LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

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

Spatially regularized estimation for the analysis of DCE-MRI data

Technical Report Number 132, 2012
Department of Statistics
University of Munich

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

Spatially regularized estimation for the analysis of DCE-MRI data

Julia C. Sommer[†], Jan Gertheiss[†] & Volker J. Schmid[*†]

October 10, 2012

Abstract

Competing compartment models of different complexities have been used for the quantitative analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging data. We present a spatial Elastic Net approach that allows to estimate the number of compartments for each voxel such that the model complexity is not fixed *a priori*. 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 a specific compartment. Using a spatial Elastic Net estimator, we chose a sparse set of basis functions per voxel, and hence, rate constants of compartments. The spatial penalty takes into account the voxel structure of an image and performs better than a penalty treating voxels independently. The proposed estimation method is evaluated for simulated images and applied to an *in-vivo* data set.

Keywords: DCE-MRI; Elastic Net; Model Selection; Multi-compartment Model; Spatially Penalized Estimation

1 Introduction

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) provides an imaging series of contrast agent (CA) concentration in some tissue of interest. For example, Fig. 1 depicts parts of a DCE-MRI series of a breast cancer patient. After injection of a CA, the CA concentration at each time point is computed from the MR signal on each voxel (Buckley and Parker, 2005). Thus, CA perfusion

[*]To whom correspondence should be addressed: `volker.schmid@lmu.de`.

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

in the tissue is recorded. The dynamic behavior of CA uptake allows to specify malignancy, type and grading of tumors and to assess therapy success (Padhani et al., 2005).

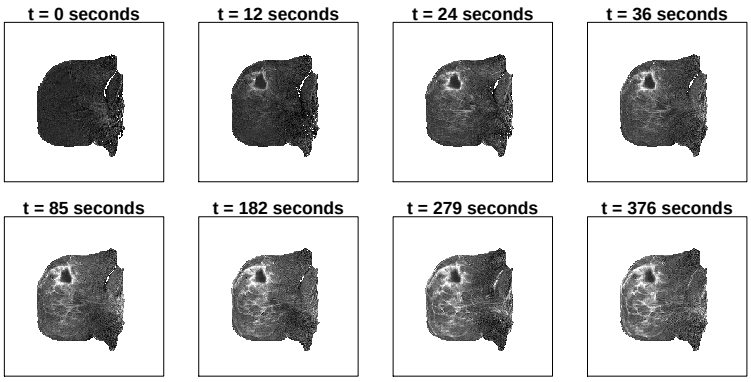

Figure 1: DCE-MRI scan of a breast cancer patient showing the contrast agent concentration over time

Quantitative characterization of CA uptake can be done with model-driven methods or with data-driven methods. With data-driven methods like nonparametric regression no *a priori* compartment-structure has to be defined (Schmid et al., 2009). Model-driven methods are based on pharmacokinetic models describing the exchange of CA between different, well-mixed compartments (Tofts and Kermode, 1991). Such compartment models provide quantitative physiological parameters characterizing the amount and rate of capillary leakage (Padhani et al., 2005).

However, several compartmental models with different, *a priori* fixed, numbers of compartments have been proposed and it remains unclear which model to use. This is particularly the case when the imaged tissue is heterogeneous as often observed in cancerous tissue. Therefore, several authors propose more complex models to describe perfusion in tissue. For example, the two compartment exchange model (2CXM) has separate compartments for arterial plasma and interstitial plasma (Brix et al., 2009; Sourbron and Buckley, 2011). Multi-compartment models allow for two to three kinetically distinct tissue compartments to describe CTCs on a region of interest level (Port et al., 1999); and in Kärcher and Schmid (2010), a hierarchical Bayesian two-compartment model has been applied to DCE-MRI on a voxel level. However, the adequate number of compartments might be different in different types of tissue and—as cancerous tissue is often heterogeneous—it might even vary over a field of voxels. That is why, in this paper, we aim to estimate the number of compartments per voxel from the data. This is important as the degree of tissue heterogeneity itself is diagnostically informative.

To this end, a multi-compartment model is fitted using likelihood based regularization techniques. We use a bundle of exponential basis functions, each of

which is derived from the differential equation describing the tracer uptake of a tissue compartment. Like this, corresponding coefficients remain interpretable and linked to physiological parameters. The 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. With this approach we combine the advantages of model-driven and data-driven methods, *i*.e. , a good fit to the data and biologically interpretable parameter.

An approach of sparse basis selection—not considering the spatial image structure though—has also 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. However, compared to the Lasso, Ridge regression (Hoerl and Kennard, 1970) has often been shown to produce better results in case of highly correlated covariates. Therefore, we use a restricted Elastic Net, combining the advantages of Ridge and Lasso estimation (Zou and Hastie, 2005).

Spatial information is frequently used in image processing, most prominently in neuroimaging (Gössl et al., 2001; Zou et al., 2004; Christensen and Yetkin, 2005). For DCE-MRI, several authors have proposed Bayesian hierarchical models (assuming a fixed number of compartments) to account for the spatial structure intrinsic in an image (Schmid et al., 2006; Kelm et al., 2009; Sommer and Schmid, 2012).

For our approach we need to (1) restrict the parameters to be non-negative to ensure the positiveness of the physiological parameters; (2) we include prior spatial information by assuming spatial smoothness of the parameters. We do so by penalizing quadratic differences of neighboring coefficients. With this spatially penalized maximum likelihood (ML) approach we use the intrinsic spatial information given by the voxel structure of the image.

The paper is organized as follows. In Section 2, we introduce a multi-compartment model as a generalization of commonly used compartment models. In Subsection 2.3 and 2.4 the proposed estimation techniques—the voxelwise Elastic Net and spatial Elastic Net—are introduced. Then, the proposed approach is evaluated for simulated data (Section 3) as well as for *in vivo* data from a breast cancer study (Section 4). Section 5 concludes.

2 Methods

2.1 Compartment Models

Several compartmental architectures of different complexities have been proposed for the analysis of DCE-MRI data. A common model is the so-called "extended" Tofts model, assuming a plasma compartment and an interstitial space compart-

ment (Tofts and Kermode, 1991),

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

where $C_T(t)$ is the CA concentration of interest at time point $t = 1, \ldots, T$; $*$ denotes the convolution operator, that is,

$$C_p(t) * \exp(-k_{\text{ep}} t) = \int_0^t C_p(t-\tau) \exp(-k_{\text{ep}} \tau) d\tau. \tag{2}$$

The arterial input function (AIF) $C_p(t)$ is assumed to be known 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. In the original, "simple" Tofts model, no additional plasma contribution is assumed (*i.e.*, $v_p = 0$).

2.2 Multi Tissue Compartment Model

As a generalization of the model with one tissue compartment (1) and its extensions described above, the multi-compartment model with q tissue compartments can be defined by

$$C_T(t) = v_p C_p(t) + \sum_{k=1}^{q} C_p(t) * K_k^{\text{trans}} \exp(-k_{\text{ep}_k} t). \tag{3}$$

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)

$$C_p(t) = D(a_1 \exp(-m_1 t) + a_2 \exp(-m_2 t)), \tag{4}$$

with $a_1 = 3.99\ kg/l$, $a_2 = 4.78\ kg/l$, $m_1 = 0.144\ min^{-1}$, $m_2 = 0.0111\ min^{-1}$. 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 $t = 0, \ldots, T$ in voxel $i = 1, \ldots, N$ is modeled as:

$$C_T^i(t) = v_{p,i} C_p(t) + \sum_{k=1}^{q} K_{k,i}^{\text{trans}} \Psi_k(t) + \epsilon_{i,t}, \tag{5}$$

where $\epsilon_{i,t} \sim N(0, \sigma_i^2)$ is a Gaussian noise term and $\Psi_k(t) = C_p(t) * \exp(-k_{\text{ep}_k} t)$ are basis functions. Each tissue compartment k is characterized by how fast it exchanges with the plasma compartment, expressed by its rate constant k_{ep_k}. As candidate values we consider $\log(k_{\text{ep}_k}) \in \{-3, -2.9, -2.8, \ldots, 2.9, 3\}$, and the adequate values need to be selected. Moreover, each compartment is characterized by its transfer constant K_k^{trans}. The transfer constant is obtained by the

product of the volume fraction v_k and the rate constant of the compartment and is hence non-negative: $K_k^{\text{trans}} = k_{\text{ep}_k} \cdot v_k \geq 0$. Thus, for voxel i, the vector $\theta_i = (\theta_{i,0}, \ldots, \theta_{i,q})^T = (v_{p,i}, K_{i,1}^{\text{trans}}, \ldots, K_{i,q}^{\text{trans}})^T$ is unknown and has to be estimated. Wherever the estimated transfer constant is positive ($\hat{K}_{i,k}^{\text{trans}} > 0$), the corresponding compartment, resp. k_{ep_k} value, is selected. From (5) it can be seen that a regression problem with predictors $C_p(t), \Psi_k(t)$, $k = 1, \ldots, q$, is to be solved. The Appendix depicts a subset of those predictors.

When estimating the parameter vector θ_i with simple ML 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 need to be non-negative, and, hence, the pure ML-estimate is

$$\hat{\theta}_i^{\text{ML}} = \text{argmin}_{\theta_i \geq 0} \left\{ \sum_t \left(C_T^i(t) - z(t)^T \theta_i \right)^2 \right\}, \tag{6}$$

with

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

Since, however, we have a large number of θ_i parameters and adjacent entries of $z(t)$ are highly correlated (due to construction), pure ML-estimates are unstable or even not unique. Therefore, we use two penalized approaches.

2.3 Voxelwise Regularized Estimation

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

$$\hat{\theta}_i = \text{argmin}_{\theta_i \geq 0} \left\{ \sum_t \left(C_T^i(t) - z(t)^T \theta_i \right)^2 + \lambda J(\theta_i) \right\}, \tag{7}$$

with $z(t)$ as given in (6). The strength of penalization is controlled by λ. The crucial point, however, is to choose an appropriate penalty $J(\theta_i)$. For example, Vega-Hernandez et al. (2008) 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 was constructed in particular 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

$$J(\theta_i) = \alpha \sum_{k=0}^{q} \theta_{i,k}^2 + (1 - \alpha) \sum_{k=0}^{q} |\theta_{i,k}|. \tag{8}$$

Alternatively, the corresponding estimate can be written as

$$\hat{\theta}_i = \text{argmin}_{\theta_i \geq 0} \left\{ \sum_t \left(C_T^i(t) - z(t)^T \theta_i \right)^2 + \lambda \sum_{k=0}^{q} \theta_{i,k}^2 \right\},$$
$$\text{subject to } \sum_{k=0}^{q} |\theta_{i,k}| \leq s. \tag{9}$$

This optimization problem can be solved using quadratic programming methods, for example, the R package `quadprog` (Turlach and Weingessel, 2011). Before running the optimization algorithm, entries of z are scaled to have unit variance over time, because otherwise θ_k corresponding to entries of z with smaller variance would implicitly undergo higher penalization.

Seeking for a sparse solution in a high dimensional predictor space, we use a two-stage estimation procedure that separates model selection from parameter shrinkage similar to the relaxed Lasso (Meinshausen, 2007). In a first step, a set of basis functions is selected with the aid of restrictive L_1- and L_2-penalties. Due to the L_1-type penalty term in (9), coefficients from $\{\hat{\theta}_{i,0}, \ldots, \hat{\theta}_{i,q}\} = \{\hat{v_{p,i}}, \hat{K}_{i,1}^{\text{trans}}, \ldots, \hat{K}_{i,q}^{\text{trans}}\}$ may be set to zero (see *e.g.* Zou and Hastie (2005)), which means that corresponding arterial plasma or tissue compartments are excluded. As the basis functions are highly correlated non-vanishing coefficients mostly appear in clusters. From each cluster only the basis function with maximal contribution is selected to obtain a sparse basis. Like this, the selected predictors correspond to compartments the exchange rates of which are different enough to be biologically meaningful. In a second step, we refit the sparse model without penalization (pure ML-estimates).

Tuning parameters λ and s can, for example, be determined using the Bayesian information criterion (BIC) as described in Hastie et al. (2009). For given (λ, s) the BIC can be calculated as

$$\text{BIC}_{(\lambda,s)} = -2 \log(L(C_T, \hat{\theta}_{(\lambda,s)})) + \log(N \cdot T) \cdot (p).$$

Here, $p = \hat{q} + \mathbf{1}(\hat{v}_p > 0)$ corresponds to the number of selected predictors and $L(C_T, \hat{\theta}_{(\lambda,s)})$ is the Likelihood of the observed concentration C_T evaluated at the Elastic Net estimates (9) of all voxels.

2.4 Spatially Regularized Estimation

So far, voxels have been treated separately, and for each voxel i parameters θ_i have been fitted independently of each other. Since, however, there is some spatial structure across voxels, parameters $\theta = (\theta_1, \ldots, \theta_N)^T$ from all voxels should be estimated taking this structure into account. We assume a neighborhood structure where adjacent voxels are neighbors. That is, each voxel has four direct

neighbors unless it lies at the edge of the image. Therefore, parameters for voxel i are now estimated by

$$\hat{\theta}_i = \text{argmin}_{\theta_i \geq 0} \left\{ \sum_t \left(C_T^i(t) - z(t)^T \theta_i \right)^2 + \lambda J_i(\theta) \right\}, \tag{10}$$

with penalty

$$J_i(\theta) = \alpha Q_i(\theta) + (1-\alpha) \sum_{k=0}^{q} |\theta_{i,k}|. \tag{11}$$

Similar to the voxelwise regularization, the L_1-type penalty term only penalizes parameters of voxel i and favors some of the coefficients to be set to zero. The quadratic penalty term Q_i, however, enforces some spatial smoothness of parameters θ by penalizing differences of neighboring θ_j:

$$Q_i(\theta) \;=\; \sum_{j \in \partial(i)} \|\theta_i - \theta_j\|^2 = \sum_{j \in \partial(i)} \sum_{k=0}^{q} \left(\theta_{i,k} - \theta_{j,k}\right)^2.$$

Here, $\partial(i)$ is the set of voxels that are direct neighbors of voxel i. Note that the penalty terms J_i and Q_i do depend only on parameters corresponding to neighboring voxels: $Q_i(\theta) = Q(\theta_i, \theta_{\partial(i)})$ and $J_i(\theta) = J(\theta_i, \theta_{\partial(i)})$ with $\theta_{\partial(i)} = (\theta_j)_{j \in \partial(i)}$.

Computational Issues

For the computation of the spatially regularized estimates, we introduce pseudo-observations; see also Zou and Hastie (2005). More precisely, for voxel i, we define the vector of "response values"

$$y^i = (C_T^i(0), \ldots, C_T^i(T), \sqrt{\lambda} \cdot \xi^i)^T,$$

where $\xi^i = \left(\theta_j \right)_{j \in \partial(i)}$ are the parameter values of neighboring voxels used as pseudo-observations. The design matrix

$$X = \begin{pmatrix} Z \\ \sqrt{\lambda} \cdot D_i \end{pmatrix}$$

consists of

$$Z = \begin{pmatrix} z(0)^T \\ z(1)^T \\ \vdots \\ z(T)^T \end{pmatrix}$$

and

$$D_i = \left.\begin{pmatrix} I \\ \vdots \\ I \end{pmatrix}\right\} |\partial(i)| \text{ times}$$

where I is the identity matrix of dimension $(q+1) \times (q+1)$. Thus, we have

$$\hat{\theta}_i = \text{argmin}_{\theta_i \geq 0} \left\{ (y^i - X\theta_i)^T (y^i - X\theta_i) \right\},$$
$$\text{subject to } \sum_{k=0}^{q} |\theta_{i,k}| \leq s.$$

Since parameters for neighboring voxels, $\theta_{\partial(i)}$, are unknown but needed for the calculation of ξ^i, we plug in current estimates $\hat{\theta}_{\partial(i)}$ and iterate this procedure over all voxels until convergence. For efficiency we use a parallel update of voxels following a checkerboard pattern that uses conditional independence from other voxels given all neighboring voxels. As starting values we use the estimates which describe the mean concentration time curve (CTC) over all voxels. Predictors are selected using the same two-step estimation procedure as for the voxelwise regularized estimation.

3 Simulation Study

3.1 Simulation Setup

To evaluate the proposed voxelwise and spatial Elastic Net, we simulated a DCE-MR image of 75×75 voxels. We chose different parameter combinations of the multi-compartment model with blocks simulated from one, two or three compartments ($q = 1, 2, 3$) with or without contribution of the plasma compartment ($v_p = 0.1$ or $v_p = 0$). In the left block, the true underlying parameters for the CTCs are $k_{\text{ep}_1} = 0.2$ and $K_1^{\text{trans}} = 0.2$; in the central block $k_{\text{ep}_1} = 0.2$, $k_{\text{ep}_2} = 4$, $K_1^{\text{trans}} = 0.1$, and $K_2^{\text{trans}} = 2$; in the right block $k_{\text{ep}_1} = 0.2$, $k_{\text{ep}_2} = 1$, $k_{\text{ep}_3} = 4$, $K_1^{\text{trans}} = 0.07$, $K_2^{\text{trans}} = 0.3$, and $K_3^{\text{trans}} = 1$. Gaussian noise was added to the CTCs for each voxel with standard deviation $\sigma = 0.05$. The Appendix shows the simulated CTCs and gives a more detailed sketch of the simulation design. The optimal penalization parameters λ^* and s^* for the proposed methods were chosen corresponding to lowest BIC values.

3.2 Results

Fig. 2 summarizes the model selection for the simulated image comparing the extended Tofts model and the voxelwise and spatially regularized multi-compartment models; it shows the number q of selected basis functions Ψ_k and if the AIF C_p

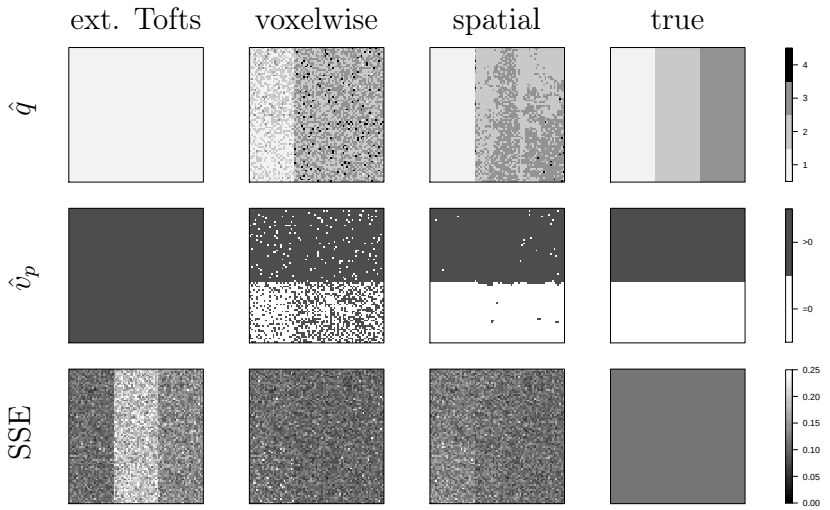

Figure 2: Results of simulation study: estimated number of tissue compartments $\hat{q}$, estimated inclusion of plasma compartment $\hat{v}_p$ and sum of squared errors (SSE)

was included as predictor. The extended Tofts model has a fixed number of compartments (one tissue and one plasma compartment) for all voxels. With the voxelwise and spatial Elastic Net one to four tissue compartments ($\hat{q}$) are selected as well as—optionally—a plasma compartment ($\hat{v}_p$). The spatial regularization helps to select the number of parameters more sparsely compared to the voxelwise model. The estimated model complexity corresponds very well to the true underlying model.

Fig. 2 also depicts the sum of squared errors (SSE) per voxel. For the voxelwise and spatial Elastic Net the SSE is at a similar level. In the left block, the spatial Elastic Net has a slightly increased SSE. For the extended Tofts model the SSE is much higher in the second and third blocks, as the Tofts model does not account for the contribution of additional tissue compartments. This is why both restricted multi-compartment models have considerably lower BIC values compared to the extended Tofts model. In Table 1 the BIC as well as the average number of selected coefficients per voxel is shown. The BIC is lowest for the spatial Elastic Net.

Let's now discuss the model selection with the spatial Elastic Net in more detail. Fig. 3 depicts the results of the spatial Elastic Net for three of the simulated voxels: one from the left, center and right block respectively. For the voxel from the left block ($q = 1$) only one basis function is selected and the true estimated coefficient is close to the true underlying coefficient (Fig. 3). The number of compartments q is correctly estimated to be one for almost all voxels in the left block (Fig. 2). In the center block q is sometimes overestimated—occasionally three instead of two basis functions are selected. This is mainly because the first basis

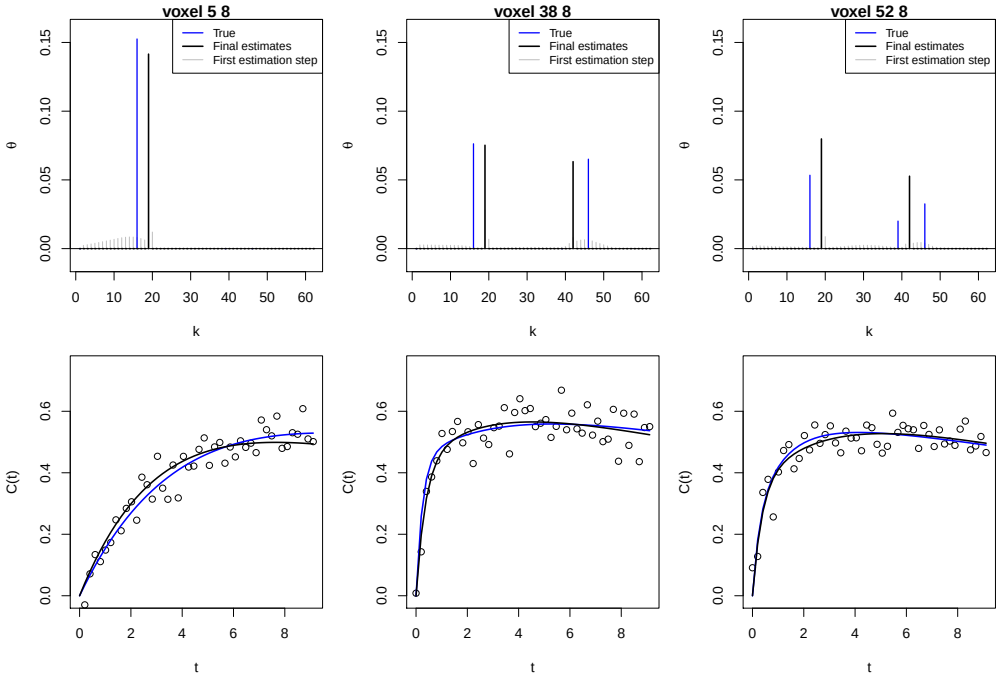

Figure 3: Results of spatially regularized estimation: Selection of basis functions and corresponding curve fits for three simulated voxels from the left ($q = 1$), center ($q = 2$) and right ($q = 3$) block

function is often additionally selected. For a CTC simulated from three compartments (right block, $q = 3$), the contributing basis functions are too similar and—in the voxel depicted—only two basis functions are selected (Fig. 3). Also for other voxels in the right block the estimate for q is often two, sometimes three. This result could be expected. In all of the three blocks, the fitted CTCs (gray) and the true underlying CTCs (blue) coincide (Fig. 3). Model selection for the v_p-term matches the true underlying models: in the upper half the AIF is mostly included in the model, in the lower half it is mostly excluded (Fig. 2). Furthermore, with the spatial regularization smooth parameter maps are obtained that match the true underlying parameters very well (see the Appendix).

4 DCE-MRI Breast Cancer Study

4.1 Data Description

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 two weeks of chemotherapy. The scans were acquired with a 1.5 T Siemens MAGNETOM Symphony scanner, $TR = 11$ ms and TE

$= 4.7\ ms$. Each scan consists of three slices of 230×256 voxel, but only the central slice was used in our analysis. A dose of $D = 0.1\ mmol/kg$ 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.

4.2 Results

For the voxelwise and spatially penalized multi-compartment models as well as for the extended Tofts model the BIC and the average number of selected coefficients per voxel are listed in Table 1. For all scans the spatial Elastic Net has the lowest BIC. This indicates that this approach is suitable to sparsely select parameters in DCE-MR images and still allowing for increased model complexity where needed.

The estimation results of the spatial Elastic Net for the pre- and post-treatment scans of patient 6 are depicted in Fig. 4. The estimated number of tissue compartments $\hat{q}$ and the inclusion of the AIF ($\mathbf{1}(\hat{v}_p > 0)$) are shown separately and sum up to the total number of compartments. The estimated number of tissue compartments $\hat{q}$ ranges between one and four and is two for a large number of voxels. In the pre-treatment scan, inside the tumor (top of the image), the tissue is relatively homogeneous ($\hat{q} = 1$ and $\hat{v}_p = 0$), whereas in surrounding tissue and especially at tumor margins the tissue is more heterogeneous ($\hat{q} \geq 1$). In the post-treatment scan, tumor size is reduced and the model complexity in the shrunk tumor has increased. From the difference in SSE compared to the extended Tofts model we find that the maximal benefit of additional compartments is at tumor margins. Here, the tissue is too heterogeneous to be adequately described by the extended Tofts model.

For most voxels, the AIF is not selected as predictor. It is selected more frequently at tissue borders, indicating a relevant amount of vascular space. If selected, the corresponding estimated v_p terms are rather small compared to contributions of other compartments: the median of non-zero v_p estimates ranges from 0.01 to 0.05 in the different scans. Summarizing, the average number of plasma and tissue compartments per voxel ($\hat{q} + \mathbf{1}(\hat{v}_p > 0)$) in a single scan ranges from 1.55 to 2.24 depending on the volume of homogeneous tissue, see Table 1.

The spatial Elastic Net estimates provide good fit to the observed CA concentration in different tissue regions. For exemplar voxels inside the tumor, at the tumor edge and in surrounding tissue the selection of basis functions and the final curve fits are shown in Fig. 5. Exemplar voxels are taken from the pre-treatment scan of patient 2, see the Appendix for precise location of the voxels. Parameter estimates for the entire scan are also shown in the Appendix.

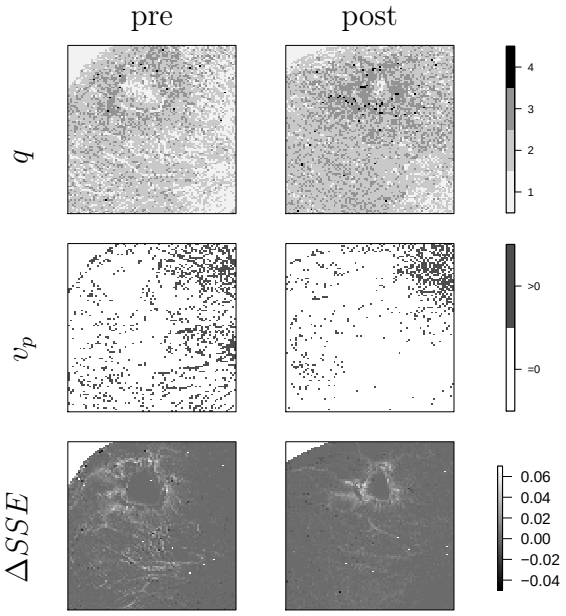

Figure 4: Results of spatially regularized estimation. Parameter maps for the mid-slice of patient 6 pre-treatment and post-treatment scans: estimated number of tissue compartments $\hat{q}$, estimated inclusion of plasma compartment $\hat{v}_p$ and difference in SSE ($\Delta SSE = SSE_{ext.Tofts} - SSE_{spatial}$)

5 Summary and Conclusions

Based on a multi-compartment model, we have proposed two penalized ML based approaches for data driven model choice and parameter estimation in DCE-MRI. Choosing basis functions based on the contribution of corresponding compartments, both approaches combine the advantages of data driven and model based approaches and parameters remain interpretable. The spatial Elastic Net which incorporates the spatial structure intrinsic in an image performed better than the voxelwise Elastic Net. Due to "borrowing strength" from neighbouring pixels, the spatial Elastic Net is more robust compared to the independent voxelwise Elastic Net. Hence, the spatial Elastic Net provides a sparser solution with similar fit to the data.

With a simulation study and the analysis of twelve DCE-MRI scans we found that the spatially penalized multi-compartment model outperforms the commonly used extended Tofts model as well as the voxelwise Elastic Net. Our results indicate that additional model complexity is needed to adequately describe observed CTCs, especially at tumor margins and surroundings. As the number of contributing tissue compartments (and the plasma compartment) is estimated per voxel, important information about the tissue heterogeneity is gained that cannot be obtained with *a priori* fixed model architectures.

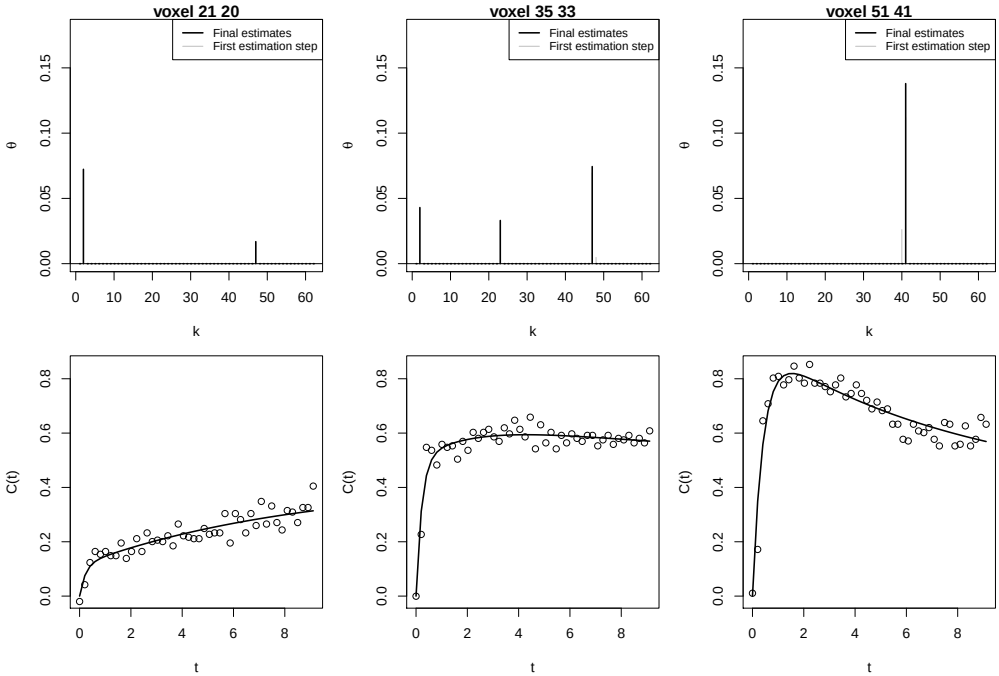

Figure 5: Results of spatially regularized estimation: Selection of basis functions and corresponding curve fits for a voxel in normal tissue $(21, 20)$, at the tumor edge $(35, 33)$ and inside the tumor $(51, 41)$; voxels from the mid-slice of patient 2 pre-treatment scan

Acknowledgements

Clinical data were graciously provided by Dr. A.R. Padhani, PSSC, Mount Vernon Hospital, Northwood, U.K.

References

Brix, G., S. Zwick, F. Kiessling, and J. Griebel (2009). Pharmacokinetic analysis of tissue microcirculation using nested models: Multimodel inference and parameter identifiability. *Medical Physics 36*(7), 2923–2933.

Buckley, D. L. and G. J. M. Parker (2005). Measuring Contrast Agent Concentration in T1-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.

Christensen, W. F. and F. Z. Yetkin (2005). Spatio-temporal analysis of auditory cortex activation as detected with silent event related fMRI. *Statistics in Medicine 24*(16), 2539–2556.

Gössl, C., D. P. Auer, and L. Fahrmeir (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics 57*(2), 554–562.

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.

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.

Kelm, B. M., B. H. Menze, O. Nix, C. M. Zechmann, and F. A. Hamprecht (2009). Estimating kinetic parameter maps from dynamic contrast-enhanced MRI using spatial prior knowledge. *IEEE Trans Med Imaging 28*(10), 1534–47.

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis 52*(1), 374–393.

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.

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., B. Whitcher, A. Padhani, and G. Yang (2009). Quantitative analysis of Dynamic contrast-enhanced MR images based on Bayesian P-Splines. *IEEE Transactions on Medical Imaging 28*(6), 789–798.

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.

Sommer, J. C. and V. J. Schmid (2012). Spatial two tissue compartment model for DCE-MRI. *Pre-print*.

Sourbron, S. P. and D. L. Buckley (2011). On the scope and interpretation of the Tofts models for DCE-MRI. *Magnetic Resonance in Medicine 66*(3), 735–745.

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. and A. Weingessel (2011). *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.5-4.

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.

Zou, K. H., W. M. Wells, R. Kikinis, and S. K. Warfield (2004). Three validation metrics for automated probabilistic image segmentation of brain tumours. *Statistics in Medicine 23*(8), 1259–1282.

Table 1: Comparison of extended Tofts model and the voxelwise or spatially penalized multi-compartment model: BIC and average number of compartments per voxel ($\hat{q} + \mathbf{1}(\hat{v}_p > 0)$) for simulated image and scans from breast cancer study.

		extended Tofts	voxelwise	spatial
simulation	BIC	462,317	438,313	436,423
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.90	2.51
scan 1	BIC	758,745	764,613	717,703
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.22	1.80
scan 2	BIC	619,243	587,659	561,450
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	1.86	1.53
scan 3	BIC	539,332	535,391	508,484
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.20	1.86
scan 4	BIC	638,952	630,269	610,377
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.26	1.89
scan 5	BIC	296,972	287,248	276,943
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.30	2.05
scan 6	BIC	440,293	428,462	407,239
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.00	1.67
scan 7	BIC	866,187	805,909	781,456
	$\hat{q} + \hat{v}_p$	2.00	1.78	1.59
scan 8	BIC	672,523	618,126	584,098
	$\hat{q} + \hat{v}_p$	2.00	2.00	1.69
scan 9	BIC	855,938	796,163	763,135
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.23	1.77
scan 10	BIC	777,398	704,287	675,509
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.53	2.13
scan 11	BIC	567,411	567,171	534,742
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.35	2.02
scan 12	BIC	582,249	587,292	559,053
	$\hat{q} + \mathbf{1}(\hat{v}_p > 0)$	2.00	2.53	2.24

Appendix / Supplementary Material

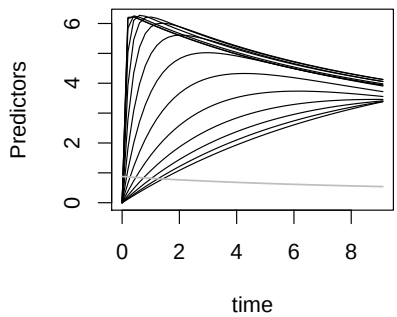

Figure 6: Subset of predictors (every fifth predictor is shown). Grey line: $C_p(t)$, black lines: $\Psi_1(t), \ldots, \Psi_q(t)$

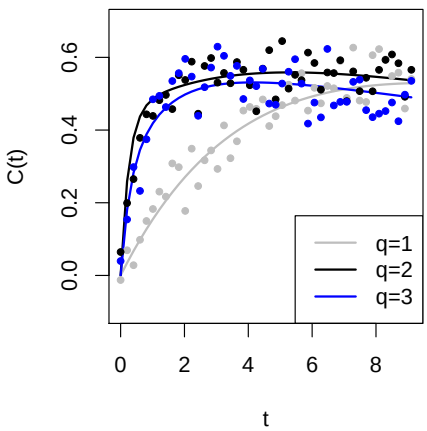

Figure 7: Simulated CTCs with underlying parameters $v_p = 0$, $k_{\mathrm{ep}_1} = 0.2$, $k_{\mathrm{ep}_2} = 1$, $k_{\mathrm{ep}_3} = 4$. $q = 1$: $K_{\mathrm{trans}_1} = 0.2$. $q = 2$: $K_{\mathrm{trans}_1} = 0.1$, $K_{\mathrm{trans}_3} = 2$. $q = 3$: $K_{\mathrm{trans}_1} = 0.07$, $K_{\mathrm{trans}_2} = 0.3$, $K_{\mathrm{trans}_3} = 1$

$q = 1$ $v_p > 0$	$q = 2$ $v_p > 0$	$q = 3$ $v_p > 0$
$q = 1$ $v_p = 0$	$q = 2$ $v_p = 0$	$q = 3$ $v_p = 0$

Figure 8: Sketch of simulation design. The simulated image consists of 75×75 voxels divided into three main blocks ($q = 1, 2, 3$). Each main block is split up into a block with $v_p > 0$ or $v_p = 0.1$.

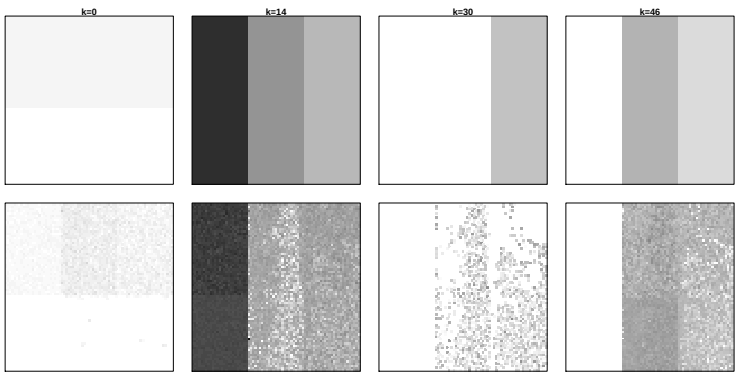

Figure 9: Maps of true underlying coefficients Ψ_k of the simulation study and corresponding estimated coefficients (spatially regularized)

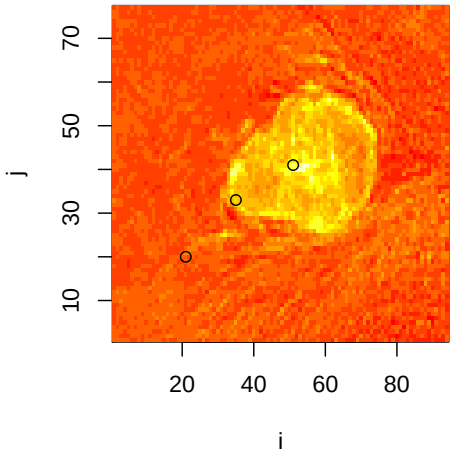

Figure 10: Concentration map at a fixed time point $C(t_{10})$ (patient 2 pre-treatment). Voxels $(21, 20)$, $(35, 33)$ and $(51, 41)$ are marked.

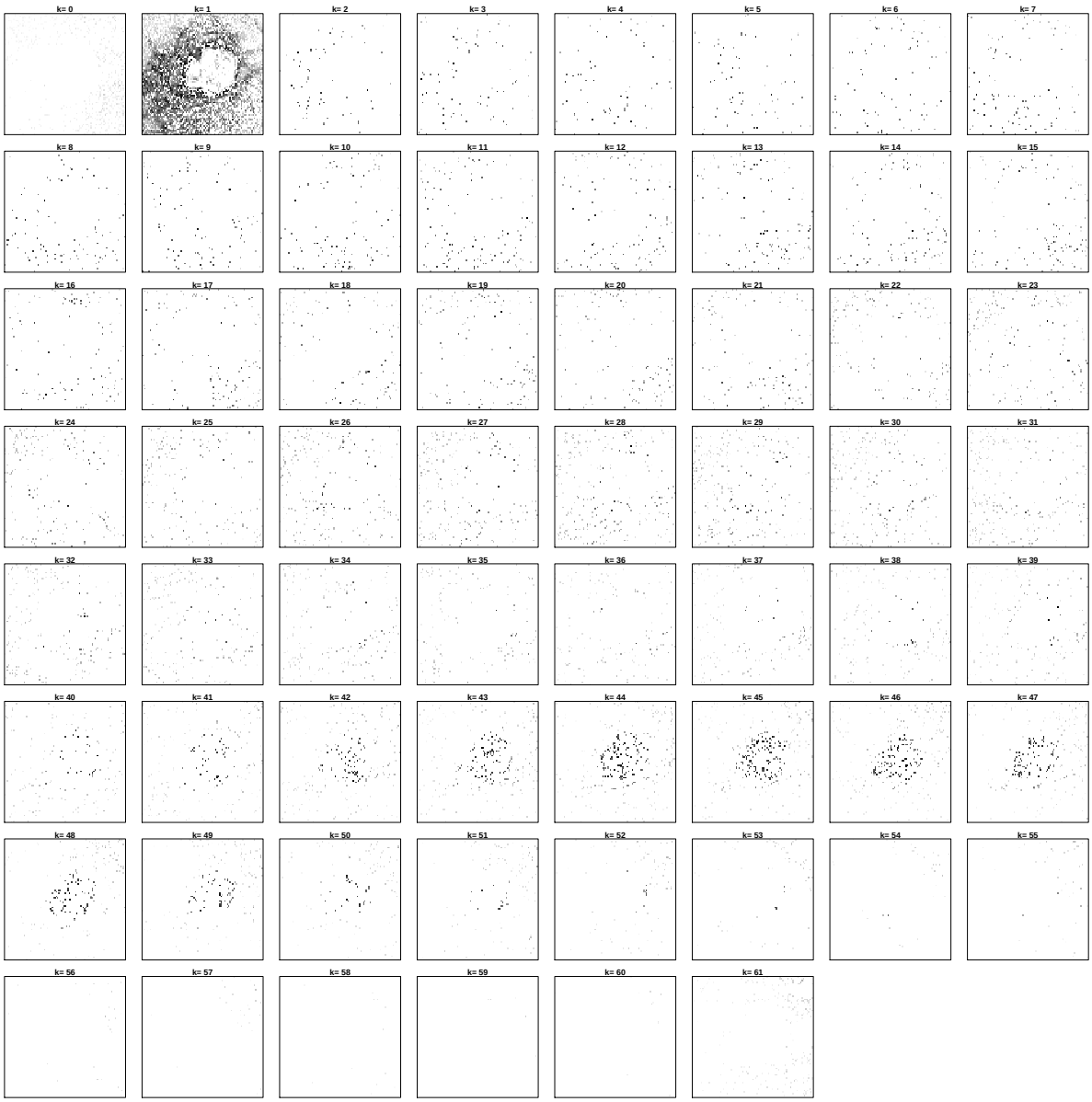

Figure 11: Maps of estimated coefficients Ψ_k, $k = 0, \ldots, 61$ (patient 2 pre-treatment)