Boosting in Nonlinear Regression Models with an Application to DCE-MRI Data*

M. Feilke1; B. Bischl1,2; V. J. Schmid1; J. Gertheiss3,4

1Institut für Statistik, Ludwig-Maximilians-Universität München, München, Germany;
2Fakultät Statistik, Technische Universität Dortmund, Dortmund, Germany;
3Department für Nutztierwissenschaften, Georg-August-Universität Göttingen, Göttingen, Germany;
4Zentrum für Statistik, Georg-August-Universität Göttingen, Göttingen, Germany

Methods: The observed uptake of contrast agent in a voxel of an image of some tissue is described by a concentration time curve. This curve can be modeled using a nonlinear regression model. We present a boosting approach with nonlinear regression as base procedure, which allows us to estimate the number of compartments and the related parameters for each voxel of an DCE-MR image. In addition, a spatially regularized version of this approach is proposed.

Results: With the proposed approach, the number of compartments and the related parameters for each voxel of an DCE-MR image can be fitted for a varying, a priori fixed number of compartments. It is shown that this approach is more flexible than the one-compartment model and outperforms other competing methods.

Conclusions: The proposed boosting approaches outperform all competing methods considered in this paper regarding the correct localization of tumors in DCE-MR images as well as the spatial homogeneity of the estimated number of compartments across the image, and the definition of the tumor edge.

1. Introduction

DCE-MRI is an imaging technique by which the blood supply of a tissue of interest can be recorded in vivo. A series of images is acquired by an MRI scanner that captures images of the tissue at several time points after the injection of a contrast agent (CA). An excerpt of the DCE-MRI series of a breast cancer patient showing the concentration of contrast agent at eight different time points can be found in Figure 2 in the Appendix. For each voxel of an image, the concentration of CA at each time point can be computed from the MR signal [1]. From the dynamic behavior of CA uptake, tumors can be located, malignancy and types of tumors can be inferred, tumors can be graded and therapy success can be assessed [2, 3].

For the analysis of the CA uptake behavior, model-based or data-driven methods can be used. An advantage of model-based methods is that they result in quantitative physiological parameters which characterize the amount and rate of capillary leakage [2], as they are based on pharmacokinetic models describing the exchange of CA between different, well-mixed compartments [4]. For data-driven methods such as non-parametric regression, usually no a priori compartment-structure is assumed [5].

So far, several compartment models with a varying, a priori fixed, number of compartments have been proposed and it is not clear which model should be used in which situation. The two-compartment exchange models proposed by [6] and [7], for example, consists of two different compartments for arterial plasma and interstitial plasma. Furthermore, a hierarchical Bayesian two-compartment model for the...
analysis of DCE-MRI data on voxel level has been proposed by [8], and a multi-tissue compartment model with a fixed number of compartments has been proposed by [9].

In different types of tissue, however, different numbers of compartments might be needed. Moreover, as tumor tissue is often heterogeneous, the adequate number of compartments might even vary over a field of voxels. Therefore, we propose to estimate the number of compartments for each voxel from the data.

Sommer et al. [10] recently proposed a spatially regularized estimation method based on a multi-tissue compartment model for the estimation of the number of compartments and the related parameters per voxel. Thereby, they combined the advantages of model-based and data-driven methods, as the number of compartments is chosen data-dependently, but biologically interpretable parameter estimates are obtained.

Similar to [10], we base our approach on a multi-tissue compartment model. We, however, propose a more straightforward method to estimate the number of compartments and the parameters of this model: a boosting approach with nonlinear regression as base procedure. The original boosting algorithm [11] arose in the machine learning community and was mainly used for classification [12]. Later, the concept of boosting was adapted to the field of regression modeling [13–17], where it can be used in various settings to select predictors and estimate their effects on a univariate continuous response [18]. Gradient boosting algorithms are currently gaining attention, as they can be very useful to address important research questions in modern biomedicine [18].

The boosting algorithm that we will be introduced in subsection 2.4 together with the voxelwise or the spatially regularized estimation procedure presented in subsections 2.5 and 2.6 is a completely novel approach for a regression setting where the univariate response variable is described by a nonlinear parametric function. Specifically, it tackles the problem of the estimation of the number of compartments in a multi-tissue compartment model and the related parameter estimation. To the best of our knowledge, in contrast to the extensive literature on boosting in additive regression, boosting used in nonlinear parametric regression is described for the first time in this paper.

The remainder of the paper is organized as follows. In Section 2, the compartment model and the nonlinear regression model built upon it are introduced. Moreover, gradient boosting in general, the novel boosting algorithm, and the associated voxelwise and spatially regularized estimation procedures are described. In Section 3, the simulation setup and the results of the simulation studies are presented. The evaluation of the proposed method for in vivo data from a breast cancer study is found in Section 4. Section 5 gives a discussion on the paper and concludes.

With the R package dcemriboost [19], which is available as supplementary file, voxelwise and spatially regularized boosting for DCE-MRI data as described in this paper can be performed. This package contains the simulated data of Simulation 1 introduced in Section 3 of this paper. Moreover, we uploaded two accompanying R files to GitHub. These files provide code for performing voxelwise boosting and spatially regularized boosting ($\lambda = 10^{-10}$) for the data of Simulation 1.

2. Methods
2.1 Compartment Model

For the analysis of DCE-MRI data, several compartment models with a varying number of compartments have been used. A fundamental compartment model is the so-called Tofts model [20]. A block diagram for the Tofts model is given in Figure 2. In this model, the perfusion of CA is described with the help of an arterial plasma compartment (APC) and an interstitial space compartment, which is also called extravascular extracellular space (EES) [20, 21]. It is assumed that the contrast agent enters only the interstitial space and not the cells. This assumption holds typically when using low-molecular-weight tracers such as gadolinium diethylenetriaminepentaacetic acid (Gd-DTPA). CA is exchanged between the APC and the EES at constant rate $k_{ep}$.

In the Tofts model, the concentration $C(t)$ of CA at time $t$ is the product of the concentration $C_{ep}(t)$ of CA in the EES at time $t$ and the fractional volume of the EES $v_{ep}$. $C(t) = e(t) = C_{ep}(t) \cdot K_{trans} \exp(-k_{ep} t)$, where $*$ is the convolution operator such that $C_{ep}(t)^* \exp(-k_{ep} t) = \int_{-\infty}^{t} C_{ep}(t - \tau) \exp(-k_{ep} \tau) d\tau$ [10].

The parameters of interest, which have to be estimated, are $k_{ep}$ and $K_{trans}$. $k_{ep}$ is the rate constant at which the EES exchanges CA with the APC, and $K_{trans} := v_{ep}K_{trans}$ is the associated volume transfer constant. Observed tracer kinetics can often be well described by the Tofts model for healthy tissue despite the simplifying assumptions made [22]. However, as the microvasculature in tumors often is highly heterogeneous [23], the Tofts model fails to describe its observed uptake dynamics [9, 24], and therefore, more complex compartment models are needed [22].

We use a multi-tissue compartment model with $q$ tissue compartments, which is a generalization of the Tofts model:

$$C(t) = C(t)^* = \sum_{i=1}^{q} C_{ep}(t) \ast K_{trans}^i \exp(-k_{ep}^i t).$$

This model can also be obtained by solving a system of differential equations derived from the compartment model in Figure 3 with some initial conditions [25]. With this model, it is possible to map tissue heterogeneity on the voxel level. For each voxel, we assume the compartment model in Equation 1, and estimate the number of compartments $q$. Therefore, the number of compartments can vary over the voxels of an image. Moreover, each tissue compartment $k, k = 1, \ldots, q$, has a unique rate constant $k_{ep^k}$ at which CA is exchanged between the APC and the EES, as well as an individual volume transfer constant $K_{trans^k}$. Of course, these parameters have to be estimated in addition to $q$.

2.2 Nonlinear Regression Model

Let the observed concentration of CA in voxel $i$, $i = 1, \ldots, N$, at time $x_i$, $t = 1, \ldots, T$, be denoted by $Y_i(x_i)$, and let $C_i(x_i)$ be the expected concentration of CA in voxel $i$ at time $x_i$. We assume that $C_i(x_i)$ can be described by the nonlinear multi-tissue com-
partment model with $q_i$ tissue compartments

$$C_i(x) = \sum_{t=1}^{\hat{w}} C_{p}(x_i) \cdot K^{\text{trim}}_{q_i} \cdot \exp(-k_{q_i} x_i) \quad (2)$$

Furthermore, we assumed that the observed concentration of CA in voxel $i$ can be modeled by

$$f_i(x_i) = C_i(x_i) + \epsilon_i = \sum_{t=1}^{\hat{w}} C_{p}(x_i) \cdot K^{\text{trim}}_{q_i} \cdot \exp(-k_{q_i} x_i) + \epsilon_i \quad (3)$$

where $\epsilon$ is independent Gaussian noise terms with mean zero and variance $\sigma^2$ [26].

For $k_{q_i}$ values between 0.05 and 20 are feasible, considering that the rate is fast and does typically not exceed 20 [26], and for $K^{\text{trim}}_{q_i}$ values in the interval [0.01, 20] are considered [26, 27], as these ranges of values are biologically realistic. Since in breast DCE-MRI there are no big vessels in the captured from which $C_p(x_i)$ could be measured, we use a bi-exponential population based arterial input function as proposed by [4]: $C_p(x_i) = D(a_i \cdot \exp(-m_i x_i) + a_2 \cdot \exp(-m_2 x_i))$, with constant dose D of tracer (in mmol/kg) depending on the experimental conditions and fixed values $a_1 = 3.99 \text{ kg/l}$, $a_2 = 4.78 \text{ kg/l}$, $m_1 = 0.144 \text{ min}^{-1}$, $m_2 = 0.0111 \text{ min}^{-1}$ [4, 28].

### 2.3 Gradient Boosting

We use the boosting algorithm that is introduced in the following subsection to fit model (3), and thus to estimate the model complexity for each voxel of an MR image. We use a notation similar to the notation in [29], and adapt the general methodology presented there to the voxelwise DCE-MRI framework in this paper.

We consider the response $Y$ and the predictor function $C = f$. The aim is to receive an optimal prediction of $Y$, which is achieved by minimizing the loss function $P(Y, f) = \hat{\epsilon}$, over all admissible prediction functions $f$. This means, the optimal prediction function

$$j^* = \arg \min_f \mathbb{E}[P(Y, f)] \quad (4)$$

has to be estimated, where $P$ is assumed to be differentiable with respect to $f$. In our approach, we consider a nonlinear regression with response $Y \in \mathbb{R}$ and use the $L_2$ loss function $P(Y, f) = \frac{1}{2}(Y - f)^2$. As in practice, we deal with realizations $Y(x_i)$, $t = 1, \ldots, T$, of $Y$, the empirical risk $\hat{P} = \frac{1}{T} \sum_{t=1}^{T} P(Y(x_i), f(x_i))$ has to be minimized instead of the expected loss in Equation 4.

### 2.4 Boosting Algorithm

For each voxel $i$, we use the boosting algorithm in Figure 1 to minimize the empirical risk $\hat{P}$ over $f$.

By step 4(c), model choice is performed, because compartments are iteratively accepted into the model only if the model is improved by the acceptance. A new compartment is accepted only if its $k_{q_i}$-value differs substantially (i.e., by a factor of at least 5) from the $k_{q_i}$-values of the already accepted compartments. We set a factor for the relative difference of the $k_{q_i}$-values in the multi-tissue compartment model, because redundancy issues may occur if the exponential rates of two compartments are too similar. In [33], it has been shown that parameters in a sum of exponentials model are highly redundant if the exponential rates differ by less than a factor of 5 [22]. Therefore, the stepwise increments of the final estimator $f$ are multiplied by the step length factor $u$. Each update of compartments, the final boosting algorithm could overshoot the minimum of the empirical risk. We use $u = 0.1$, which is a typical value [29]. As $u_{[m]}$ depends on $x_i$, it is from now on denoted by $u_i(x_i)[m]$.

A substantial issue in the gradient boosting framework is the choice of the stopping iteration $m_{\text{stop}}$. Boosting algorithms should not be run until convergence, because otherwise, overfitting would be the consequence [17, 35]. Here, the optimal $m_{\text{stop}}$ is chosen according to the Bayesian Information Criterion (BIC) [36, 37]. For the computation of the BIC, the degrees of freedom of the model have to be defined. As we fit a compartment model where each compartment is defined by two parameters that are biologically interpretable, we think that the number of parameters in the model is a somewhat natural definition of the degrees of freedom here.

### 2.5 Voxelwise Estimation

In a first approach, we estimate the parameters $k_{q_i} = \{k_{q_1}, \ldots, k_{q_n}\}$ and $K^{\text{trim}}_{q_i} = \{K_{q_1}^{\text{trim}}, \ldots, K_{q_n}^{\text{trim}}\}$ independently for all voxels in an image, i.e., for each voxel $i$, we fully run the boosting algorithm described in Section 2.4.

In each iteration $m$, we have to minimize the residual sum of squares $\sum_i \left((x_i(x_i)) - \hat{\beta}_i(x_i)^{\text{trim}}\right)^2$. Hence, the non-linear least squares estimate is

$$\hat{k}_{\text{est}} = \left(\hat{k}_{\text{est}}[m], \ldots, \hat{k}_{\text{est}}[m]\right)^T \quad (5)$$

with

$$\hat{u}^{*} = C_{p}(x_i) \cdot K \cdot \exp(-k_{x_i})$$

and side conditions $0.05 \leq \hat{k}_{q_i} \leq 20$ and $0.01 \leq K^{\text{trim}}_{q_i} \leq 20$.

### 2.6 Spatially Regularized Estimation

To take into account the spatial structure of the voxels in an image, we perform, as a second approach, a spatially penalized estimation. We use a two-dimensional neighborhood structure where adjacent voxels are neighbors, i.e., each voxel has up to four neighbors. Voxels at the edge of an image...
1. Initialization: Set \( m = 0 \). Set \( \hat{f}_i^{[m]} = 0 \), because we fit a model without intercept. Generally, \( \hat{f}_i^{[m]} \) denotes the vector of function estimates in iteration \( m \) for voxel \( i \).

2. Specify a set of base-learners. Here, the base-learners are single compartments

\[
C_p(x_i) = K_{\text{trans}}^{x_i} \exp(-k_{\text{trans}}^{x_i} x_i)
\]

from the multi-tissue compartment model in (2).

3. Set \( m = m + 1 \).

4. 
   a) Compute the negative gradient of the loss function evaluated at the function estimates of the previous iteration:

\[
u_i^{[m]} = Y_i - \hat{f}_i^{[m-1]}, \]

where \( Y_i \) is the original response vector.

   b) Fit the nonlinear model in (5) to the negative gradient vector \( \nu_i^{[m]} \). This is done by a nonlinear least squares method. As optimization procedure, a Levenberg-Marquardt algorithm implemented in the R [30] function \( \text{nls} \) in package \text{lms} [31] is used [32]. Hereby, the best base-learner, i.e., the base-learner that fits \( \nu_i^{[m]} \) best according to the residual sum of squares (RSS) criterion is selected. The function estimate of the best base-learner, which minimizes the RSS, is denoted by \( \hat{u}_i^{[m]} \).

   As the output of the optimization procedure may depend on the starting values for the unknown parameters, ten initial guesses are drawn by latin hypercube sampling from the grid stretched by the biologically realistic ranges for the two parameters \( (k_{\text{trans}} \in [0.05, 20], K_{\text{trans}} \in [0.01, 20]) \). The optimization procedure is performed with all ten initial guesses and the base-learner that leads to the smallest RSS is chosen. Details on the choice of the number of initial guesses are given in Section 1 in the Appendix.

   c) Decide if a new compartment is accepted into the model. We use the fact that a compartment \( k \) is characterized by its \( k_{\text{trans}}^{x_i} \)-value and the new \( k_{\text{trans}}^{x_i} \)-value, i.e., the new compartment, if the estimated \( k_{\text{trans}}^{x_i} \)-value resulting from the optimization procedure in b) differs from the \( k_{\text{trans}}^{x_i} \)-values of the compartments that have already entered the model in the previous iterations by a factor of at least 5.

   - do not accept the new compartment otherwise. In this case, the model is fitted again with the previously accepted \( k_{\text{trans}}^{x_i} \)-value, i.e., only the parameter \( K_{\text{trans}} \) is estimated. The estimator \( K_{\text{trans}} \) is determined analytically. Details can be found in Section 1 in the Appendix.

   d) Update \( \hat{f}_i^{[m]} = \hat{f}_i^{[m-1]} + \nu\hat{u}_i^{[m]} \), with a step length factor \( \nu \), \( 0 < \nu \leq 1 \).

5. Iterate steps 3 and 4 until \( m = m_{\text{stop}} \), the stopping iteration.

Figure 1  Boosting algorithm

The nonlinear penalized least squares estimate for voxel \( i \) in iteration \( m \) is as in (6) with \( u^* = (C_p(x_i) * K \exp(-k_{\text{trans}}^{x_i}) + \lambda(k)) \), and side conditions \( 0.05 \leq k_{\text{trans}}^{x_i} \leq 20 \) and \( 0.01 \leq K_{\text{trans}}^{x_i} \leq 20 \), penalty parameter \( \lambda \), and penalization term

\[
J(k) = \sum_{i=1}^{q} \min \left( \tilde{k}_{\text{trans}}^{x_i} - k_{\text{trans}}^{x_i} \right)^2
\]

where \( \delta_i \) is the set of voxels in the neighborhood of pixel \( i \). This estimate corresponds to the estimate in the voxelwise approach, except for the penalization term \( \lambda(k) \), where \( J(k) \) corresponds to a Markov random field on the \( k_{\text{trans}}^{x_i} \)-values. By replacing the least squares base-learners by penalized variants, i.e., by the introduction of the penalization term \( \lambda(k) \), we ensure that neighboring voxels share similar \( k_{\text{trans}}^{x_i} \)-values, and therefore, spatial smoothness of these parameters is achieved.

Similar to [10], for efficiency reasons, we do a parallel update of the estimates of the voxels following a checkerboard pattern, using conditional independence from the remaining voxels given all neighboring voxels. By the checkerboard pattern, the voxels in an image are divided into “even” and “uneven” voxels. We do not fully run the boosting algorithm for each voxel as in 2.5. Rather, the updates of the parameter have less than four neighbors, whereas all other voxels have four neighbors.

For the penalization, we use again the fact that a compartment is characterized by its \( k_{\text{trans}}^{x_i} \)-value. While the contribution of a specific tissue compartment, i.e., the fractional volume \( v \), and consequently the volume transfer constant \( K_{\text{trans}}^{x_i} \) of this compartment, may vary considerably over a field of neighboring voxels, the rate constant \( k_{\text{trans}}^{x_i} \) at which the compartment exchanges CA with the APC should be similar for neighboring voxels [21, 22]. Thus, we penalize differences in the \( \tilde{k}_{\text{trans}}^{x_i} \)-values of adjacent voxels.
estimates are done alternately for the “even” and the “uneven” voxels. This means, in each iteration, the “even” and the “uneven” voxels are updated each once. In the first iteration, the $k_{ep}$-values of the “even” and the “uneven” voxels are penalized to some starting $k_{ep}$-values. To receive these starting values, we apply the voxelwise boosting algorithm (subsections 2.4 and 2.5) to the mean concentration time curve (CTC), which results from averaging the CTCs of all voxels in one image. From the second iteration on, the $k_{ep}$-values of the “even” voxels are penalized to the current $k_{ep}$-values of the “uneven” voxels and vice versa.

2.7 Refit

After the execution of the boosting algorithm with the voxelwise or the spatially regularized estimation procedure, for each voxel, the number of compartments and the $k_{ep}$-values are considered as fixed and a refit of the model is done in order to get the final $k_{ep}$-values. Using fixed $k_{ep}$-values, the nonlinear regression model is linearized, and therefore, the refit is done with the R function solve.QP (package quadprog [42]).

2.8 Competing Methods

Within the framework of the simulation and for the analysis of real DCE-MRI data, we compare the results of our two boosting approaches (voxel-wise and spatially regularized) with the results of fitting a Tofts model (subsection 2.1), as well as the results of a voxelwise and a spatially regularized estimation approach as described in [10]. The two approaches in [10] are based on basis functions $\psi_i(x) = C_i(x) \exp(-k_{ep} x)$. Using those, Equation 2 becomes $C_i(x) = \sum_{i=1}^{\text{num}} k_{ep}^{\text{num}} \psi_i(x)$. A set of candidate values for $k_{ep}$ is considered such that $\log(k_{ep}) \in [-3.0, -2.9, \ldots, 3.0]$, and suitable values, i.e., compartments, have to be chosen. Therefore, the unknown parameters $k_{ep}$ are estimated, and the $k_{ep}$-values related to positive $k_{ep}$-values are selected. As the usual ML-estimates are unstable or even not unique, two penalized approaches are proposed. In the first approach, the voxelwise regularized estimation, a penalized maximum likelihood estimator is used, with the (positive) elastic net [43] being chosen as penalty. In the second approach, the spatially regularized estimation, the penalty term used in the first approach is modified in such a way that it enforces spatial smoothness of parameters of neighboring voxels. Differences in the $k_{ep}$-values of adjacent voxels are penalized by a quadratic penalty term, which in our opinion, however, is not as straightforward as penalizing the $k_{ep}$-values of adjacent voxels as done in the spatially regularized boosting approach proposed in this paper.

In both approaches from [10] as well as in the spatially regularized boosting approach, the tuning parameters are determined according to the BIC. The neighborhood structure in [10] and in this paper coincides. In [10], the BIC is computed according to [36]. For the DCE-MRI data, the variance of the assumed normal distribution is estimated by the mean-squared error of a low-bias model [36]. We, however, compute the BIC for this data according to the formula

$$n \cdot \log \left( \sum_{i=1}^{n} \sum_{t} \left( y_i(t) - \hat{y}_i(t) \right)^2 \right) + p \cdot \log(n),$$

where $p$ is the number of parameters, i.e., the sum of the estimated number of compartments over all voxels in an image, and $n = N \cdot T$ is the number of voxels in an image multiplied by the number of time points at which an MR signal has been measured for this image. The approaches in [10] are from now on called ‘elastic net approaches’ (voxelwise and spatial elastic net approach, respectively).

3. Simulation Studies

3.1 Simulation Setup

For each of the simulations listed in Table 1, we simulated 10 DCE-MRI images, each consisting of 40 × 50 voxels. We

---

**Table 1 Simulation setup**

<table>
<thead>
<tr>
<th>Simulation</th>
<th>outside tumor</th>
<th>inside tumor</th>
<th>tumor edge</th>
<th>$k_{ep1}$</th>
<th>$k_{trans}$</th>
<th>$k_{ep2}$</th>
<th>$k_{trans}$</th>
<th>$k_{ep3}$</th>
<th>$k_{trans}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation 1</td>
<td>0.2</td>
<td>0.05</td>
<td>1.4</td>
<td>0.15</td>
<td>0.05</td>
<td>1.3</td>
<td>0.6</td>
<td>9.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Simulation 2</td>
<td>0.15</td>
<td>0.1</td>
<td>1.3</td>
<td>0.8</td>
<td>0.08</td>
<td>1.5</td>
<td>0.6</td>
<td>15</td>
<td>2.6</td>
</tr>
<tr>
<td>Simulation 3</td>
<td>0.25</td>
<td>0.15</td>
<td>1.25</td>
<td>0.25</td>
<td>0.15</td>
<td>1.25</td>
<td>0.4</td>
<td>6.25</td>
<td>1.6</td>
</tr>
</tbody>
</table>
simulated images which are similar to DCE-MR images revealing breast cancer, and therefore, three kinds of typical CTCs were simulated. For the tissue located outside the tumor, we simulated from a compartment model with one tissue compartment. Inside the tumor, we assumed a compartment model with two tissue compartments. For the voxels at the edge of the tumor, we simulated from a compartment model with three tissue compartments. We used different $k_{ep}$-values for the simulations, and the factor by which the $k_{ep}$-values differ in each of the three simulations is chosen differently as well. The $k_{ep}$-values differ by a factor of 7 in Simulation 1, by a factor of 10 in Simulation 2, and by a factor of 5 in Simulation 3. For all simulations, each simulated CTC consists of CA concentrations at 46 time points every 12 seconds, and we added Gaussian noise (standard deviation = 0.05) to the CTCs of all voxels. A figure showing the simulated CTCs (without noise) for Simulations 1–3 can be found in Figure 3 in the ▶Appendix.

In ▶Figure 4, some results of Simulation 1 are shown for one of the ten simulated images. The results for the remaining simulated images are similar and can be found in the ▶Appendix, as well as the results for the remaining simulations. The first image in the first row of ▶Figure 4 shows the true number of compartments $\hat{q}$ and MSE for each voxel of the image.
imated number of compartments is given for all methods under consideration. In general, when applying the Tofts model, for each voxel in an image, a compartment model with one tissue compartment is assumed. Therefore, the average number of compartments in an image is always 1 and the true average number of compartments is underestimated if the image contains voxels for which a compartment model with more than one tissue compartment is assumed, as it is the case for Simulation 1. It is obvious from the images in the first row of Figure 4 that for the boosting approaches, the estimated number of compartments across the image is spatially more homogeneous than for the elastic net approaches. This is most obvious for the tumor edge and the tissue located outside the tumor. Moreover, with the boosting approaches, the number of compartments is not as often overshooted as with the elastic net approaches, where in many cases a number of 4 or 5 compartments is estimated. In the second row of Figure 4, the voxelwise mean squared error (MSE) \( \text{MSE} = \sum \left[ C(x_i) - F_i(x_i) \right]^2 \) is plotted for all methods under consideration. For the Tofts model, it can be clearly seen that the MSE is the highest at the tumor edge, followed by the MSE for the inner region of the tumor. The MSE is the lowest for the tissue outside the tumor. This shows that especially for the voxels at the tumor edge, the Tofts model is not sufficient and more complex compartment models are needed. For the boosting approaches, the MSE is comparatively large for the voxels inside the tumor, whereas for the elastic net approaches, voxels with a comparatively large MSE bit more scattered across the image.

In Table 2, for each simulation, the average MSE for the 10 simulated images is given for each considered method. The MSE for image \( k \) is calculated according to and the average MSE over all \( K \) images according to \( \text{MSE}_{avg} = \frac{1}{K} \sum_{k=1}^{K} \text{MSE}_k \). Additionally, in Table 2, the true average number of compartments \( \hat{q}_{true} \), the estimated average number of compartments over all ten simulated images \( \hat{q}_{av} \), and the percentage \( q_c \) of voxels for which the correct number of tissue compartments is estimated is given. It can be seen that the average MSE is approximately the same for all methods except for the Tofts model, where it is considerably larger than for the other four methods. For Simulation 1, boxplots of the voxelwise MSE for all considered methods for all 10 simulated images are available in Figure 4 in the Appendix. When comparing the average MSE for the two voxelwise methods (boosting and elastic net) and the two spatial methods (boosting and elastic net) for Simulations 1–3, we can see that it is most of the times slightly larger for the elastic net approaches than for the boosting approaches. In all three simulation settings, the boosting approaches perform better than the elastic net approaches with regard to and \( q_c \), as is at most as large as with the elastic net approaches, and \( q_c \) is always larger as with the elastic net approaches. For each simulation setting, the performance of the voxelwise and the spatially regularized boosting approach regarding \( q_c \) is similar, whereas for the elastic net approaches, the spatial approach always performs considerably better regarding \( q_c \) compared to the voxelwise approach.

### 4. Application to DCE-MRI Data

#### 4.1 Description of the Data

For the clinical evaluation of our approach we used data of six breast cancer patients who have participated in a breast cancer study which has previously been reported on and which has already been analyzed [44, 45, 26, 10]. Per breast cancer patient, we used two scans recorded by a 1.5 T Siemens MAGNETOM Symphony scanner (Repetition time (TR) = 11 ms, Echo Time (TE) = 4.7 ms). One of the scans has been recorded at the beginning of the treatment (pre-treatment scan) and the second one after two weeks of chemotherapy (post-treatment scan). Each scan comprises three slices. For our analysis, we used only the central slice. Per patient and recording, the MR signal has been recorded every 12 seconds at 36 or 46 time points, respectively. At the start of the fifth MR signal recording, Gd-DTPA at a dose of 0.1 mmol per kg body weight has been injected with a power injector. The regions of interest used for the analyses cover the tumor as well as surrounding non-tumorous tissue. The smallest analyzed image consists of 48 × 81 voxels, and the largest one of 118 × 115 voxels.

#### 4.2 Results

In Fig. 5 and Figs. 34–38 in the Appendix, for both the boosting and the elastic net approaches the estimated number of compartments \( q \) is plotted for the two images of each patient from the breast cancer study. In the last row of these figures, the concentration of CA one min-
ute after the injection of the CA is plotted pre- and post-treatment as a reference. We selected the figure for patient 2 (Figure 5) for the main part of this paper, as the difference in the results for the boosting and elastic net approaches is particularly evident for this patient. However, the results for the remaining patients are of similar quality. Patient 2 is a responder to the chemotherapy [26]. Therefore, the tumor is considerably smaller in the post-treatment image than in the pre-treatment image. We can see that with the boosting approaches, the tumor can be located much better than with the elastic net approaches. This is especially obvious for the post-treatment scan of patient 2, where the tumor can not be located correctly with the spatial elastic net approaches, but with the boosting approaches it can. Furthermore, with the boosting approaches, the estimated number of compartments across the image is spatially more homogeneous, but the tumor edge is still more clearly defined than with the elastic net approaches, which is especially obvious for the pre-treatment scan of patient 2.

The BIC and the estimated average number of compartments \( \hat{q} \) for the elastic net and the boosting approaches for all analyzed images can be found in Table 3 together with the corresponding tuning parameters. For all images except for the post-treatment scan of patient 3, the BIC for the elastic net approaches is smaller than the BIC for the boosting approaches, based on which one would usually favor the elastic net approaches. However, the BIC does not reflect the spatial structure of the image and the correct localization of the tumor. When comparing the two boosting approaches presented, we see that the BIC for spatial boosting is smaller or approximately in the same range as the BIC for voxelwise boosting.

The tuning parameters \( \lambda \) and \( s \) for the elastic net approaches and \( \lambda \) for the spatially regularized boosting approach can be found in the last three rows of Table 3. Mostly, a small value of \( 10^{-10} \) or \( 10^{-7} \) is chosen for the penalization parameter \( \lambda \) in all of these approaches. An exception are the images of patients 5 and 6 and the post-treatment image of patient 4 (Figs. 36–38).
in the ▶Appendix), for which the value of the penalization parameter λ is $10^{-3}$ for the spatially regularized boosting approach. Therefore, for these five images, the biggest influence of the spatial regularization can be found, and a clear difference between the spatial structure of the number of compartments for the voxelwise and the spatially regularized boosting approach is apparent. The estimated number of compartments is distributed more homogeneously across the image for the spatially regularized approach.

### 5. Conclusions and Discussion

We proposed two boosting approaches for data-driven model choice and parameter estimation for DCE-MRI data. Both approaches build on a multi-tissue compartment model and combine the advantages of data-driven and model-based approaches, as the number of compartments is estimated from the data for each voxel of an image, and the parameters nevertheless remain interpretable. Based on the results of the simulation studies and the results of the application to DCE-MRI data of six breast-cancer patients, we conclude that the boosting approaches outperform the Tofts and elastic net approaches regarding the correct localization of the tumor as well as the spatial homogeneity of the estimated number of compartments across the image, and the definition of the tumor edge. Therefore, we favor the boosting approaches over the elastic net approaches, even though the BIC was smaller for the elastic net approaches than for the boosting approaches for most images of the breast-cancer patients.

When comparing the plots of the estimated number of compartments $\hat{q}$ (▶Figure 5 and Figs. 34–38 in the ▶Appendix), the BIC, and the estimated average number of tissue compartments $\hat{q}_{av}$ (▶Table 3) for the two boosting approaches applied to the DCE-MRI data, it appears that there is no major difference between the results of these two approaches, except for the images of patients 5 and 6 and the post-treatment image of patient 4. For these images, the number of compartments is distributed more homogeneously across the image for the spatially regularized approach, and for the images of patients 5 and 6, the BIC is considerably smaller for the spatially regularized approach compared to the voxelwise boosting approach. It was indeed a surprising result that the proposed penalization did not have a large impact on the nonlinear boosting. The voxelwise boosting approach already turned out to deter-

---

**Table 3** Results of the application to the DCE-MRI data: BIC and estimated average number of tissue compartments for all images of all six patients.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pre</td>
<td>post</td>
<td>pre</td>
<td>post</td>
<td>post</td>
<td>pre</td>
<td>post</td>
<td>post</td>
<td>post</td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 1</td>
<td>3904845</td>
<td>2971558</td>
<td>2258271</td>
<td>3314848</td>
<td>1099939</td>
<td>2103958</td>
<td>5256399</td>
<td>2821669</td>
<td>3640991</td>
</tr>
<tr>
<td>Patient 2</td>
<td>3169567</td>
<td>2290706</td>
<td>2309692</td>
<td>2138783</td>
<td>1122943</td>
<td>1928639</td>
<td>3438782</td>
<td>2130934</td>
<td>2683162</td>
</tr>
<tr>
<td>Patient 3</td>
<td>3147522</td>
<td>2275503</td>
<td>2294001</td>
<td>2131346</td>
<td>1116738</td>
<td>1912825</td>
<td>3332692</td>
<td>2111086</td>
<td>2667538</td>
</tr>
<tr>
<td>Patient 4</td>
<td>3185805</td>
<td>2304293</td>
<td>2310659</td>
<td>2143496</td>
<td>1124535</td>
<td>1879643</td>
<td>3355232</td>
<td>2148980</td>
<td>2731534</td>
</tr>
<tr>
<td>Patient 5</td>
<td>3185398</td>
<td>2304577</td>
<td>2310615</td>
<td>2143351</td>
<td>1124490</td>
<td>1879275</td>
<td>3355391</td>
<td>2144395</td>
<td>2714865</td>
</tr>
<tr>
<td>Patient 6</td>
<td>3185398</td>
<td>2304577</td>
<td>2310615</td>
<td>2143351</td>
<td>1124490</td>
<td>1879275</td>
<td>3355391</td>
<td>2144395</td>
<td>2714865</td>
</tr>
</tbody>
</table>

© Schattauer 2016
mine the number of compartments quite well and found numbers of compartments similar to the spatial elastic net approach in all of the simulations and most of the real situations. Therefore, we think that the spatial penalization could not strengthen the boosting approach in the same way it strengthened the elastic net approach. As there is also a large difference in computation time, in order to get results as quickly as possible, we would recommend to favor the voxel-wise boosting approach over the spatially regularized version. In order to get results where the spatial structure is taken into account properly, however, we would recommend to use the spatially regularized version. The computation time for the spatially regularized boosting for all 12 images was approximately 19 days on a Linux server with 64 cores and 512 GB memory using 40 cores in parallel, whereas the voxelwise boosting for all 12 images took approximately 2 hours on the same machine using 40 cores in parallel (For comparison: the computation times for the competing methods on the same machine using 40 cores in parallel were: spatial elastic net: approximately 2 hours, voxelwise elastic net: approximately 3 minutes.). For the spatially regularized boosting, the updates of the parameter estimates are done alternately for the voxels in an image, which increases the runtime considerably in comparison to the voxelwise boosting, where the parameter estimates are updated independently of each other for the voxels in an image. All code for the analyses was written in R. For the parallelization of R code, we used the packages parallel, BatchJobs and BatchExperiments [46]. For the independent parameter updates and the refit in the voxelwise boosting approach, the parallelization was done with the packages BatchJobs and BatchExperiments. For the alternating updates in the estimation process in the spatially regularized boosting approach, we used the package parallel. For the refit, we used the packages BatchJobs and BatchExperiments. The fact that we had to use a different package for the estimation process in the spatially regularized boosting approach might also contribute to the increased computation time for this approach.

In order to avoid redundancy issues in the boosting algorithm, we set a factor by which the $k_{\alpha}$-values in the multi-tissue compartment model must differ at least. We chose a factor of 5, as we know from [33] that parameters in a sum of exponentials model are highly redundant if the exponential rates differ by less than a factor of 5, and a generalization of that for convoluted exponentials is difficult to obtain. Table 2 in the Appendix shows the influence of the choice of this factor within the boosting algorithm on the average MSE ($MSE$), the estimated average number of compartments ($\hat{q}_{av}$), and the percentage of voxels for which the correct number of tissue compartments is estimated ($q_{c}$) for alternative values of 3 and 4 for Simulation 1. From this table, we see that the smaller the chosen factor, the larger the $MSE$, the larger $\hat{q}_{av}$, and the smaller $q_{c}$. However, also with factors 3 or 4, the $MSE$ is at most as large as the $MSE$ for the spatial elastic net approach (Table 2). But with factors 3 or 4, $q_{c}$ is considerably reduced compared to factor 5. The influence of the chosen factor on $q_{c}$ with the spatial boosting approach is not as strong as with the voxelwise boosting approach, which is probably due to the spatial penalization.

In order to avoid overfitting of the boosting algorithm, we used early stopping. If we run the voxelwise boosting algorithm for example on the 10 simulated images of Simulation 1 without early stopping, the estimated average number of tissue compartments $\hat{q}_{av}$ is 2.68. With early stopping, $\hat{q}_{av}$ is 2.04 (Table 2), which makes clear that without early stopping, the number of compartments is clearly overestimated.

The results of the simulation studies and the application to real world DCE-MRI data indicate that additional complexity is needed especially at tumor edges, and the Tofts model is not capable of reflecting this complexity. Using the approaches presented, the number of compartments is estimated per voxel. Thus, important information about the tissue heterogeneity is gained. This can not be done with a priori fixed model architectures.

### Acknowledgments

We thank Dr. A.R. Padhani, PSSC, Mount Vernon Hospital, Northwood, U.K., for the provision of the clinical data. Martina Feilke and Volker J. Schmid were supported by Deutsche Forschungsgemeinschaft (DFG SCHM 2747/1–1).

### References

12. Mayr A, Binder H, Gefeller O, Schmid M. Extending statistical boosting. An overview of recent...
13. Friedman J, Hastie T, Tibeshi-rani R. Additive logis-
tic regression: a statistical view of boosting (with
14. Friedman J. Greedy function approximation: A
gradient boosting machine. Ann Stat 2001; 29:
1189–1232.
15. Buhlmann P, Yu B. Boosting with the L2-L0SS:
Regression and classification. J Am Stat Assoc
16. Buhlmann P. Boosting for high-dimensional linear
17. Buhlmann P, Hothorn T. Boosting algorithms:
Regularization, prediction and model fitting. Stat
evolution of boosting algorithms. From machine
learning to statistical modelling. Methods Inf Med
19. Feilke M. dcmrboost: Boosting for DCE-MRI
data.; 2015. R package version 0.1. Available from:
20. Tofte P. Modeling tracer kinetics in dynamic Gd-
from dynamic contrast-enhanced T1-weighted MRI of
a diffusible tracer: Standardized quantities and
symbols. J Magn Reson Imaging 1999; 10:
223–232.
22. Sommer J. Regularized estimation and model se-
lection in compartment models. Dissertation at
the Faculty of Mathematics, Computer Science
and Statistics of the LMU Munich, Munich, Ger-
macy: Verlag Dr. Hut; 2013.
23. Yang X, Kopp M. Quantifying tumor vascular
heterogeneity with dynamic contrast-enhanced
magnetic resonance imaging: a review. J Biomed
GZ. A Bayesian hierarchical model for the analysis
of a longitudinal dynamic contrast-enhanced MRI
163–174.
26. Schmid VJ, Whitcher B, Padhani AR, Taylor NJ,
Yang GZ. Bayesian methods for pharmacokinetic
models in dynamic contrast-enhanced magnetic
resonance imaging. IEEE Trans Med Imaging
27. Parker G, Suckling J, Tanner S, Padhani A, Hus-
band J, Leach M. MRIV: Parametric analysis soft-
ware for contrast-enhanced dynamic MR imaging
28. Weinmann H, Lianiado M, Mittel W. Pharnma-
ketics of Gd-DTPA/dimeglumine after intra-
venous injection into healthy volunteers. Physiol
29. Hofner B, Mayr A, Robinzonov N, Schmid M.
Model-based boosting in R: a hands-on tutorial using
29: 3–35.
30. R Core Team. R: A language and environment for
statistical computing. Vienna, Austria; 2013. ISBN
3-900051-07-0. Available from: http://www.R-
project.org/.
31. Elzhov T, Mullken K, Spiess AN, Bolker B. min-
pack.lm: R interface to the Levenberg-Marquardt
nonlinear least-squares algorithm found in MIN-
PACK, plus support for bounds; 2013. R package
version 1.1.8. Available from: http://CRAN.R-
project.org/package=minpack.lm.
32. Ahearn T, Staff R, Redpath T, Semple S. The use
of the Levenberg-Marquardt curve-fitting algorithm
in pharmacokinetic modelling of DCE-MRI data.
33. Reich J. On parameter redundancy in curve fitting
of kinetic data. Kinetic data analysis: Design and
analysis of enzyme and pharmacokinetic experi-
34. Schmid M, Hothorn T. Boosting additive models
using componentwise P-splines. Comput Stat Data
35. Mayr A, Hofner B, Schmid M. The importance of
knowing when to stop. A sequential stopping rule
for component-wise gradient boosting. Methods
of statistical learning. 2nd ed. New York: Springer;
2009.
37. Hansen M, Yu B. Model selection and the principle
of minimum description length. J Am Stat Assoc
2001; 96 (454): 746 –774.
38. Candes E, Tao T. The Dantzig selector: Statistical
estimation when p is much larger than n. Ann Stat
39. Gertheiss J, Tutz G. Sparse modeling of categorial
2150–2180.
40. Tutz G, Groll A. Likelihood-based boosting in bi-
(nary and ordinal random effects models. J Comput
41. Groll A. GMMBoost: Likelihood-based Boosting
for Generalized mixed models; 2013. R package
version 1.1.2. Available from: http://CRAN.R-
project.org/package=GMMBoost.
42. Turlach B, Weingessel A. quad-prog: Functions to
solve Quadratic Programming Problems.; 2013. R
CRAN.R-project.org/package=quadprog.
43. Zou H, Hastie T. Regularization and variable selec-
tion via the elastic net. J Roy Stat Soc B 2005; 67:
301–320.
44. Ah-See ML, Makris A, Taylor N, Burcombe R,
Harrison M, Stirling J, et al. Does vascular imaging
with MRI predict response to neoadjuvant chemo-
therapy in primary breast cancer? J Clin Oncol
2004; 22 (14S): 582.
A. Statistical analysis of pharmacokinetical models
in dynamic contrast-enhanced magnetic resonon-
ance imaging. In: Duncan J, Gerig S, editors. Medi-
cal Imaging Computing and Computer-Assisted
46. Bischof B, Lang M, Mersmann O, Rahnenfuehrer
J, Wehns C. BatchJobs and BatchExperiments: Ab-
straction mechanisms for using R in batch en-