



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK



Jan Gertheiss, Julia C. Kärcher & Volker J. Schmid

Analysis of DCE-MRI Data using a Nonnegative Elastic Net

Technical Report Number 090, 2010
Department of Statistics
University of Munich

<http://www.stat.uni-muenchen.de>



Analysis of DCE-MRI Data using a Nonnegative Elastic Net

Jan Gertheiss[†], Julia C. Kärcher[†] & Volker J. Schmid^{*†}

September 29, 2010

Abstract

We present a nonnegative Elastic Net approach for the analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging data. A multi-compartment approach is considered, which is translated into a (restricted) least square model selection problem. This is done by using a set of basis functions for a given set of candidate rate constants. The form of the basis functions is derived from a kinetic model and thus describes the contribution of some compartment. Using the Elastic Net estimator, we chose clusters of basis functions, and hence, rate constants of compartments. As further challenge, the estimator has to be restricted to positive regression parameters, which correspond to transfer rates of the compartments. The proposed estimation method is applied to an *in-vivo* data set.

Keywords: Compartment Model, DCE-MRI, Elastic Net, Regularized Estimation

1 Introduction

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) provides an imaging series of contrast agent concentration in some tissue of interest. The dynamic behavior of contrast agent uptake is important for the specification of malignancy, type and grading of tumors (Parker and Padhani, 2003). Pharmacokinetic models describe the exchange of contrast agent between different, well-mixed compartments. Those compartment models provide quantitative

^{*}To whom correspondence should be addressed: volker.schmid@lmu.de.

[†]Department of Statistics, Ludwig-Maximilians-Universität Munich, Germany.

physiological parameters characterizing the amount and rate of capillary leakage (Padhani et al., 2005).

A common modeling approach is the so-called extended Tofts-Kermode model, assuming a plasma compartment and an interstitial space compartment (Tofts and Kermode, 1991). In many situations, however, this model cannot adequately describe the measured concentration (Schmid et al., 2009), which indicates that more complex models are needed. This is the case when the imaged tissue is heterogeneous as often observed in cancerous tissue. Port et al. (1999) have suggested, that a kinetic model should allow for more than one interstitial space compartment in order to adequately describe the uptake dynamics on a region of interest level. Tissue heterogeneity can, however, even be observed on a voxel level. In Kärcher and Schmid (2010), a two-compartment model has been applied to DCE-MRI data to account for heterogeneity within voxel, and parameters have been estimated using Bayesian methods.

In this paper, a multi-compartment model is fitted using likelihood based regularization techniques. We use a bundle of exponential functions as basis, each of which is derived from the differential equation describing the tracer uptake of a tissue compartment. The corresponding coefficients are sparsely selected and estimated while penalizing for an increasing number of parameters. By selecting clusters of nonzero coefficients, the number of used compartments is implicitly selected as well. A similar approach of sparse basis selection has been proposed for compartment models used in positron emission tomography (PET) (Gunn et al., 2002). The basis pursuit approach proposed there corresponds to unrestricted Lasso (Tibshirani, 1996) estimation. The Lasso, however, tends to be unstable in case of highly correlated covariates. In addition, in applications of such type, Ridge regression (Hoerl and Kennard, 1970) has often been shown to produce better results in terms of prediction accuracy. Therefore, we use a restricted Elastic Net, combining the advantages of Ridge and Lasso estimation. The estimates of an Elastic Net are known to be stable even for highly correlated predictors. In our case, however, estimated parameters need to be nonnegative to ensure the positiveness of the physiological parameters. Hence our approach is called *nonnegative* Elastic Net.

The advantage of this approach is that the number of compartments (corresponding to the number of nonzero coefficient clusters) is estimated from the data and no model choice has to be done *a priori*. In addition, the plasma volume fraction will be estimated as zero when there is no considerable contribution. Moreover, the number of compartments is not fixed, but may be varying over a field of voxel. Briefly, this approach combines the advantages of model-driven methods (parameter estimates correspond to compartmental structure and are interpretable) and data-driven methods (no *a priori* compartment-structure has to be defined).

The paper is organized as follows. In the following section, DCE-MRI data and the standard compartment model are described. Subsection 2.2 describes

the multi-compartment model and in Subsection 2.3 the proposed estimation technique – the nonnegative Elastic Net – is introduced. Finally, the proposed method is applied to the data and results are discussed in Section 3.

2 Methods

2.1 DCE-MRI Data and Compartment Models

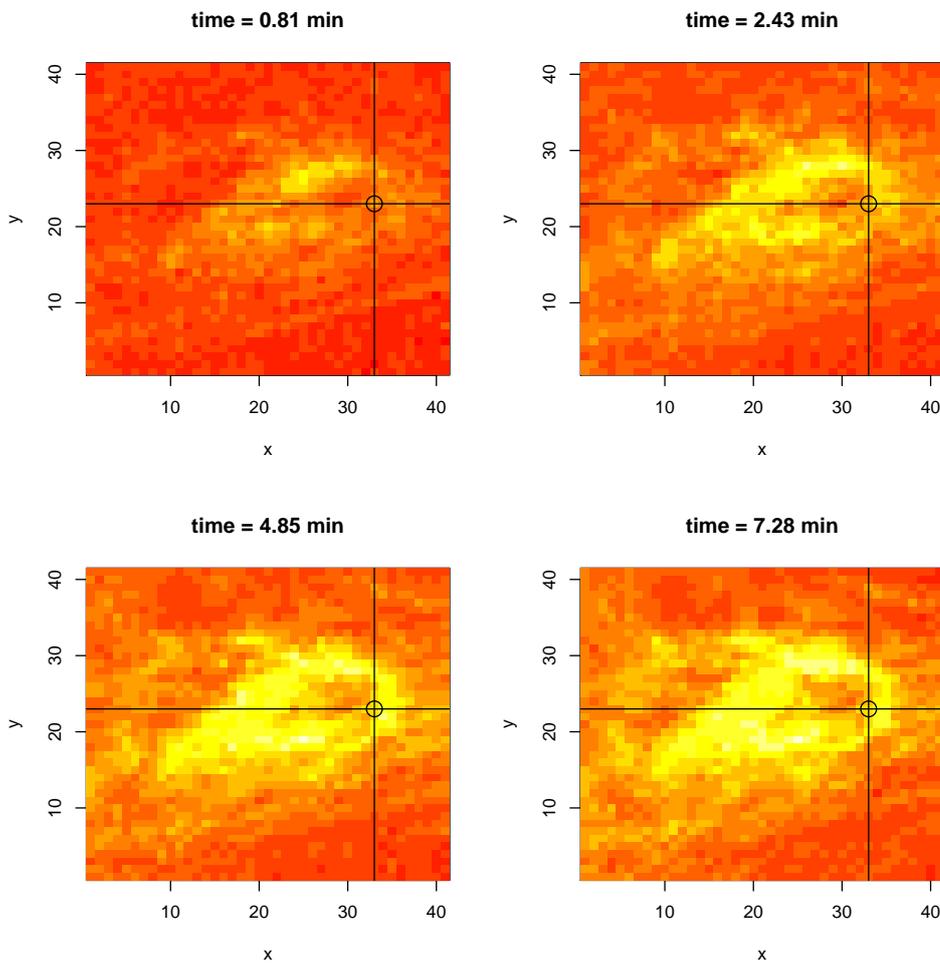


Figure 1: Color coded concentration curves $C_T(t)$ at 41×41 voxel for different time points; voxel at coordinates $\{33, 23\}$ is indicated by crosslines.

To evaluate the clinical use of our approach we use a subset of a previously analyzed DCE-MRI study on breast cancer (Schmid et al., 2006). The dataset consists of six patients with breast tumors, scanned once at the beginning of

treatment and again after six weeks. The scans were acquired with a 1.5 T Siemens MAGNETOM Symphony scanner, $TR = 11 \text{ ms}$ and $TE = 4.7 \text{ ms}$. Each scan consists of three slices of 230×256 voxel. A dose of $D = 0.1 \text{ mmol/kg}$ body weight Gd-DTPA was injected at the start of the fifth acquisition using a power injector. Regions of interest cover the tumor and surrounding normal tissue. The contrast agent concentration at time t is computed from the MR signal on each voxel (Buckley and Parker, 2005).

In Figure 1, color coded concentration maps for a 41×41 voxel region of interest for a mid tumor slice of one scan are shown for different time points. This is a pre-treatment scan, where the treatment has shutdown most of the angiogenic property, *i.e.*, the enhanced blood flow towards the tumor. Over all, concentrations (for each voxel) are reported for 46 nearly equally spaced time points from 0 to 9.1 minutes; more generally, at time points $t = 0, \dots, T$. The idea is to model these concentration time curves $C_T(t)$ separately for each voxel (as *e.g.* done by Schmid et al., 2006). For illustration, the voxel at coordinates $\{33, 23\}$ is indicated by crosslines in Figure 1.

A common modeling approach for DCE-MRI data is the (extended) Tofts-Kermode model (Tofts and Kermode, 1991) with one interstitial space compartment,

$$C_T(t) = v_p C_p(t) + C_p(t) \otimes K^{trans} \exp(-k_{ep}t), \quad (1)$$

where \otimes denotes the convolution operator and thus

$$C_p(t) \otimes \exp(-k_{ep}t) = \int_0^t C_p(t - \tau) \exp(-k_{ep}\tau) d\tau. \quad (2)$$

The arterial input function (AIF) $C_p(t)$ is assumed as fixed and parameters K^{trans} , k_{ep} and the plasma volume fraction v_p need to be estimated. The parameters do have a biological meaning: k_{ep} is the rate constant at which the interstitial space compartment exchanges with the blood plasma and K^{trans} is the corresponding volume transfer constant.

2.2 Multi Tissue Compartment Model

As a generalization of the one-compartment model (1), the multi-compartment model with q compartments can be defined by

$$C_T(t) = v_p C_p(t) + \sum_{j=1}^q C_p(t) \otimes K_j^{trans} \exp(-k_{ep_j}t). \quad (3)$$

The volume fraction of the plasma compartment is denoted by v_p . As arterial input function (AIF) we use a bi-exponential function of the form proposed by Tofts and Kermode (1991)

$$C_p(t) = D(a_1 \exp(-m_1t) + a_2 \exp(-m_2t)), \quad (4)$$

with $a_1 = 2.4 \text{ kg/l}$, $a_2 = 0.62 \text{ kg/l}$, $m_1 = 3.01 \text{ min}^{-1}$, $m_2 = 0.016 \text{ min}^{-1}$, as suggested by Fritz-Hansen et al. (1996). The constant D is the actual dosage of tracer in mol/kg .

Due to measurement error, the concentration $C_T^i(t)$ at time points $t = 0, \dots, T$ in voxel $i = 1, \dots, N$ is modeled as:

$$C_T^i(t) = v_{p,i}C_p(t) + \sum_{j=1}^q C_p(t) \otimes K_{j,i}^{trans} \exp(-k_{ep_j}t) + \epsilon_{i,t}, \quad (5)$$

where $\epsilon_{i,t} \sim N(0, \sigma_i^2)$ is a Gaussian noise term. That means

$$C_T^i(t) = v_{p,i}C_p(t) + \sum_{j=1}^q K_{j,i}^{trans} \Psi_j(t) + \epsilon_{i,t}, \quad (6)$$

with basis functions

$$\begin{aligned} \psi_j(t) &= DC_p(t) \otimes \exp(-k_{ep_j}t) \\ &= \frac{Da_1(\exp(-k_{ep_j}t) - \exp(-m_1t))}{m_1 - k_{ep_j}} + \frac{Da_2(\exp(-k_{ep_j}t) - \exp(-m_2t))}{m_2 - k_{ep_j}}. \end{aligned} \quad (7)$$

Each compartment j is characterized by how fast it exchanges with the plasma compartment, expressed by its rate constant k_{ep_j} . As candidate values we consider $\log(k_{ep_j}) \in \{-2, -1.99, -1.98, \dots, 1.99, 2\}$, and the adequate values need to be selected. Moreover, each compartment is characterized by its transfer constant K_j^{trans} . The transfer constant is obtained by the product of the volume fraction v_j and the constant rate of the compartment and is hence nonnegative: $K_j^{trans} = k_{ep_j} \cdot v_j \geq 0$. Thus, the vector $\theta_i = (v_{p,i}, K_{1,i}^{trans}, \dots, K_{q,i}^{trans})^T$ is unknown and to be estimated. Wherever the estimated transfer constant is positive ($\hat{K}_{j,i}^{trans} > 0$), the corresponding compartment, resp. k_{ep_j} value, is selected. From (6) it can be seen that a regression problem with predictors $C_p(t), \Psi_j(t), j = 1, \dots, q$, is to be solved. Figure 2 depicts the subset $\{C_p(t), \Psi_j(t)\}$ of predictors with $\log(k_{ep_j}) \in \{-2, -1.9, -1.8, \dots, 1.9, 2\}$, shown together with the observed response $C_T^i(t)$ at voxel i with coordinates $\{33, 23\}$ (blue line).

When estimating the parameter vector θ_i with simple maximum likelihood inference, under the assumption of independent Gaussian distributed observation errors $\epsilon_{i,t}$, the residual sum of squares $\sum_t (C_T^i(t) - \hat{C}_T^i(t))^2$ has to be minimized. However, here the θ_i with $\theta = K_i^{trans}$ for $i \geq 1$ and $\theta_0 = v_p$ need to be nonnegative, and, hence, the pure ML-estimate is

$$\hat{\theta}_i^{ML} = \operatorname{argmin}_{\theta \geq 0} \left\{ \sum_t (C_T^i(t) - z(t)^T \theta)^2 \right\}, \quad (8)$$

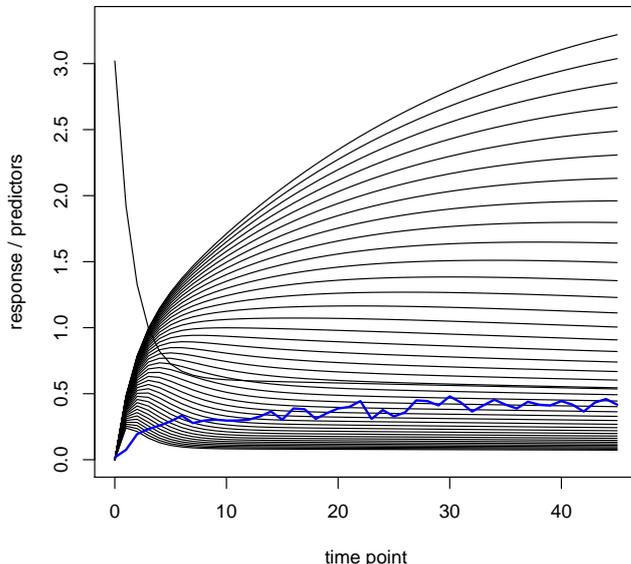


Figure 2: Subset of predictors $\{C_p(t), \Psi_j(t)\}$ (black lines), together with the observed response $C_T^i(t)$ at voxel i with coordinates $\{33, 23\}$ (blue line).

with

$$z(t) = D(C_p(t), \Psi_1(t), \dots, \Psi_q(t))^T, \quad t = 1, \dots, T.$$

Since, however, we have a large number of θ parameters and adjacent entries of $z(t)$ are highly correlated (due to construction), pure maximum likelihood estimates are unstable or even not unique. Therefore we use a penalized approach as described in the following.

2.3 Regularized Estimation

In order to stabilize the estimation of parameter vector θ , the log-likelihood which is to be maximized is additively corrected by a penalty term $J(\theta)$. More precisely, for a fixed voxel i , we use the estimator

$$\hat{\theta}_i = \operatorname{argmin}_{\theta \geq 0} \left\{ \sum_t (C_T^i(t) - z(t)^T \theta)^2 + \lambda J(\theta) \right\}, \quad (9)$$

with $z(t)$ as given in (8). The strength of penalization is controlled by λ . The crucial point, however, is to choose an appropriate penalty $J(\theta)$. Vega-Hernandez et al. (2008), for example, discussed the use of different penalties for solving the so-called inverse problem of the electroencephalography (EEG) in neuroscience.

A well established regularization technique which in particular was constructed for high-dimensional problems with highly correlated explanatory variables (as found in z) is the so-called Elastic Net (Zou and Hastie, 2005), with penalty

$$J(\theta) = \alpha \sum_{j=1}^q \theta_j^2 + (1 - \alpha) \sum_{j=1}^q |\theta_j|. \quad (10)$$

Alternatively, the corresponding estimate can be written as

$$\hat{\theta}_i = \operatorname{argmin}_{\theta \geq 0} \left\{ \sum_t (C_T^i(t) - z(t)^T \theta)^2 + \lambda \sum_{j=1}^q \theta_j^2 + \gamma \sum_{j=1}^q |\theta_j| \right\},$$

or

$$\hat{\theta}_i = \operatorname{argmin}_{\theta \geq 0} \left\{ \sum_t (C_T^i(t) - z(t)^T \theta)^2 + \lambda \sum_{j=1}^q \theta_j^2 \right\},$$

subject to $\sum_{j=1}^q |\theta_j| \leq s. \quad (11)$

Due to the L_1 -type penalty term in (11), coefficients from $\{\hat{\theta}_{i1}, \dots, \hat{\theta}_{iq}\} = \{\hat{K}_{1,i}^{trans}, \dots, \hat{K}_{q,i}^{trans}\}$ may be set to zero (see *e.g.* Zou and Hastie, 2005), which means that corresponding k_{ep_j} are excluded. Remaining compartments (with $\hat{K}_{j,i}^{trans} > 0$) are selected.

For practical estimation the R package `quadprog` (Turlach, 2009) is used. Tuning parameters λ and s can, for example, be determined using K-fold cross-validation; for an introduction to cross-validation see, *e.g.*, Hastie et al. (2009). Moreover, entries of z are scaled to have unit variance over time, because otherwise θ_j corresponding to entries of z with smaller variance would implicitly undergo higher penalization.

3 Results

For illustration, the contrast agent concentration curve in voxel $\{33, 23\}$ was fitted and estimated K^{trans} values are shown in Figure 3 for $s = 1/8$ and $\lambda = 10^{-5}$ (top) or $\lambda = 10^{-2}$ (bottom). With smaller λ values a sparser solution is obtained. In the extreme case $\lambda = 0$ the (positive) Lasso (Tibshirani, 1996) results, which, however, is quite unstable for the given data. With $\lambda = 10^{-5}$ and $\lambda = 10^{-2}$ the quality of the fit is very similar, as shown in Figure 4.

In order to find adequate λ and s values for each voxel, 5-fold cross-validation can be done. However, we prefer sparse solutions, *i.e.*, solutions with just a few $\hat{K}_{j,i}^{trans} > 0$, and we observed that cross-validation scores just slightly changed for

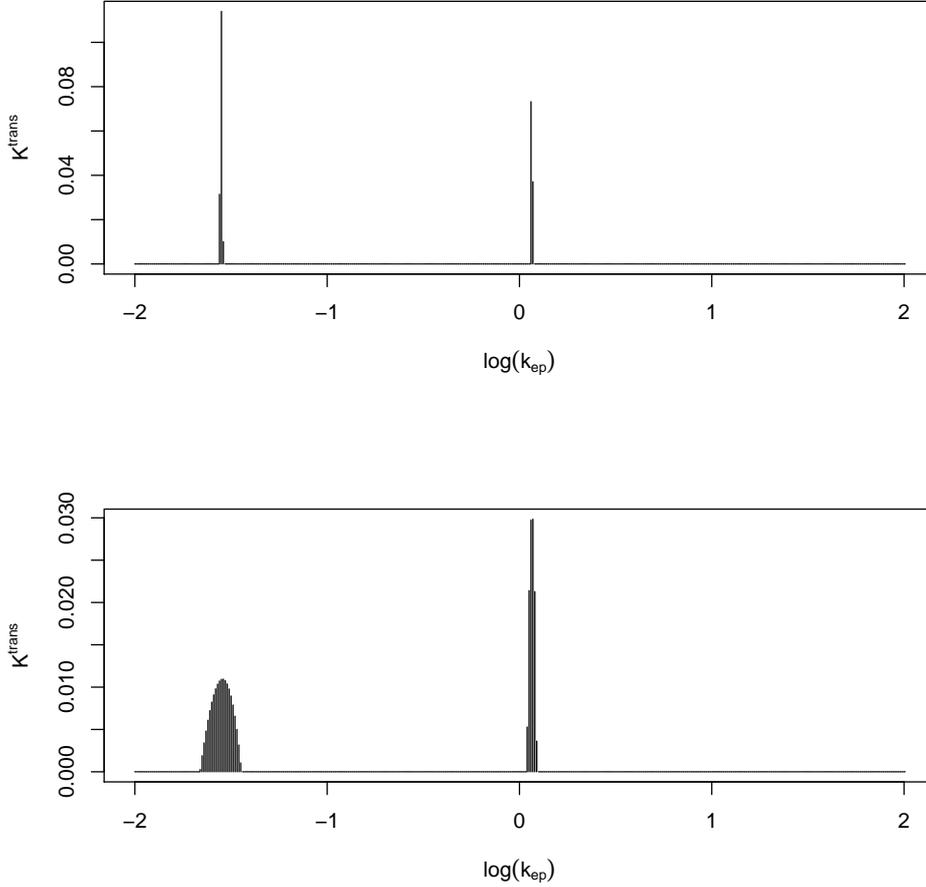


Figure 3: Estimated K^{trans} values for voxel $\{33, 23\}$, if $s = 1/8$ and $\lambda = 10^{-5}$ (top) or $\lambda = 10^{-2}$ (bottom) are chosen.

different small λ values. So we fixed $\lambda = 10^{-8}$, and only s values were chosen via (5-fold) cross-validation.

For four voxel ($\{10, 5\}, \{18, 27\}, \{29, 30\}, \{33, 23\}$) observed (black points) and fitted (solid red lines) concentration curves are shown in Figure 5. For comparison, the fitted concentration curves for the standard one-compartment model are shown, too (dashed black). In case of voxel $\{29, 30\}$ the results of both models are equal. For the other voxels (especially $\{10, 5\}$ and $\{30, 23\}$) the fast uptake is more adequately described by the multi-compartment model. However, for voxel $\{18, 27\}$, it is unclear whether this feature always results in a better overall fit.

In Figure 6 the estimated variances of error terms $\epsilon_{i,t}$ are visualized. We see that variances tend to be larger at the boundary of the tumor. The observation error $\epsilon_{i,t}$ has been assumed to be independent over time. In order to verify

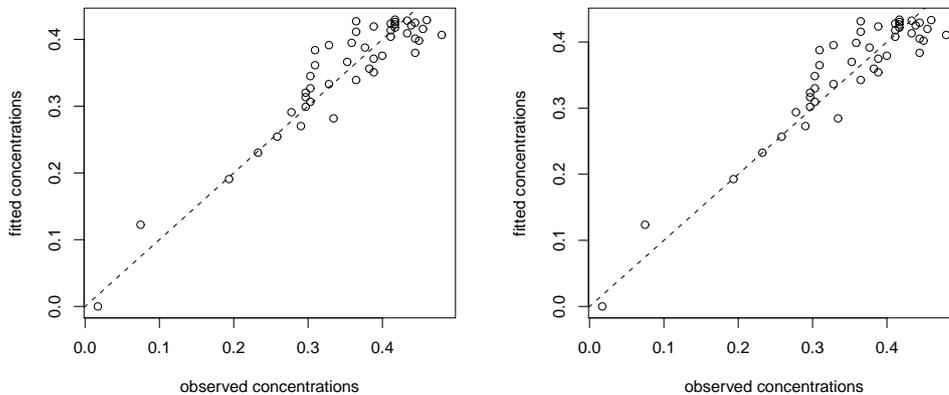


Figure 4: Fitted against observed concentration $C_T(t)$ for voxel $\{33, 23\}$, if $s = 1/8$ and $\lambda = 10^{-5}$ (left) or $\lambda = 10^{-2}$ (right) are chosen.

this assumption observed (first-order) autocorrelations are summarized in Figure 7. From the left panel we can see that autocorrelation values are modest and the distribution is rather symmetric and centered at zero, which supports the assumption of independent errors. Moreover, no spatial pattern of correlation coefficients is observed (right panel).

In Figure 8 a map of the number of compartments selected by the algorithm is shown. A one-compartment model is indicated in red, two compartments in orange, and three or more in white. The number of compartments is defined as the number of disjoint clusters of estimated nonzero K^{trans} values for the respective voxel. The model fitted to the concentration curve of voxel $\{33, 23\}$, for example, is a two-compartment model (see also Figure 3). The number of compartments is increased towards the upper left of the region of interest, which in this case roughly corresponds to a region with relatively healthy tissue.

4 Summary and Discussion

In this paper we presented a nonnegative Elastic Net estimator for fitting compartmental models. The proposed estimation approach allows to estimate the number of compartments along with the kinetic parameters.

The proposed approach uses the basis pursuit approach by Gunn et al. (2002), but accounts for the high correlation of the basis functions. In contrast to the Lasso, the Elastic Net allows to chose clusters of basis functions, and hence, is more stable for correlated basis functions. As additional challenge, the algorithm has to account for the fact that the estimated parameters are rates and therefore have to be non-negative.

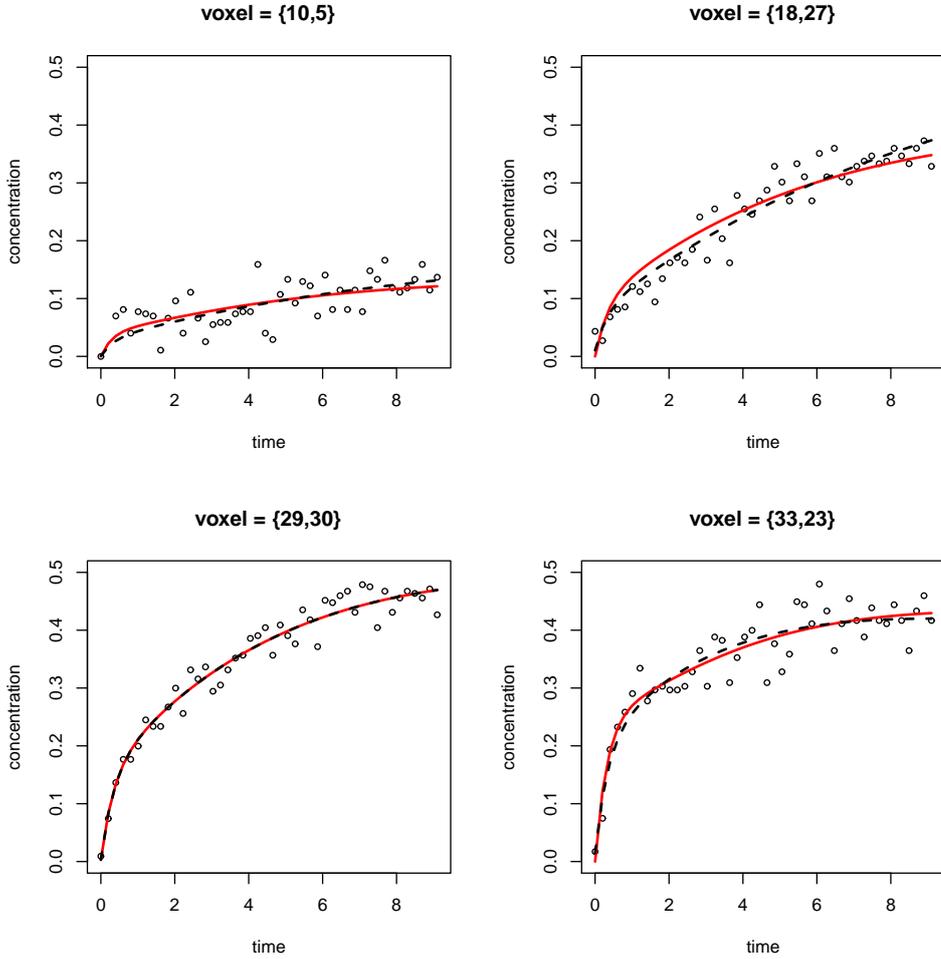


Figure 5: Observed (black points) and fitted concentration curves for four voxel if the multi-compartment model (solid red line) or one-compartment model (dashed black line) is applied.

The proposed algorithm is able to give better fits to the observed contrast agent concentration. However, the map of estimated compartments is not completely conclusive. The number of compartments in a pixel can be an indicator for the heterogeneity of the tissue, however, a higher number of compartments can also be caused by incorrect definition of the input function. Here, we used a literature AIF, which may not be adequate for the data. Hence, further investigations on the causes of the number of compartments will be necessary.

The proposed approach is however not limited to DCE-MRI data, but can be used for all applications, where kinetic models with more than one compartment are considered. This is the case for a variety of applications in biology and medicine.

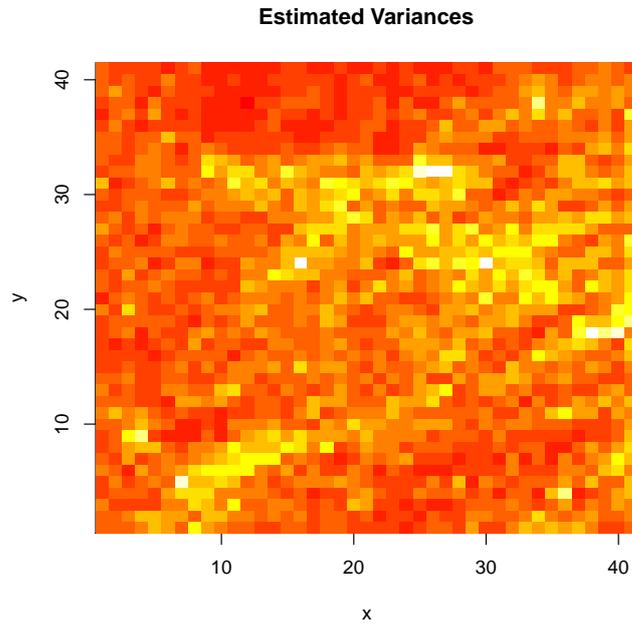


Figure 6: Spatial pattern of estimated variances of error terms $\epsilon_{i,t}$.

References

- Buckley, D. L. and G. J. M. Parker (2005). Measuring Contrast Agent Concentration in T 1 -Weighted Dynamic Contrast-Enhanced MRI. In A. Jackson, D. Buckley, and G. Parker (Eds.), *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*, Chapter 5, pp. 69–79. Springer.
- Fritz-Hansen, T., E. Rostrup, K. B. Larsson, L. Sondergaard, P. Ring, and O. Henriksen (1996). Measurement of the arterial concentration of Gd-DTPA using MRI: A step toward quantitative perfusion imaging. *Magnetic Resonance in Medicine* 36, 225–231.
- Gunn, R. N., S. R. Gunn, F. E. Turkheimer, J. A. D. Aston, and V. J. Cunningham (2002). Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. *Journal of Cerebral Blood Flow and Metabolism* 22, 1425–1439.
- Hastie, T., R. Tibshirani, and J. H. Friedman (2009). *The Elements of Statistical Learning* (2nd ed.). New York: Springer.
- Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67.

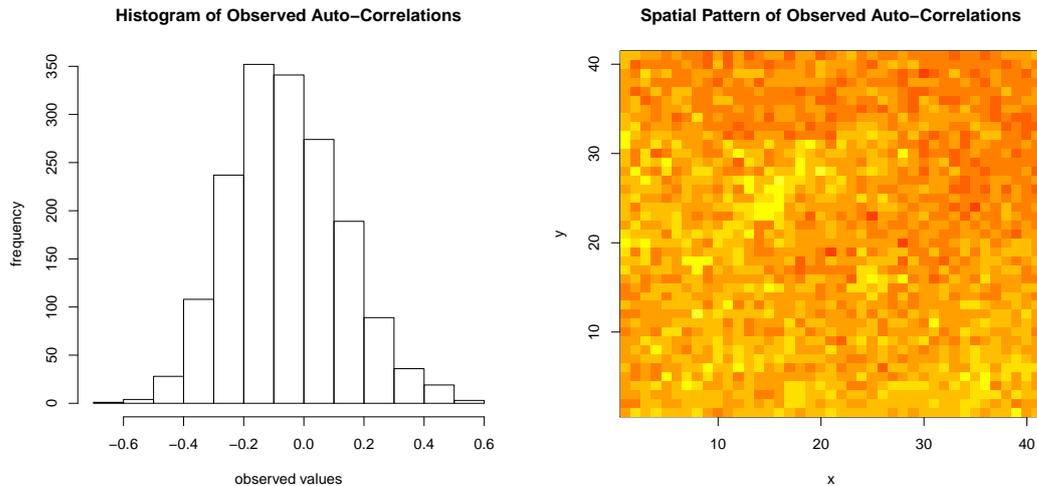


Figure 7: Histogram of observed (first-order) autocorrelations (left) and spatial pattern (right).

- Kärcher, J. C. and V. J. Schmid (2010). Two Tissue Compartment Model in DCE-MRI: A Bayesian Approach. In *IEEE International Symposium on Biomedical Imaging. From Nano to Macro*, Number 3, pp. 724–727.
- Padhani, A. R., M.-L. W. Ah-See, and A. Makris (2005). MRI in the detection and management of breast cancer. *Expert Reviews in Anticancer Therapy* 5(2), 239–252.
- Parker, G. J. M. and A. R. Padhani (2003). T1-w dce-mri: T1-weighted dynamic contrast-enhanced mri. In P. Tofts (Ed.), *Quantitative MRI of the Brain*, pp. 341–364. Chichester, England: Wiley.
- Port, R. E., M. V. Knopp, U. Hoffmann, S. Milker-Zabel, and G. Brix (1999). Multicompartment analysis of gadolinium chelate kinetics: blood-tissue exchange in mammary tumors as monitored by dynamic MR imaging. *Journal of Magnetic Resonance Imaging* 10, 233–241.
- Schmid, V. J., B. Whitcher, A. R. Padhani, N. J. Taylor, and G.-Z. Yang (2006). Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Transactions on Medical Imaging* 25(12), 1627–1636.
- Schmid, V. J., B. Whitcher, A. R. Padhani, N. J. Taylor, and G.-Z. Yang (2009). A bayesian hierarchical model for the analysis of a longitudinal dynamic contrast-enhanced MRI oncology study. *Magnetic Resonance in Medicine* 61(1), 163–174.

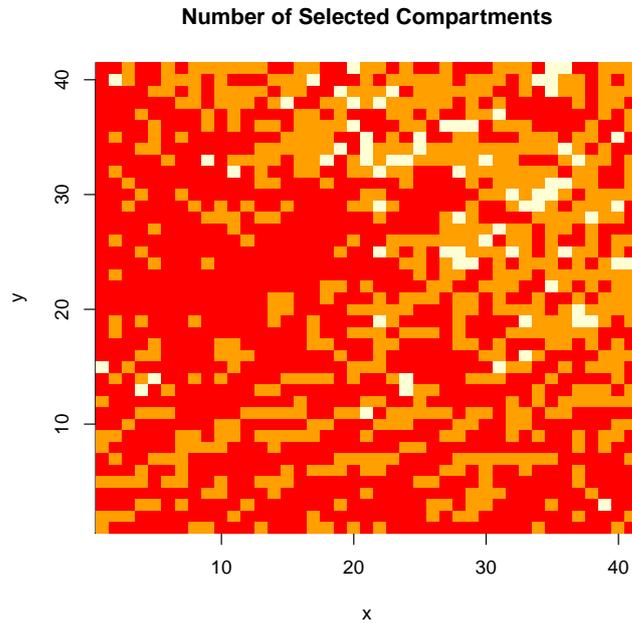


Figure 8: Spatial pattern of the number of selected compartments, where 'red' means one compartment, 'orange' two compartments, and 'white' three or more.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 58, 267–288.

Tofts, P. S. and A. G. Kermode (1991). Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging - 1. Fundamental concepts. *Magnetic Resonance in Medicine* 17, 357–367.

Turlach, B. A. (2009). *quadprog: Functions to solve Quadratic Programming Problems*.

Vega-Hernandez, M., E. Martinez-Montes, J. M. Sanchez-Bornot, A. Lage-Castellanos, and P. A. Valdes-Sosa (2008). Penalized Least Squares Methods for Solving the EEG Inverse Problem. *Statistica Sinica* 18, 1535–1551.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B* 67, 301–320.