
[3] Author to whom any correspondence should be addressed.

E-mail: Andreas.Mittermeier@med.lmu.de

Abstract

Tracer-kinetic analysis of dynamic contrast-enhanced magnetic resonance imaging data is commonly performed with the well-known Tofts model and nonlinear least squares (NLLS) regression. This approach yields point estimates of model parameters, uncertainty of these estimates can be assessed e.g. by an additional bootstrapping analysis. Here, we present a Bayesian probabilistic modeling approach for tracer-kinetic analysis with a Tofts model, which yields posterior probability distributions of perfusion parameters and therefore promises a robust and information-enriched alternative based on a framework of probability distributions. In this manuscript, we use the quantitative imaging biomarkers alliance (QIBA) Tofts phantom to evaluate the Bayesian tofts model (BTM) against a bootstrapped NLLS approach. Furthermore, we demonstrate how Bayesian posterior probability distributions can be employed to assess treatment response in a breast cancer DCE-MRI dataset using Cohen's *d*. Accuracy and precision of the BTM posterior distributions were validated and found to be in good agreement with the NLLS approaches, and assessment of therapy response with respect to uncertainty in parameter estimates was found to be excellent. In conclusion, the Bayesian modeling approach provides an elegant means to determine uncertainty via posterior distributions within a single step and provides honest information about changes in parameter estimates.

1. Introduction

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a noninvasive imaging technique used to quantify microvascular tissue perfusion with the help of a contrast agent (CA) (Ingrisch and Sourbron 2013). In MRI, a gadolinium-based CA is used most commonly and injected intravenously after the acquisition of pre-contrast baseline scans. The CA increases T1 and T2 relaxation rates of surrounding water protons and causes signal enhancement in a T1-weighted acquisition. By measuring multiple T_1-weighted images during the passage of the CA through the tissue of interest, a time-dependent CA concentration can be extracted from the signal-time course of each voxel. Besides determining semi-quantitative and descriptive parameters from the concentration curves, e.g. time to peak, area under curve, or maximum, quantitative perfusion parameters can be obtained by fitting pharmacokinetic (PK) models to the data (Roberts *et al* 2006, Sourbron and Buckley 2012, 2013). Popular PK models that characterize CA transport from DCE-MRI data are the classical Tofts model (TM) (Tofts 1997), the extended Tofts model and the two compartment exchange model (Sourbron and Buckley 2011).

The standard approach for estimating PK parameters from DCE-MRI data is using non-linear regression to determine a maximum likelihood estimator by non-linear least squares (NLLS) analysis (Seber and Wild 2003). For this purpose, an optimizing algorithm aims to minimize the sum of squared residuals between model and data and yields, if successful, a *point estimate* of model parameters. The NLLS approach is widely used, and a number of software packages provide non-linear regression implementation of a range of PK models (Huang *et al* 2014a, Beuzit *et al* 2016). Bayesian probabilistic modeling, on the other hand, offers an alternative modeling approach within a framework of probability distributions. Briefly, a *prior* belief about model parameters is formulated as a probability distribution; this allows to incorporate domain expertise, e.g. physical constraints. With

dedicated algorithms, this prior belief is then updated with the measured data and yields the *posterior* probability distributions of the parameters given the data (McElreath 2015). Through recent algorithmic developments (Hoffman and Gelman 2011) and the increasing availability of computational power, the use of Bayesian modeling approaches is spreading in various disciplines and has already shown to be a robust and accurate alternative for the analysis of MR imaging data (Schmid *et al* 2006, Orton *et al* 2007, Woolrich *et al* 2009, Dikaios *et al* 2017, Tietze *et al* 2018, Hansen *et al* 2019). The posterior probability distributions that result from Bayesian modeling greatly increase the interpretability of analysis results. Compared to simple point estimates, entire parameter probability distributions allow a straightforward assessment of, e.g. whether a parameter has truly changed in the course of a therapy, or whether the parameter change has only occurred within the uncertainty of the estimation (Shukla-Dave *et al* 2018).

In the present manuscript, we investigated Bayesian tracer-kinetic modeling in the context of DCE-MRI. To this end, we implemented a Bayesian TM (BTM) with the purpose to (i) evaluate accuracy against a NLLS approach using a digital reference object, (ii) validate uncertainty estimates against a bootstrapped NLLS approach to assess the precision and (iii) demonstrate how Bayesian posterior probability distributions can be used to assess treatment response in a breast cancer DCE-MRI dataset.

2. Materials and methods

2.1. Signal conversion and pharmacokinetic models

In a typical DCE-MRI experiment, time-resolved signal intensity curves $S(t)$ are extracted voxel-wise from multiple T_1-weighted images. To derive quantitative information, the measured signal intensities need to be converted to CA concentration curves. For this purpose, the signal equation for the spoiled gradient echo (SPGR) sequence in steady state can be used with the baseline signal $S_0(t)$, flip angle α, repetition time T_R and relaxation rate $R_1(t)$ as:

$$S(t) = S_0 \sin(\alpha) \frac{1 - e^{-T_R R_1(t)}}{1 - \cos(\alpha) e^{-T_R R_1(t)}}. \tag{1}$$

One can solve equation (1) for the time-dependent relaxation rate $R_1(t)$:

$$R_1(t) = -\frac{1}{T_R} \log\left(\frac{1 - A}{1 - \cos(\alpha)A}\right), \tag{2}$$

with the auxiliary variable

$$A = \frac{S(t)}{S_0} \frac{1 - e^{-T_R R_{10}}}{1 - \cos(\alpha) e^{-T_R R_{10}}}. \tag{3}$$

A time-dependent concentration can then be calculated from the linear relation to the change in relaxation rates during and before administration of CA, $R_1(t)$ and R_{10}, respectively:

$$c(t) = (R_1(t) - R_{10})/r_1, \tag{4}$$

with the specific *relaxivity* of the gadolinium-based CA r_1 (Pintaske *et al* 2006).

A standard approach for the analysis of concentration-time curves in DCE-MRI data is the TM (Tofts and Kermode 1991, Tofts 1997, Sourbron and Buckley 2011) which assumes a negligible amount of intravascular tracer and describes CA transportation as:

$$c_t(t) = K^{\mathrm{trans}} e^{-t k_{\mathrm{ep}}} * c_p(t). \tag{5}$$

Here, $c_t(t)$ is the time-dependent concentrations of CA in the tissue of interest; $c_p(t)$ is the concentration in the blood plasma of the tissue-feeding artery, often referred to as arterial input function (AIF). $c_t(t)$ and $c_p(t)$ are connected with a convolution, expressed as '$*$'. The parameter v_e is the volume fraction of the interstitium, the extravascular extracellular space (EES). K^{trans} is defined as the transfer constant of CA between blood plasma and EES. The rate constant $k_{\mathrm{ep}} = K^{\mathrm{trans}}/v_e$ is the ratio of the transfer constant to the EES (Tofts *et al* 1999, Sourbron and Buckley 2011).

The tissue concentration $c_t(t)$ can be calculated from the measured signal $S(t)$ with equations (1)–(4) using the relaxation time T_{10} in tissue via $R_{10} = 1/T_{10}$. Plasma concentration $c_p(t)$ in the AIF can be calculated likewise using the relaxation time T_{10} of blood and the additional transformation from blood to plasma concentration via the hematocrit *hct*:

$$c_p(t) = c_b(t) \cdot (1 - hct). \tag{6}$$

Table 1. Parameters for the conversion from signal to concentration.

	T_{10} (Tissue)	T_{10} (Blood) (ms)	r_1 (Lmmol^{-1} ms^{-1})	α (°)	T_R (ms)	T_E	*hct*
QIBA DRO[a]	1000 ms	1440	0.0045	30	5	—	0.45
QIN breast[b]	1666 ms[c]	1440	0.0045	10	6.2	2.9 ms	0.45

[a] Quantitative imaging biomarker alliance digital reference object *QIBA_v6_Tofts*.
[b] Quantitative imaging network breast cancer dataset.
[c] Personal communication with the author of Huang *et al* (2014a).

The standard TM is used in the following within a classical NLLS likelihood framework and a Bayesian framework to quantify perfusion in simulated and measured DCE-MRI data. To account for noise in any observed data, an error term is added to the TM from equation (5) and an observation i is given as

$$y_i = c_t(t_i, \theta) + \epsilon_i \tag{7}$$

where the model parameters K^{trans} and v_e are summarized in the vector θ.

2.2. Data

2.2.1. Validation: QIBA DCE-MRI phantom

To evaluate accuracy of estimates and compare results of different fitting approaches, a simulated phantom with known PK parameters was investigated first. The quantitative imaging biomarkers alliance (QIBA)[3] provides several freely available test images for DCE-MRI analysis, known as digital reference objects (DRO). These have been used previously to validate various fitting algorithms and analysis toolkits (Ortuño *et al* 2013, Smith *et al* 2015, Debus *et al* 2019). The noise-free *QIBA_v6_Tofts* version was chosen here. The DRO contains simulated DCE-MRI data generated with the standard TM in equation (5) for a study duration of $t = 660$ s with a temporal resolution $\Delta t = 0.5$ s. Tissue concentration-time curves $c_t(t)$ have been created for all combinations of $K^{\text{trans}} \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.35\}$ min^{-1} and $v_e \in \{0.01, 0.05, 0.1, 0.2, 0.5\}$, filling a 10×10 pixel patch for each combination. Table 1 lists the parameters stated in the QIBA description[4], following QIBA's DCE MRI quantification profile[5] to convert signal intensities to concentrations (compare equations (1)–(4)). For a more realistic setting, complex Gaussian noise with standard deviation $\sigma = 0.2$ relative to the pre-contrast baseline signal S_0 was added to the original noise-free test data. No noise was added to the AIF for simplicity and to be able to reliably relate our results to published work of Smith *et al* (2015) and Ortuño *et al* (2013). Figure 1 shows a snapshot of the DRO signal intensities at $t = 100s$, the AIF and an exemplary voxel with parameters $K^{\text{trans}} = 0.2$ min^{-1} and $v_e = 0.2$, respectively.

2.2.2. Application: breast cancer DCE-MRI data

The quantitative imaging network (QIN) aims at improving quantitative imaging and does so by sharing data which was acquired as part of various QIN studies, collected in the cancer imaging archive (TCIA)[6] (Clark *et al* 2013). A set of breast cancer DCE-MRI data (Huang *et al* 2014b) in DICOM format acquired from 10 patients was used to demonstrate the performance of the BTM on clinical data. The dataset contains DCE-MRI measurements acquired before (visit 1) and during (visit 2) preoperative neoadjuvant chemotherapy (NACT), respectively. For three patients, pathologic complete response (pCR) was reported, the remaining seven patients were classified as non-pCR. In addition, the dataset includes a region of interest (ROI) per patient, drawn by an experienced breast radiologist. A sample-averaged AIF is provided as blood concentration $c_b(t)$ and was converted to plasma concentration $c_p(t)$ using equation (6). Signal intensities within the ROI were converted to tissue concentrations using equations (1)–(4). Parameters for the conversion are specified in table 1, further details can be found in the original work by Huang *et al* (2014a).

2.3. Models and analysis

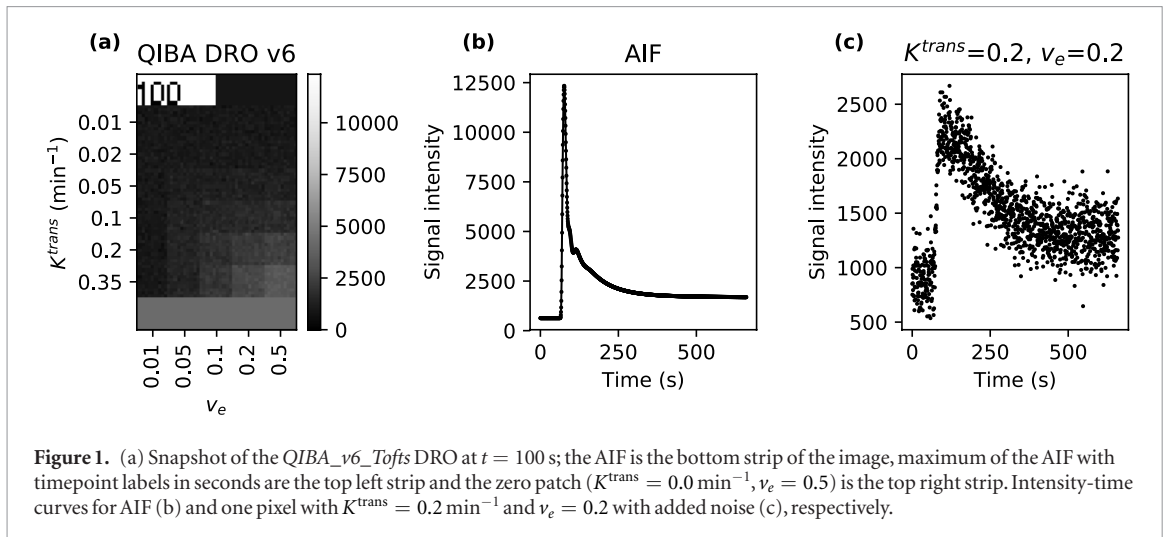#### 2.3.1. Non-linear least squares approach with bootstrapping

The standard evaluation of DCE-MRI data is performed in a likelihood framework by fitting a non-linear regression model to the concentration-time curve in every voxel. The NLLS approach minimizes the sum of squared errors between measured data y_i at timepoint t_i for $i = 0, 1, ..., N$ and the model function $c_t(t_i)$ in equation (7)

Figure 1. (a) Snapshot of the *QIBA_v6_Tofts* DRO at $t = 100$ s; the AIF is the bottom strip of the image, maximum of the AIF with timepoint labels in seconds are the top left strip and the zero patch ($K^{\text{trans}} = 0.0\,\text{min}^{-1}$, $v_e = 0.5$) is the top right strip. Intensity-time curves for AIF (b) and one pixel with $K^{\text{trans}} = 0.2\,\text{min}^{-1}$ and $v_e = 0.2$ with added noise (c), respectively.

$$\min \sum_{i=0}^{N} (y_i - c_t(t_i, \theta))^2 = \min \sum_{i=0}^{N} \epsilon_i^2 \qquad (8)$$

to infer the *best guess* parameter $\hat{\theta}$. Assuming normally distributed noise ϵ_i, the least-squares estimator $\hat{\theta}$ equals the maximum-likelihood estimator (Seber and Wild 2003).

An implementation of the Broyden–Fletcher–Goldberg–Shano (L-BFGS) algorithm (Byrd *et al* 1994, Zhu *et al* 1997) in *SciPy*[7] (Jones *et al* 2001) was used for inference of the parameters via the *optimize.minimize* function. Initial values for K^{trans} and v_e were set to 0.001; constraints for K^{trans} and v_e were set to positivity and [0,1], respectively. The concentration curves of the DRO were then fitted and parameter maps were constructed for K^{trans} and v_e. By comparing them to the true parameter maps, percentage error maps were calculated as $\theta_{\%\text{err}} = (\hat{\theta} - \theta_{\text{true}})/\theta_{\text{true}}$.

A bootstrap method was implemented to assess the uncertainty of $\hat{\theta}$ (Kershaw and Buckley 2006). For that, the residuals, i.e. the difference between the fitted and the measured curve were calculated. In a next step, the residuals were resampled by randomly drawing samples with replacement. Subsequently, the resampled residuals were added to the fitted curve and the TM was used to determine another set of estimates, equivalent to inferring the original best guess. The number of iterations was set to 1000.

Uncertainty maps were then calculated from the bootstrap samples for the NLLS approach. Denoted as σ, half the width between 17th and 83rd percentile was considered a more robust measure for the precision than the standard deviation and is used throughout this work. For samples following a Gaussian normal distribution, σ would be equal to the standard deviation.

2.3.2. Bayesian inference and implementation

The alternative evaluation is performed in a Bayesian framework which infers a full posterior distribution $P(\theta \mid y)$ of the model parameters θ given an observation of data y. The observational error ϵ_i for each measurement y_i at timepoint t_i for $i = 0, 1, ..., N$ in equation (7) is assumed to be Gaussian with standard deviation σ. Hence, the joint observations of CA concentration in each voxel, conditional on the parameters, are modeled in the *likelihood* as

$$P(y \mid \theta) = \prod_{i=0}^{N} \mathcal{N}(y_i \mid c_t(t_i, \theta), \sigma^2), \qquad (9)$$

with $\mathcal{N}$ representing a normal distribution and $c_t(t_i, \theta)$ the CA tissue concentration evaluated with the TM in equation (5).

Information about the parameters prior to the observation of data are specified in the *prior distribution* $P(\theta)$, enforcing physical or biological constraints. The likelihood of the data $P(y \mid \theta)$ and the product of the prior probability densities $P(\theta)$ are combined with the observed data to infer the joint posterior distribution via Bayes' theorem:

$$P(\theta \mid y) = \frac{P(y \mid \theta)\,P(\theta)}{P(y)}. \qquad (10)$$

The denominator in equation (10) is referred to as model evidence and calculates as $P(y) = \int P(\theta)P(y \mid \theta)\mathrm{d}\theta$. If the complexity of the model allows no analytical solution to this integral, Markov Chain Monte Carlo (MCMC)

---

[7] Python 3.6.6, scipy 1.1.0, www.scipy.org/

methods (Gilks *et al* 1995) offer a means to determine the posterior probability distribution. Briefly, a MCMC algorithm draws samples from a target distribution, which equals the desired posterior distribution. The accepted parameter proposals are stored in a chain or trace of estimates (Kruschke 2014).

The BTM was implemented in *Stan* (Carpenter *et al* 2017), an open-source software package, using *pystan*[8]. In the present analysis, weakly informative priors were chosen which are applicable to a wide range of clinical DCE-MRI data without restrictions. Appendix provides a prior predictive check on these distributions, showing their weakly informative nature by comparing generated concentration curves with real observations. In particular, for the volume fraction $v_e \in [0, 1]$ a beta prior $v_e \sim \text{Beta}(\alpha = 2, \beta = 2)$ was chosen and for $K^{\text{trans}} \in \mathbb{R}_+$ a gamma prior was specified $K^{\text{trans}}(\text{min}^{-1}) \sim \text{Gamma}(\alpha = 1.1, \beta = 1/0.002)$. The prior for the standard deviation of the observational error was set to $\sigma(\text{mmol/L}) \sim \text{LogNormal}(\mu = 0, \sigma = 1)$. MCMC samples were drawn from the posterior distribution with the No-U-Turn (NUTS) algorithm (Hoffman and Gelman 2011). The number of iterations was set to 1000, sampled in two chains simultaneously, following a warm-up period of 500 iterations. Stan also reports divergences of the sampling algorithm and indicates the need to update the default settings of NUTS, e.g. initial step size and target acceptance rate.

To monitor the convergence of the MCMC chains to the target distribution, different diagnostics are automatically run alongside in Stan. The potential scale reduction statistic, $\hat{R}$, by Gelman and Rubin (1992) compares the sample variance within and across chains, and indicates if chains have not converged to a common distribution ($\hat{R} > 1.1$). The effective sample size N_{eff} indicates the degree of uncertainty in estimates due to autocorrelation of samples (Geyer 2011).

All concentration curves of the DRO were then fitted with the BTM to obtain posterior probability distributions of the parameters θ. To be able to compare the distributions to point estimates and to generate parameter maps, two hallmarks of the posterior distributions were determined: the median and, as for the bootstrap samples, half the distance between the 17th and 83rd percentile, denoted as σ. By comparing the median parameter maps to the true parameter maps, a map of the percentage error was calculated as above to assess the accuracy of estimates.

To evaluate the breast cancer DCE-MRI datasets, the mean tissue concentration curve over the ROI $c_{t,\text{ROI}}(t)$ was calculated for each patient and both visits. Subsequently, all concentration-time curves were fitted with the BTM to infer posterior distributions for the model parameters θ. To ensure that the model adequately captured the underlying data generating process, a *posterior predictive check* (PPC) was performed. Briefly, we used the BTM to generate new predictive data $\hat{y}$ and checked if it resembled the observed data. The full posterior distribution is exploited in this way to generate a *posterior predictive distribution*

$$P(\hat{y}|y) = \int P(\hat{y}|\theta)P(\theta|y)\text{d}\theta, \tag{11}$$

which propagates the uncertainty in the parameter estimates to uncertainty about prediction (Betancourt 2015, McElreath 2015, Gabry *et al* 2017). In this way, PPCs allow to detect systematic modeling errors and violations of model assumptions. Subsequently, the posterior distributions of K^{trans} were compared across visits for all patients with the objective to discriminate between patients with pCR and non-pCR.

2.4. Statistical analysis

A quantitative statistical measure for signal fidelity is the structural similarity index (SSIM) (Wang *et al* 2004). It gives an average value over similarities of three key elements of an image: luminance, contrast and structure (Wang and Bovik 2009). To assess the accuracy of parameter estimates for the DRO, the SSIM was calculated between the estimated and the true parameter maps. As a comparison, the root-mean-squared error (RMSE) was calculated alongside. In order to get reasonable values for RMSE, outliers in K^{trans}-estimates obtained from NLLS fitting needed to be restricted to one. In addition, the SSIM was calculated between the σ-uncertainty maps determined with the BTM and the bootstrapping method to assess similarities in the precision of estimates.

To compare the K^{trans} posterior distributions between visits for the breast cancer DCE-MRI dataset, Cohen's d was calculated for each of the ten patients as:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}}. \tag{12}$$

$\bar{x}$ represents the average K^{trans} value per visit, σ its standard deviation. In this way, the width of the posterior distributions are incorporated into a single value. Compared to just reporting the percentage change of K^{trans} mean values, the uncertainty in parameter estimation is accounted for. An univariate logistic regression (ULR) model, implemented in *scikit-learn*[9] (Pedregosa *et al* 2011), was fitted to the Cohen's d values. The receiver operating characteristic (ROC) area under curve (AUC) was calculated in order to obtain a quantitative measure for the assessment of response.

Figure 2. Estimated K^{trans} with BTM (top) and NLLS approach (bottom). The left column displays median (a) and point estimates (d). The middle column ((b) and (e)) shows the calculated percentage error between estimates and ground truth. The right column illustrates the uncertainty σ of the Bayesian posterior (c) and the additional bootstrap samples (f).
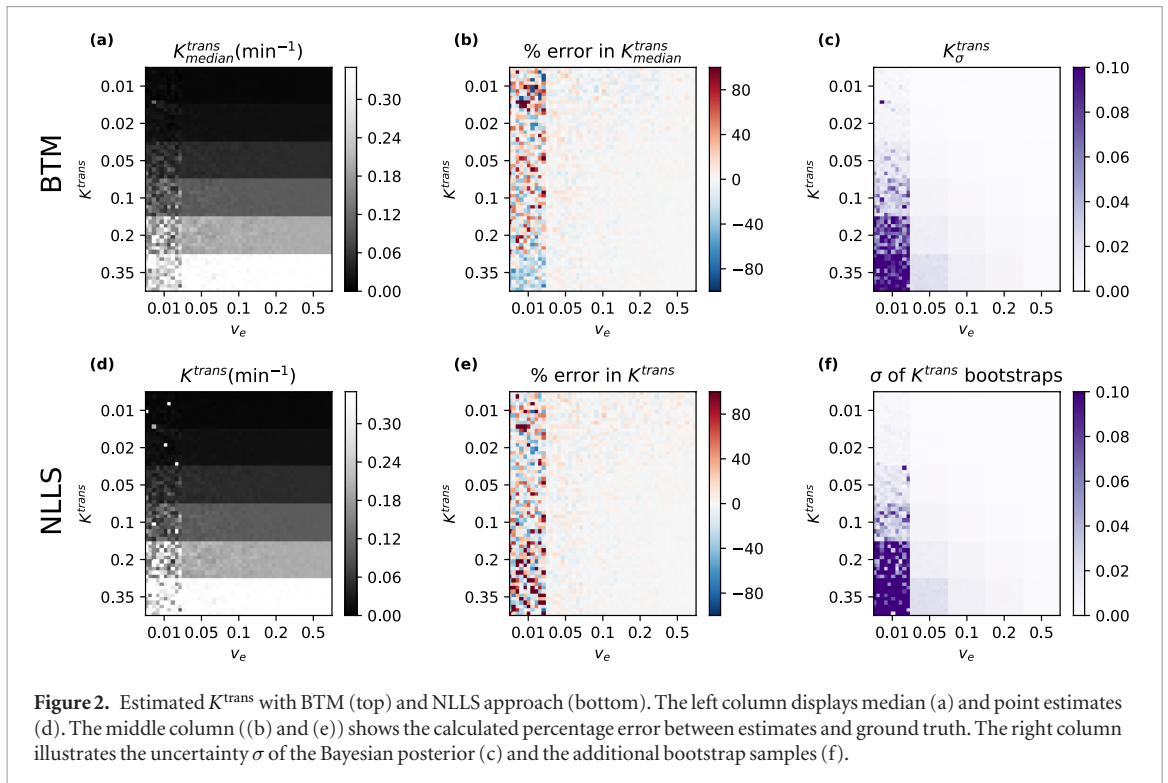
Table 2. SSIM and RMSE between estimated DRO K^{trans} and v_e parameter maps and ground truth for both approaches; SSIM of 100% indicates perfect similarity.

	K^{trans}		v_e	
	BTM (%)	NLLS (%)	BTM (%)	NLLS (%)
SSIM	96	91	92	94
RMSE	2.5	7.0	4.1	5.4

BTM = Bayesian tofts model; NLLS = Non-linear least squares approach; SSIM = Structural similarity index; RMSE = Root-mean-squared error.

3. Results

3.1. Validation: QIBA DCE-MRI phantom

Concentration-time curves of the DRO were evaluated within a Bayesian and likelihood framework. The resulting parameter estimates for K^{trans} are exemplarily shown in figure 2 for the Bayesian approach (a) and the NLLS reference (d). Note that the voxels in the Bayesian framework show median values of their respective posterior distributions while voxels in the likelihood framework represent point estimates. In general, the parameter maps show high accordance with the true values. The corresponding percentage error maps in the middle column (b) and (e) display relatively low errors for all regions with $v_e > 0.01$ for both methods. Low accuracy, hence high percentage errors are observed for regions where $v_e = 0.01$. SSIM between estimated and true K^{trans}-maps is higher for the BTM than for the NLLS approach. Furthermore, RMSE is lower for the BTM for both PK parameter maps. Details are provided in table 2.

The right column of figure 2 displays the precision of the parameter estimates evaluated with the BTM (c) and a bootstrapping method applied to the fit results of the NLLS approach (f). The visual analysis of the uncertainty maps reveals very similar patterns for both approaches, supported by a SSIM of 91%. The highest uncertainty occurs in regions with the highest percentage error for the fitting parameter estimates. The remaining parameter combinations have much greater precision. Information about divergences (BTM) and pixels where the NLLS algorithm did not find a solution can be found in table 3, together with computational times for fitting all 3000 pixels with BTM and NLLS approaches and the additional bootstrap analysis.

3.2. Application: breast cancer DCE-MRI data

Figure 3 shows representative signal intensity-time curves with the associated PPCs (a)–(c) and their corresponding K^{trans} posterior distributions (d). Here, the dark line illustrates the median and the increasingly lighter bands are the 20%, 40%, 60% and 80% highest density intervals (HDI) between the corresponding (0.4,0.6), (0.3,0.7), (0.2,0.8) and (0.1,0.9) percentiles of the posterior predictive distribution. The PPC in (a)

Table 3. Fitting process and parameter estimation of all 50×60 DRO concentration curves.

	BTM	NLLS
Divergences	17	27
Computational time: fitting	∼48 min	∼2 min
Computational time: uncertainty	*included*	∼2100 min[a]

BTM = Bayesian tofts model; NLLS = Non-linear least squares approach.

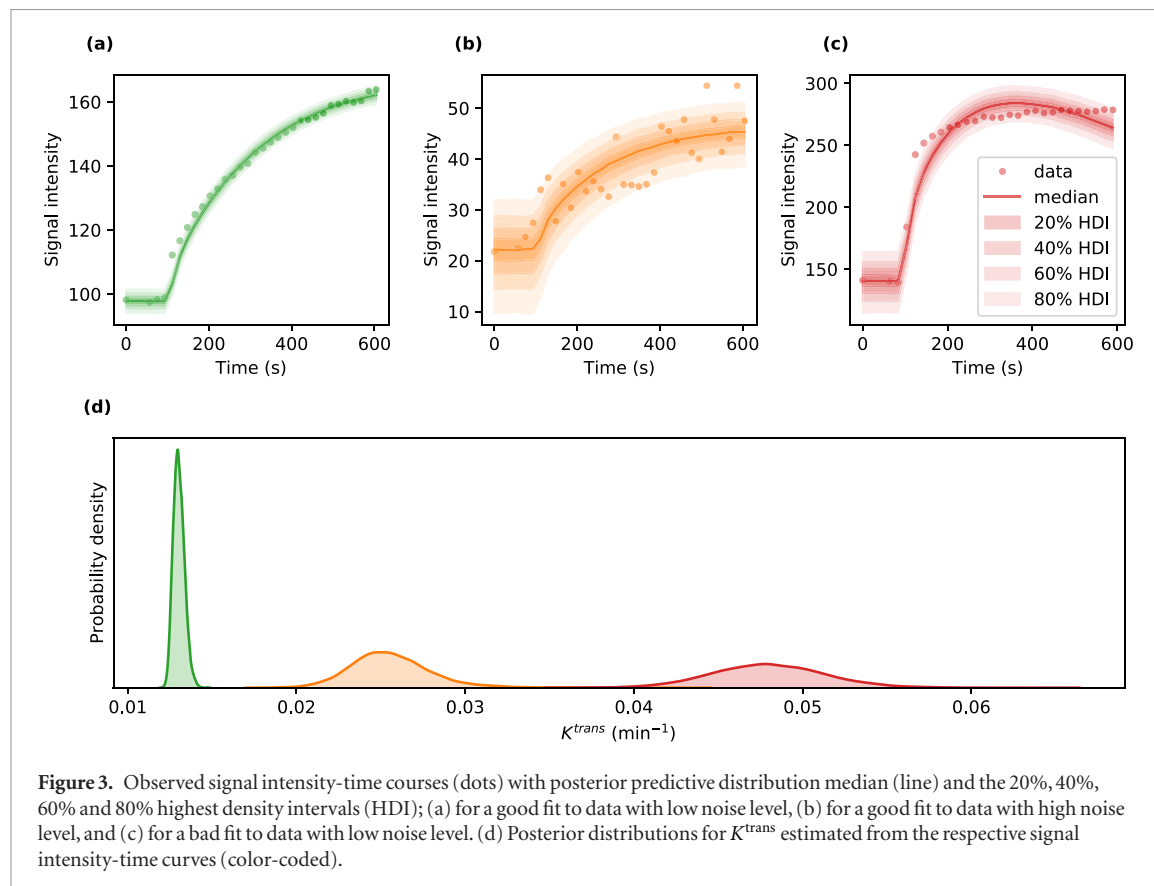[a] Based on additional bootstrap analysis.

Figure 3. Observed signal intensity-time courses (dots) with posterior predictive distribution median (line) and the 20%, 40%, 60% and 80% highest density intervals (HDI); (a) for a good fit to data with low noise level, (b) for a good fit to data with high noise level, and (c) for a bad fit to data with low noise level. (d) Posterior distributions for K^{trans} estimated from the respective signal intensity-time curves (color-coded).
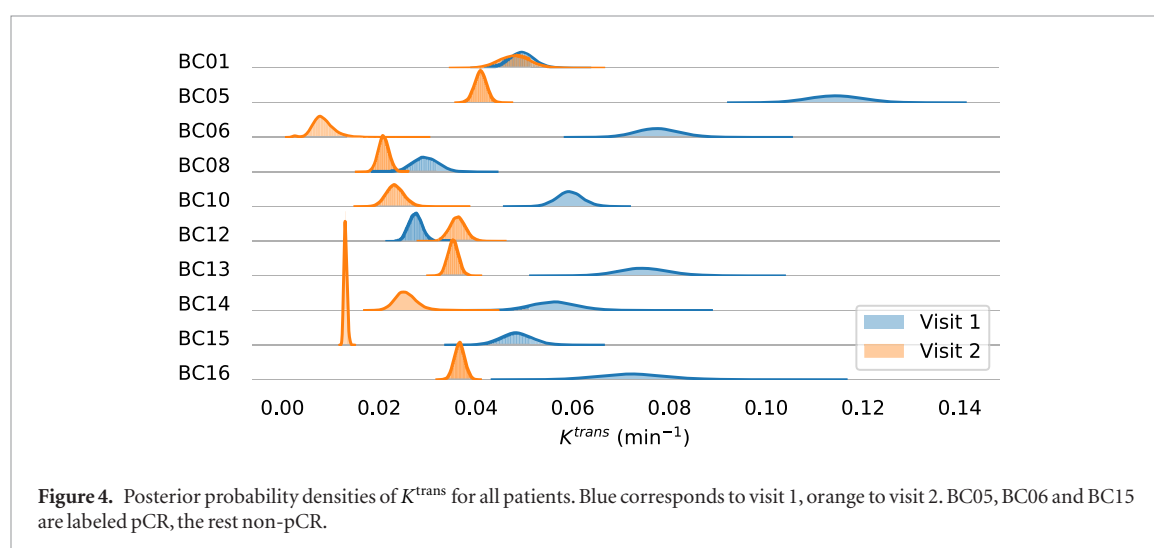

Figure 4. Posterior probability densities of K^{trans} for all patients. Blue corresponds to visit 1, orange to visit 2. BC05, BC06 and BC15 are labeled pCR, the rest non-pCR.

indicates a good fit of the model to the data, the corresponding posterior distribution (green) for K^{trans} is narrow. The PPC in (b) suggests that the chosen model provides a good fit to the data, the high noise level in the data is associated with a broader posterior distribution (orange). In (c), the noise level of the data is comparable to (a), however the PPC indicates a modeling error.

Figure 4 shows the posterior distributions of K^{trans} for all patients for visit 1 (blue) and visit 2 (orange), before and during NACT, respectively. With one exception, a general decrease in K^{trans} is observed. The degree

Figure 5. Cohen's *d* calculated from K^{trans} posterior distributions for all patients; sorted by value.
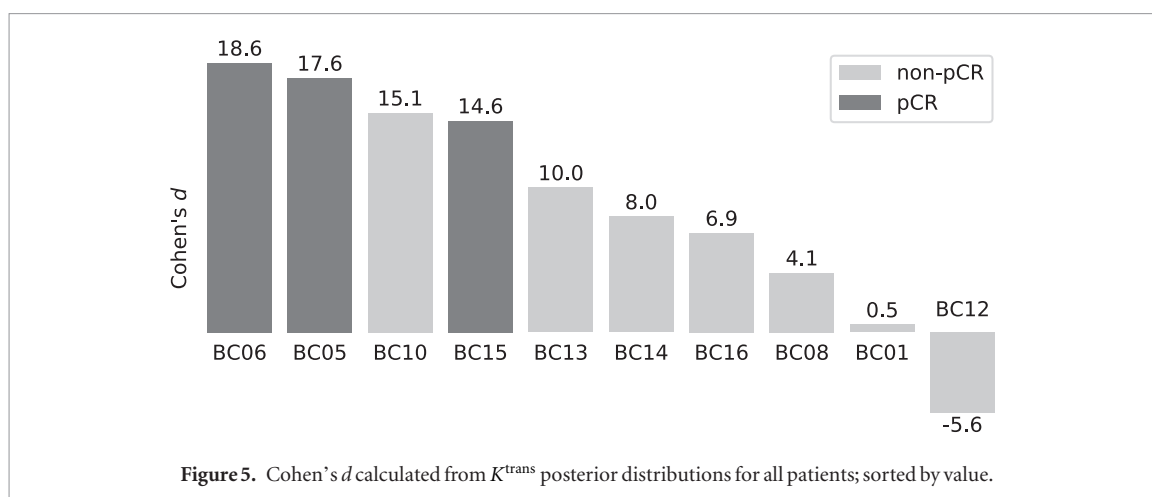
of change, dependent on the width of the posterior distributions, is summarized in Cohen's *d* values and visualized in figure 5; light-gray represents non-pCR, dark-gray pCR. The ULR analysis revealed a ROC AUC of 0.952. Computational time for fitting all 20 ROI-averaged concentration curves was ∼20 s for the BTM.
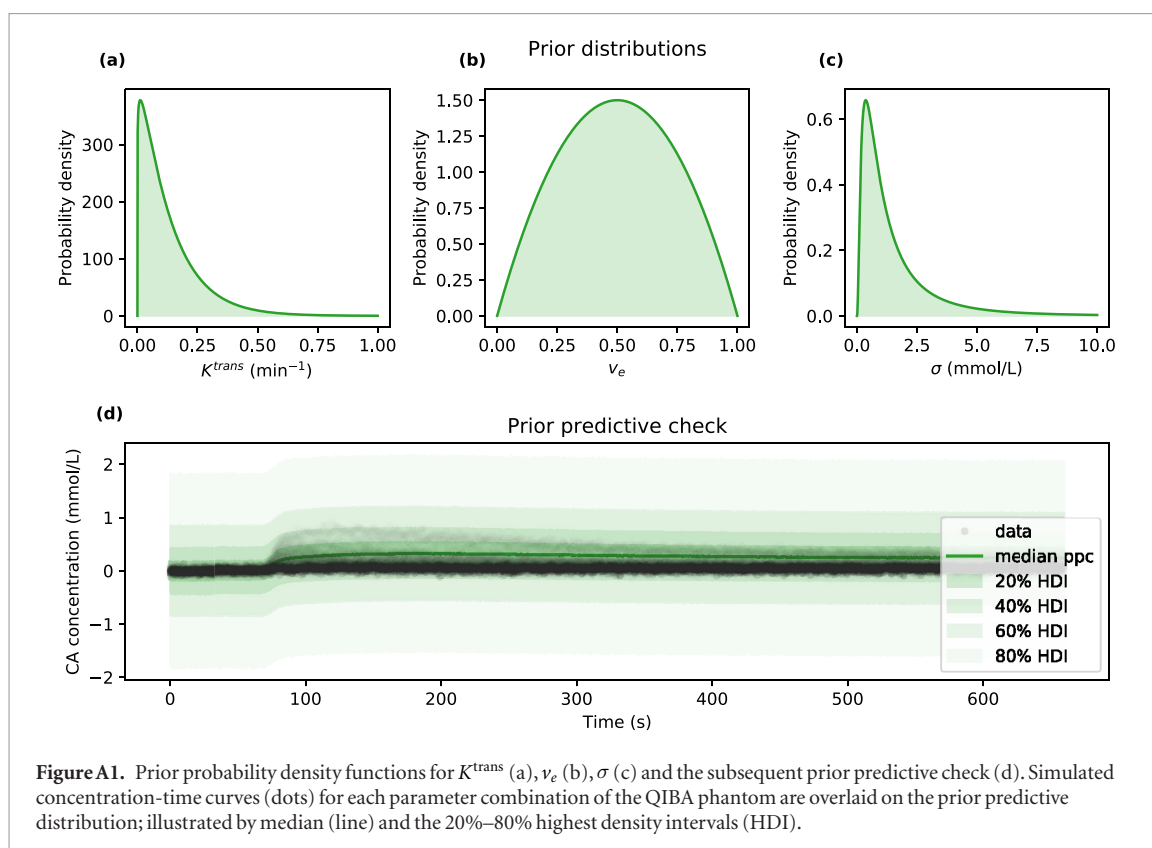
4. Discussion

In this study, we assessed posterior probability distributions of tracer-kinetic parameters obtained with a BTM against a standard NLLS approach. Validation with a DRO revealed high accuracy of BTM and NLLS approaches, indicated by strong similarity between estimated and ground truth maps. In addition, precision of estimates, assessed via the width of the posterior probability distributions and bootstrapping, respectively, was in very good agreement between both approaches. Analysis of the breast cancer DCE-MRI dataset with the BTM revealed that the degree of decrease in K^{trans} gives information about the pathologic response to NACT. The response in dependence of the uncertainty of parameter estimates was quantified with Cohen's *d*, calculated from the posterior distributions between visit 1 and 2. ULR modeling indicated excellent prediction of response.

Concerning the analysis of the DRO with the BTM, median parameter estimates were compared to the ground truth to assess the accuracy, otherwise not available with measured data. It was found that the Bayesian estimates generally have a very strong similarity with the ground truth, validating the accuracy of our BTM. The recovered parameters also have complementary regions of high and low percentage errors compared to the established NLLS fitting routine. RMS errors were lower for both implementations in the present work compared to similar DRO analysis by Smith *et al* (2015) and Ortuño *et al* (2013). Albeit, the results are in good comparison. Caution is still required for voxels with low v_e. Concentration curves with these parameter combinations have very limited intensity changes which practically vanish in the added background noise.

The variance of estimates inherent in the Bayesian posterior distribution was compared to a bootstrapping error analysis, performed likewise to the work of Kershaw and Buckley (2006). It was demonstrated that the uncertainty maps of the BTM resemble those calculated with the bootstrap analysis, validating the precision of parameters recovered with the BTM. To the best of our knowledge, only (Schmid *et al* 2006) implemented a Bayesian PK model with the objective to make use of the posterior probability distribution. Parts of the present study build up on their work and go beyond by capturing even more information from the posterior distribution. Their approach was applied to patient data only, whereas in the present work, accuracy and precision of estimates were validated with a digital phantom first.

Furthermore, we applied the BTM to the breast cancer DCE-MRI data, performed PPCs and investigated the posterior distributions. For a PPC, the observed data was compared to the posterior predictive distribution, illustrated as percentile intervals of highest density. A good fit to the data results in a posterior distribution which reflects the noise level in the data; low noise corresponds to a narrow posterior and vice versa. However, a bad fit to the data results in a broad posterior distribution despite a low noise level. This indicates a systematic modeling error which influences the information we gained about uncertainty. More complex PK models which incorporate additional assumptions about CA transport, e.g. the extended Tofts model, could be able to produce a better fit to certain data. Hence, assessing posterior distributions requires to check the corresponding data and fit before drawing any conclusions from it. While feasible for ROI-based analysis with only a handful of concentration curves, visual assessment is not possible in a pixelwise analysis. However, this can be necessary considering that tumorous tissue can be very heterogeneous (Wu *et al* 2018). An automated Bayesian model selection step as proposed in the work of Duan *et al* (2017) could be an effective means to reduce systematic modeling error but is beyond the scope of this study.

Figure A1. Prior probability density functions for K^{trans} (a), v_e (b), σ (c) and the subsequent prior predictive check (d). Simulated concentration-time curves (dots) for each parameter combination of the QIBA phantom are overlaid on the prior predictive distribution; illustrated by median (line) and the 20%–80% highest density intervals (HDI).

In order to assess therapy response for the patients in the breast cancer DCE-MRI dataset, Huang *et al* (2014a) showed in their original work that using visit 2 K^{trans} or the percentage change of K^{trans} between visits as metrics yields good to excellent results. However, the uncertainty in estimating PK parameters with tracer-kinetic models is not accounted for. For this purpose, we calculated Cohen's d as a means of quantitative change in parameter estimates which depends on the precision of estimates. In contrast, the authors of Schmid *et al* (2006) used the posterior distributions in each voxel to apply probabilistic thresholding, and then compared the mean values of K^{trans} between visits. Using Cohen's d metric, the assessment of response was found to be excellent by means of an ULR analysis. Considering the findings of the PPCs, including a model selection step as explained above could decrease the influence of systematic modeling errors on posterior distributions and hence Cohen's d values which may further improve assessment of therapy response.

Limitations of the present work include large computational time when fitting the BTM to the DRO-data. On the one hand, the MCMC sampling is time and memory consuming but necessary to avoid divergences. On the other hand, it yields a full posterior probability distribution with information about the uncertainty, and obtaining the same information with a bootstrap analysis of a NLLS fit requires even more computation time. Furthermore, the simulated DRO curves have a much higher time-resolution compared to measured data. Evaluating real DCE-MRI data increases the speed of the analysis greatly. Moreover, the influence of the chosen prior distributions on the results was not assessed in the present study.

In conclusion, we evaluated a BTM with a DRO, assessed accuracy and precision against the standard NLLS approach and showed how posterior distributions are used to assess therapy response. We demonstrated that Bayesian modeling provides an elegant means to assess posterior probability distributions, which are in good agreement with established approaches.

Acknowledgments

Appendix. Prior predictive check

To assess if the choice of prior distributions for the model parameters covers a reasonable range of concentration-time curves, it is useful to perform a *prior predictive check*. For this purpose, we generated 100 000 MCMC samples from the prior predictive distribution,

$$P(\hat{y}) = \int P(\hat{y} \mid \theta) \, P(\theta) \mathrm{d}\theta, \tag{A.1}$$

only considering the prior distributions without any actual data. This quantifies the range of possible observations $\hat{y}$, predicted by our model. In a prior predictive check, the predicted data is compared to real observations and the extent of extreme observations indicates the level of disagreement between domain expertise and model assumptions. Figure A1 shows the probability density functions of the chosen priors (a)–(c) and the prior predictive check (d). The black dots are actual observed data from the QIBA phantom, one curve for each parameter combination of K^{trans} and v_e, to assess the scope of possible phantom curves. The increasingly lighter green bands represent the 20%, 40%, 60% and 80% highest density intervals between the corresponding percentiles of the prior predictive distribution; the green line is the median thereof. We find that the model predicts observations that are more extreme than the phantom data but not too extreme to be unrealistic given the assumed observational error. Hence, we conclude that the chosen prior distributions are reasonable and weakly informative.

ORCID iDs

Andreas Mittermeier ● https://orcid.org/0000-0001-5743-473X
Olaf Dietrich ● https://orcid.org/0000-0001-6182-5039
Michael Ingrisch ● https://orcid.org/0000-0003-0268-9078

References

Betancourt M 2015 A unified treatment of predictive model comparison (arXiv:1506.02273 [stat])

Beuzit L, Eliat P A, Brun V, Ferré J C, Gandon Y, Bannier E and Saint-Jalmes H 2016 Dynamic contrast-enhanced MRI: study of inter-software accuracy and reproducibility using simulated and clinical data *J. Magn. Reson. Imaging* **43** 1288–300

Byrd R H, Lu P, Nocedal J and Zhu C 1994 A limited-memory algorithm for bound constrained optimization *SIAM J. Sci. Comput.* **16** 1190–208

Carpenter B, Gelman A, Hoffman M D, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P and Riddell A 2017 Stan: a probabilistic programming language *J. Stat. Softw.* **76** 1–32

Clark K *et al* 2013 The cancer imaging archive (TCIA): maintaining and operating a public information repository *J. Digit. Imaging* **26** 1045–57

Debus C, Floca R, Ingrisch M, Kompan I, Maier-Hein K, Abdollahi A and Nolden M 2019 MITK-ModelFit: a generic open-source framework for model fits and their exploration in medical imaging design, implementation and application on the example of DCE-MRI *BMC Bioinform.* **20** 31

Dikaios N, Atkinson D, Tudisca C, Purpura P, Forster M, Ahmed H, Beale T, Emberton M and Punwani S 2017 A comparison of Bayesian and non-linear regression methods for robust estimation of pharmacokinetics in DCE-MRI and how it affects cancer diagnosis *Comput. Med. Imaging Graph.* **56** 1–10

Duan C, Kallehauge J F, Bretthorst G L, Tanderup K, Ackerman J J H and Garbow J R 2017 Are complex DCE-MRI models supported by clinical data? *Magn. Reson. Med.* **77** 1329–39

Gabry J, Simpson D, Vehtari A, Betancourt M and Gelman A 2017 Visualization in Bayesian workflow *J. Royal Stat. Soc.* A **182** 389–402

Gelman A and Rubin D B 1992 Inference from iterative simulation using multiple sequences *Stat. Sci.* **7** 457–72

Geyer C J 2011 Introduction to Markov chain Monte Carlo *Handbook of Markov Chain Monte Carlo* ed S Brooks *et al* (London: Chapman and Hall) p 46

Gilks W R, Richardson S and Spiegelhalter D 1995 *Markov Chain Monte Carlo in Practice* (Boca Raton, FL: CRC Press)

Hansen M B, Tietze A, Haack S, Kallehauge J, Mikkelsen I K, Østergaard L and Mouridsen K 2019 Robust estimation of hemo-dynamic parameters in traditional DCE-MRI models *PLoS oNE* **14** e0209891

Hoffman M D and Gelman A 2011 The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo (arXiv:1111.4246)

Huang W *et al* 2014a Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge *Transl. Oncol.* **7** 153–66

Huang W, Tudorica A, Chui S, Kemmer K, Naik A, Troxell M, Oh K, Roy N, Afzal A and Holtorf M 2014b Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge *The Cancer Imaging Archive* (https://doi.org/10.7937/K9/TCIA.2014.A2N1IXOX)

Ingrisch M and Sourbron S 2013 Tracer-kinetic modeling of dynamic contrast-enhanced MRI and CT: a primer *J. Pharmacokinetics Pharmacodynamics* **40** 281–300

Jones E, Oliphant T and Peterson P 2001 SciPy: open source scientific tools for Python www.scipy.org/

Kershaw L E and Buckley D L 2006 Precision in measurements of perfusion and microvascular permeability with T1-weighted dynamic contrast-enhanced MRI *Magn. Reson. Med.* **56** 986–92

Kruschke J 2014 *Doing Bayesian Data Analysis* 2nd edn (New York: Academic)

McElreath R 2015 *Statistical Rethinking: a Bayesian Course with Examples in R and Stan* 1st edn (London: Chapman and Hall)

Orton M R, Collins D J, Walker-Samuel S, d'Arcy J A, Hawkes D J, Atkinson D and Leach M O 2007 Bayesian estimation of pharmacokinetic parameters for DCE-MRI with a robust treatment of enhancement onset time *Phys. Med. Biol.* **52** 2393

Ortuño J E, Ledesma-Carbayo M J, Simões R V, Candiota A P, Arús C and Santos A 2013 DCE@urLAB: a dynamic contrast-enhanced MRI pharmacokinetic analysis tool for preclinical data *BMC Bioinform.* **14** 316

Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30

Pintaske J, Martirosian P, Graf H, Erb G, Lodemann K P, Claussen C D and Schick F 2006 Relaxivity of gadopentetate dimeglumine (magnevist), gadobutrol (gadovist), and gadobenate dimeglumine (MultiHance) in human blood plasma at 0.2, 1.5, and 3 T *Investigative Radiol.* **41** 213–21

Roberts C, Issa B, Stone A, Jackson A, Waterton J C and Parker G J M 2006 Comparative study into the robustness of compartmental modeling and model-free analysis in DCE-MRI studies *J. Magn. Reson. Imaging* **23** 554–63

Schmid V J, Whitcher B, Padhani A R, Taylor N J and Yang G Z 2006 Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging *IEEE Trans. Med. Imaging* **25** 1627–36

Seber G A F and Wild C J 2003 *Nonlinear Regression* (New York: Wiley)

Shukla-Dave A *et al* 2018 Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials *J. Magn. Reson. Imaging* **49** e101–e121

Smith D S, Li X, Arlinghaus L R, Yankeelov T E and Welch E B 2015 DCEMRI.jl: a fast, validated, open source toolkit for dynamic contrast enhanced MRI analysis *PeerJ* **3** e909

Sourbron S P and Buckley D L 2011 On the scope and interpretation of the Tofts models for DCE-MRI *Magn. Reson. Med.* **66** 735–45

Sourbron S P and Buckley D L 2012 Tracer kinetic modelling in MRI: estimating perfusion and capillary permeability *Phys. Med. Biol.* **57** R1–33

Sourbron S P and Buckley D L 2013 Classic models for dynamic contrast-enhanced MRI *NMR Biomed.* **26** 1004–27

Tietze A, Nielsen A, Mikkelsen I K, Hansen M B, Obel A, Østergaard L and Mouridsen K 2018 Bayesian modeling of dynamic contrast enhanced MRI data in cerebral glioma patients improves the diagnostic quality of hemodynamic parameter maps *PLoS One* **13** e0202906

Tofts P S 1997 Modeling tracer kinetics in dynamic Gd-DTPA MR imaging *J. Magn. Reson. Imaging* **7** 91–101

Tofts P S and Kermode A G 1991 Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts *Magn. Reson. Med.* **17** 357–67

Tofts P S *et al* 1999 Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols *J. Magn. Reson. Imaging* **10** 223–32

Wang Z and Bovik A 2009 Mean squared error: Love it or leave it? A new look at signal fidelity measures *IEEE Signal Process. Mag.* **26** 98–117

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Proc.: Publ. IEEE Signal Proc. Soc.* 13 600–12

Woolrich M W, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M and Smith S M 2009 Bayesian analysis of neuroimaging data in FSL *NeuroImage* **45** S173–86

Wu J, Cao G, Sun X, Lee J, Rubin D L, Napel S, Kurian A W, Daniel B L and Li R 2018 Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy *Radiology* **288** 26–35

Zhu C, Byrd R H, Lu P and Nocedal J 1997 Algorithm 778: L-BFGS-B: fortran subroutines for large-scale bound-constrained optimization *ACM Trans. Math. Softw.* **23** 550–60