Heim, Susanne; Fahrmeir, Ludwig; Eilers, Paul H. C.; Marx, Brian D.

**Working Paper**

# Space-varying coefficient models for brain imaging

Discussion Paper, No. 455

**Provided in Cooperation with:**
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

*Suggested Citation:* Heim, Susanne; Fahrmeir, Ludwig; Eilers, Paul H. C.; Marx, Brian D. (2005) : Space-varying coefficient models for brain imaging, Discussion Paper, No. 455, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München,
https://doi.org/10.5282/ubm/epub.1824

This Version is available at:
https://hdl.handle.net/10419/31061

# Space-Varying Coefficient Models for Brain Imaging

S. Heim[1*], L. Fahrmeir[1], P. H. C. Eilers[2], and B. D. Marx[3]

[1] Department of Statistics, Ludwig-Maximilians-University, Munich, Germany

[2] Department of Medical Statistics, Leiden University Medical Center, 2300 RA, Leiden, The Netherlands

[3] Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803 USA

Summary

The methodological development and the application in this paper originate from diffusion tensor imaging (DTI), a powerful nuclear magnetic resonance technique enabling diagnosis and monitoring of several diseases as well as reconstruction of neural pathways. We reformulate the current analysis framework of separate voxelwise regressions as a 3d space-varying coefficient model (VCM) for the entire set of DTI images recorded on a 3d grid of voxels. Hence by allowing to borrow strength from spatially adjacent voxels, to smooth noisy observations, and to estimate diffusion tensors at any location within the brain, the three-step cascade of standard data processing is overcome simultaneously. We conceptualize two VCM variants based on B-spline basis functions: a full tensor product approach and a sequential approximation, rendering the VCM numerically and computationally feasible even for the huge dimension of the joint model in a realistic setup. A simulation study shows that both approaches outperform the standard method

of voxelwise regressions with subsequent regularization. Due to major efficacy, we apply the sequential method to a clinical DTI data set and demonstrate the inherent ability of increasing the rigid grid resolution by evaluating the incorporated basis functions at intermediate points. In conclusion, the suggested fitting methods clearly improve the current state-of-the-art, but ameloriation of local adaptivity remains desirable.

*Keywords:* Diffusion tensor; Brain imaging; P-splines; Varying coefficient model

*Correspondence to: Susanne Heim, Ludwigstrasse 33/II, 80539 Munich, Germany,
E-mail: susanne.heim@stat.uni-muenchen.de, Tel: +49-89-2180-2226,
Fax: +49-89-2180-5040

# 1 Introduction

Our methodological work on space-varying coefficient models results from collaborative research in neuroscience where statistical approaches in functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) are confronted with manifold challenging problems. While fMRI aims at detecting task-related neural activation which translates into signal changes of the involved cortical areas, the ultimate goal of DTI is to recover anatomical connections between brain regions. The latter which is also referred to as fiber tracking, and the physical basis of the underlying principle of directional diffusion have been comprehensively reviewed by e. g. Basser and Jones (2002). Though we will focus on DTI, the presented methodology can equally be applied to fMRI.

As described in more detail in Section 2, the basic quantity in DTI is the so-called diffusion tensor $\mathbf{D}(\mathbf{s})$ which can be interpreted as the three-dimensional (3d) covariance matrix of an anisotropic Wiener process at any location $\mathbf{s} \in \mathbb{R}^3$ of interest in the human brain. Using spectral decomposition, several tensor derived metrics can be obtained, providing insight into microstructural tissue properties and pathological alterations. While eigenvalue based measures used in diagnosing and monitoring acute and chronic neurological diseases such as stroke, brain tumours or inflammatory disease, the eigenvectors allow for fiber tracking, a promising high-end application of DTI with the outlook to improve neurosurgical planning. DTI raw data are inherently artifact-prone and recorded at limited resolution on

a discrete 3d grid of voxels, indexed by $s = 1, \ldots, n$. In a realistic setup, a typical image comprises $n_1 \times n_2 \times n_3 = 128 \times 128 \times 24$ voxels, resulting in a total of $n = 393,216$ voxels. A complete data volume consists of multiple images recorded under different conditions and can be transformed to repeated continuous measurements $y_i(s), i = 1, \ldots, r$, at each voxel $s$. These measurements can be related to the vector $\boldsymbol{\beta}(s) = (\beta_1(s), \ldots, \beta_6(s))'$ of unknown elements in $\mathbf{D}(s)$ through a regression model

$$y_i(s) \;=\; \mathbf{x}_i{}'\boldsymbol{\beta}(s) + \varepsilon_i(s)\,, \quad i = 1, \ldots, r\,. \tag{1.1}$$

The covariate vector $\mathbf{x}_i$ is determined by the design of the DTI experiment. A similar regression model can be derived for fMRI data though with different interpretation (Gössl et al., 2001).

For the purpose of estimating the coefficients $\boldsymbol{\beta}(s)$, and thus the diffusion tensor $\mathbf{D}(s)$, some regression technique is currently applied in each image voxel separately, reaching from standard least squares (Basser et al., 1994) to more sophisticated techniques such as robust (Mangin et al., 2002; Chang et al., 2005) or non-Gaussian error regression (Assaf et al., 2004). Spatial correlation induced by adjacent voxels is taken into account, if at all, in a postprocessing step, e. g. by Gaussian kernel smoothing of estimated diffusion tensors (Gössl et al., 2002).

In this work, we propose to connect the seemingly unrelated regression models in Eq. (1.1) to a joint model for all measurements $\mathbf{y} = \{y_i(s), i = 1, \ldots, r; s =$

$1, \ldots, n\}$ with $\boldsymbol{\beta} = \{\boldsymbol{\beta}(s), s = 1, \ldots, n\}$ being a high-dimensional space-varying coefficient vector. This approach considers not only spatial correlation within the model but also corroborates the tensor fitting by the diffusivity information of adjacent voxels. Moreover, the use of spline-basis functions serves to model the spatial field $\mathbf{D} = \{\mathbf{D}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^3\}$, i. e. the field $\boldsymbol{\beta} = \{\boldsymbol{\beta}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^3\}$. As a consequence, estimates for $\mathbf{D}(\mathbf{s})$ are accessible at any arbitrary position $\mathbf{s} \in \mathbb{R}^3$ in the brain, not only on the discrete grid of voxels $s = 1, \ldots, n$. This approximation property is an important pre-requisite of any DTI tracking algorithm which needs estimates of $\mathbf{D}(\mathbf{s})$ on a markedly finer 3d grid than available by the acquisition resolution in order to reconstruct fiber bundles in a biologically smooth fashion. Current techniques mostly use some sort of simple interpolation technique between voxels, see e. g. Gössl et al. (2002) and the references given therein.

At first glance, our proposed concept of a space-varying coefficient model seems to be straightforward. However, the massive dimension of the 3d array $\boldsymbol{\beta} = \{\boldsymbol{\beta}(s), s = 1, \ldots, n\}$ of coefficient vectors of length six, thus of a 4d array, implies methodological and computational challenges which cannot be solved adequately with existing methods. Originally, VCMs have been suggested by Hastie and Tibshirani (1993) for regressions with coefficients varying smoothly over a one-dimensional continuous variable such as time-varying effects. Extensions to 2d-space-varying coefficients have been developed more recently, ranging from 2d-surface smoothers to (Markov) random field models (Assunção, 2003; Eilers and

Marx, 2003; Fahrmeir et al., 2004). Further extensions from two dimensions to three- or higher-dimensional arrays are more challenging than it may seem at first sight and much less efforts have been spent on this topic, with the noteable exception of the array regression method proposed by Eilers et al. (2006). Based on a suggestion in Heim et al. (2004), array regression was used in a first attempt to analyze DTI data with a VCM model in Eilers et al. (2005). Unfortunately, the method of Eilers et al. (2005) of implementing the tensor product model does not allow for enough knots in 3d-space, thus smoothing away important details contained in DTI data.

In this work, two techniques are developed to overcome these problems. The first one (see Section 3.1) shares the theoretical model with the array regression but takes advantage of the sparsity of the spatial arrays involved. The second one (see Section 3.2) basically adapts the 'new smoothing spline' in Dierckx (1982), thus reducing the 3d (or higher-dimensional) problem to a sequence of one-dimensional smoothers. In the next section we provide a brief overview of the physical background of DTI. Section 3 describes our conceptual framework composed of the two estimation techniques. The performance of the proposed methods for DTI analysis is explored through a simulation study in Section 4, and Section 5 is devoted to a real data DTI application. Discussions and conclusions are presented in Section 6 while technical details of the implementation are deferred to the Appendix.

# 2 Diffusion tensor imaging (DTI)

DTI is named according to its fundamental quantity, the field $\{\mathbf{D}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^3\}$ of diffusion tensors

$$\mathbf{D}(\mathbf{s}) = \begin{pmatrix} D_1(\mathbf{s}) & D_4(\mathbf{s}) & D_5(\mathbf{s}) \\ D_4(\mathbf{s}) & D_2(\mathbf{s}) & D_6(\mathbf{s}) \\ D_5(\mathbf{s}) & D_6(\mathbf{s}) & D_3(\mathbf{s}) \end{pmatrix}.$$

A diffusion tensor characterizes the local diffusivity of water molecules at each location $\mathbf{s} \in \mathbb{C} \subset \mathbb{R}^3$ within a continuous subspace $\mathbb{C} \subset \mathbb{R}^3$ of the brain, and can be interpreted as the (local) covariance matrix of an anisotropic Wiener process describing the random movement of water molecules. The eigenvalues $\xi_i(\mathbf{s}), i = 1, 2, 3$, and eigenvectors of this symmetric and (theoretically) positive definite matrix correspond to the axis lengths and directions of the local diffusion ellipsoid. Among the eigenvalue-based intravoxel measures, fractional anisotropy (FA) is the most popular metric to assess the local degree of anisotropy of the diffusion process. By its definition (Basser and Pierpaoli, 1996)

$$\mathrm{FA}(\mathbf{s}) = \sqrt{\frac{3}{2} \frac{\sum_i (\xi_i(\mathbf{s}) - \bar{\xi}(\mathbf{s}))^2}{\sum_i \xi_i^2(\mathbf{s})}},$$

the FA index assumes zero in a perfect isotropic medium and equals one in a purely anisotropic medium. FA and other eigenvalue based metrics are used to characterize the physiological brain microstructure and neuropathological processes. However, both the analysis of specified brain regions and whole brain approaches such

as voxel based morphometry are hampered by the generally coarse spatial resolution of diffusion images and by error prolongation caused by noise. Therefore, sufficient approximation techniques for high resolution mapping of small scale processes and complex anatomical structures, e. g. the hippocampus, are desirable.

On the other hand, the dominant eigenvector of the diffusion ellipsoid reflects the principal diffusion direction which is particularly pronounced in white matter: Neural fibers in this brain compartment are densely packed and highly ordered such that the water molecules therein preferentially pass along the biophysiological structures instead of perpendicular to them. Hence the main eigenvector is assumed to coincide (approximately) with the local fiber orientation in space. This conceptual postulate underlies all existing tracking algorithms which, as high-end DTI analysis, aim at reconstructing neural fiber bundles (Mori and van Zijl, 2002). Figure 1 outlines this basic idea for a certain slice of a 3d data grid. The gray curve corresponds to a stylized part of a fiber bundle, and for each discrete grid point, indexed by $s$, the gray arrow indicates the dominant eigenvector of the "true" unobservable diffusion tensor $\mathbf{D}(s)$. The orientation of these true eigenvectors varies more or less randomly in isotropic parts of the volume while eigenvectors associated with the unobservable fiber bundle correspond to anisotropic diffusion tensors and provide good information about local directionality of the neural tract, that is still to be reconstructed. Linking together the eigenvectors by some suitable tracking algorithm would allow to recover the neural tract. However, as described

below, only (noisy) tensor estimates $\hat{\mathbf{D}}(s)$ can be obtained from the recorded DTI data and hence the derived principal eigenvectors will likely be erroneous as visualized by black arrows in Fig. 1. Clearly, tracking algorithms working with these unsatisfactory principal eigenvectors can yield spurious results due to fairly misleading orientation of some estimated eigenvectors. See, for instance, the severely distorted main eigenvector in the middle fiber-transit voxel of Fig.1. There is an evident need to denoise raw estimates by some kind of spatial smoothing, borrowing strength from adjacent voxels. In the following we describe the standard approach of obtaining estimates of $\mathbf{D}(\mathbf{s})$ as well as its eigenvalues and -vectors from DTI raw data.

Physical theory states that the relation between a measured magnetic resonance signal $S_i(s)$ in voxel $s$ and a certain gradient $\mathbf{g}_i = (g_{1i}, g_{2i}, g_{3i})'$ (see Mori and Barker, 1999, for physical background) is given by the (deterministic) Stejskal-Tanner equation (Stejskal and Tanner, 1965):

$$S_i(s) \;=\; S_0(s) \exp\left\{-b\, \mathbf{g}_i{}' \mathbf{D}(s) \mathbf{g}_i\right\}, \quad i = 1, \ldots, r, \quad s = 1, \ldots, n. \tag{2.1}$$

Here, $b$ is a scalar comprising several acquisition parameters such as magnetic gradient strength and duration. Yet in practice, the relation (2.1) is disturbed by thermal and physiological noise. Following Papadakis et al. (1999), Eq. (2.1) can be reformulated, for voxel $s$, $s = 1, \ldots, n$, as

$$y_i(s) = -\frac{1}{b} \log\left(\frac{S_i(s)}{S_0(s)}\right) \;=\; \mathbf{x}_i{}' \boldsymbol{\beta}(s) + \varepsilon_i(s), \quad i = 1, \ldots, r, \tag{2.2}$$

with noise $\varepsilon_i(s) \overset{\text{iid}}{\sim} N(0, \sigma^2)$, the vector $\boldsymbol{\beta}(s) = (D_1, D_2, D_3, D_4, D_5, D_6)'(s)$ of the unknown elements of the diffusion tensor and the design vector $\mathbf{x}_i = (g_{1i}^2, g_{2i}^2, g_{3i}^2, 2g_{1i}g_{2i}, 2g_{1i}g_{3i}, 2g_{2i}g_{3i})'$, constructed from the known values of $\mathbf{g}_i$. In Eq. (2.2) we recognize a linear regression problem with the unknown $p$-dimensional vector $\boldsymbol{\beta}(s)$. Note that $\mathbf{x}_i$ is the same for all voxels, and thus we face a repeated measures design with $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_r)' \sim (r, p)$ which plays a role for efficient implementation (see Appendix).

The current standard approach uses least squares (Basser et al., 1994) or more sophisticated techniques (Mangin et al., 2002; Chang et al., 2005; Assaf et al., 2004) to estimate $\boldsymbol{\beta}(s)$ through separate regressions at each voxel $s$, requiring at least $r \geq 6$ independent images obtained at different magnetic gradients. Usually, measurements are repeated for the same gradient set or the number of different gradients is chosen to be over determined in order to mitigate the effects of noise. Due to ethical as well as financial reasons, the clinical time frame is however limited.

In general, the first step of the standard tensor fitting is thus multiple regression on a voxel-by-voxel basis. In a second step, spatial smoothing (of the estimated) diffusion tensor field is often performed, for example by applying a Gaussian kernel as in Gössl et al. (2002). If required some subsequent interpolation serves to increase the resolution in a third step.

# 3 The space-varying coefficient model

Our concept combines the $n$ separate regression models of the standard approach to a joint space-varying coefficient model by a suitable (spatial) design of the 3d array $\{\boldsymbol{\beta}(s), s = 1, \ldots, n\}$. Thus, spatial correlation and information from adjacent voxels are taken into account. In addition, the number $r$ of repetitions at each voxel can be kept small as desirable to avoid long acquisition times.

## 3.1 Multidimensional smoothing with tensor products

The elements of $\boldsymbol{\beta}_j$ $(j = 1, \ldots, p = 6)$ are modelled non-parametrically by projecting them onto penalized tensor product B-splines or multidimensional P-splines. We first present a direct VCM presentation. Consider

$$\boldsymbol{\beta}_j(s) = \sum_{v=1}^{KLM} \mathbf{B}(s, v)\gamma_j(v) = \mathbf{B}(s, \cdot)\boldsymbol{\gamma}_j,$$

where the $(n \times KLM)$-matrix $\mathbf{B} = \mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1$ contains the 3d tensor products of 1d B-splines, i. e. of $(n_1 \times K)$-matrix $\mathbf{B}_1$, $(n_2 \times L)$-matrix $\mathbf{B}_2$, and $(n_3 \times M)$-matrix $\mathbf{B}_3$ evaluated at x-, y-, and z-coordinates, respectively. Regarding the full $j$th coefficient surface we equivalently write

$$\boldsymbol{\beta}_j = \mathbf{B}\boldsymbol{\gamma}_j\,. \tag{3.1}$$

Note that $\mathbf{B}$ is the same for all coefficient surfaces $\boldsymbol{\beta}_j$ and has to be calculated only once. The vector $\boldsymbol{\gamma}_j$ denotes the unknown amplitudes of the basis functions, and

$K \times L \times M$ is determined by the (generous and regularly gridded) knot partition and degree of the basis functions. To ensure sufficiently fine reconstruction of essential features in DTI, one knot is required at each 1.25 voxel as suggested by a pilot study. Furthermore, the spline degree is set linear in order to keep the influencing regions of the basis functions as locally restricted as possible. There is nothing prohibative in the general methodology from using other choices of basis degree.

Using the tensor coefficient expression in (3.1), the least squares term conforms to

$$LS(\boldsymbol{\gamma}) = \|\mathbf{y} - \sum_{j=1}^{p} \boldsymbol{\Upsilon}_{\mathbf{X}(\cdot,j)} \boldsymbol{\beta}_j\|^2 = \|\mathbf{y} - \sum_{j=1}^{p} \boldsymbol{\Upsilon}_{\mathbf{X}(\cdot,j)} \mathbf{B} \boldsymbol{\gamma}_j\|^2 \,,$$

where the observations in $\mathbf{y} \sim (rn, 1)$ are ordered according to repeats first and voxel number last, $\boldsymbol{\beta}_j = (\beta_j(1), \ldots, \beta_j(n))'$, $\mathbf{B}$ is of dimension $n \times KLM$, $\boldsymbol{\gamma}_j$ of $KLM \times 1$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_p')'$. The $(rn \times n)$-matrix $\boldsymbol{\Upsilon}_{\mathbf{X}(\cdot,j)} = \mathbf{X}(\cdot, j) \otimes \mathbf{I}_n$ with $(n \times n)$-identity matrix $\mathbf{I}_n$ consists of $n$ blocks, each one containing the $r$-dimensional regressor $j$. We then aim to find a practical solution to the penalized objective

$$\begin{aligned} LS_{pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) &= LS(\boldsymbol{\gamma}) + \text{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) \\ &= \|\mathbf{y} - (\mathbf{B} \otimes \mathbf{X})\boldsymbol{\gamma}\|^2 + \text{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) \end{aligned}$$

Note that the second equality of $LS(\boldsymbol{\gamma})$ in (3.2) holds because the $p$ interaction variables are space-invariant in the present application of DT imaging and, hence,

the $n$ blocks in $\boldsymbol{\Upsilon}_{\mathbf{X}(\cdot,j)}$ are the same. This restatement plays an important role in the implementation.

The penalty term $\text{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma})$ serves to avoid overfitting. For each regressor $j$, difference penalties are placed on the rows, columns, and layers of tensor product coefficients, such that

$$\begin{aligned}
\text{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) &= \sum_{j=1}^{p}\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{m=1}^{M}\{\lambda_1(\Delta_1^d\gamma_{jklm})^2 + \lambda_2(\Delta_2^d\gamma_{jklm})^2 + \lambda_3(\Delta_3^d\gamma_{jklm})^2\} \\
&= \lambda_1\|(\mathbf{I}_{LM} \otimes \boldsymbol{\Delta}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma}\|^2 + \lambda_2\|(\mathbf{I}_M \otimes \boldsymbol{\Delta}_2 \otimes \mathbf{I}_{Kp})\boldsymbol{\gamma}\|^2 \\
&\quad +\lambda_3\|(\boldsymbol{\Delta}_3 \otimes \mathbf{I}_{KLp})\boldsymbol{\gamma}\|^2 \\
&= \lambda_1\|\mathbf{P}_1\boldsymbol{\gamma}\|^2 + \lambda_2\|\mathbf{P}_2\boldsymbol{\gamma}\|^2 + \lambda_3\|\mathbf{P}_3\boldsymbol{\gamma}\|^2 . \qquad (3.2)
\end{aligned}$$

Here $\Delta_1^d, \Delta_2^d$, and $\Delta_3^d$ denote the $d$-th order differences across a row, down a column, and along a layer of the $K \times L \times M$ matrix of tensor product B-spline coefficients, $\boldsymbol{\Gamma}_j = [\gamma_{klm}]$, respectively. Note that, in general, different values of $d$ are conceivable for the three penalized dimensions. In the present application to DTI data, $d = 1$ proved most appropriate with respect to satisfactory detail sustainment. The $(K-d\times K)$-matrix $\boldsymbol{\Delta}_1$, $(L-d\times L)$-matrix $\boldsymbol{\Delta}_2$, and $(M-d\times M)$-matrix $\boldsymbol{\Delta}_3$ indicate the corresponding matrices of difference penalties, while $\mathbf{P_1}$, $\mathbf{P_2}$, and $\mathbf{P_3}$ consists of a carefully arranged, full matrix representation using block diagonal matrices of right Kronecker products. If $\lambda_1 \neq \lambda_2 \neq \lambda_3$, the smoothing parameter is dimension-specific. In contrast, the smoothing parameter is global if $\lambda_1 = \lambda_2 = \lambda_3$. In both cases, it determines the trade-off between smoothness and

fidelity to the data. Details on data-driven optimization of the tuning parameter are given in the Appendix.

In theory, an explicit solution to $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_p')'$ can be found using

$$\hat{\boldsymbol{\gamma}} = (\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}\mathbf{U}'\mathbf{y}, \tag{3.3}$$

where $\mathbf{U} = [\boldsymbol{\Upsilon}_{\mathbf{X}(\cdot,1)}\mathbf{B}, \ldots, \boldsymbol{\Upsilon}_{\mathbf{X}(\cdot,p)}\mathbf{B}] = \mathbf{B} \otimes \mathbf{X}$ is $(rn \times pKLM)$-dimensional, and $\mathbf{P} = \lambda_1\mathbf{P_1} + \lambda_2\mathbf{P_2} + \lambda_3\mathbf{P_3}$ is derived from (3.2). Given $\boldsymbol{\gamma}$, then the varying coefficient volumes can be built. Since the basis functions can routinely be evaluated at a large number of (intermediate) points, the resolution can be increased straightforwardly without an additional interpolation method.

For practical brain imaging applications, we may need $K \times L \times M = 32 \times 32 \times 8$ knots for a region of interest sized $40 \times 40 \times 10$ voxels. Thus $\mathbf{B}$ has approximately $3 \times 10^9$ elements: if each floating point takes 8 bytes, then $\mathbf{B}$ will use several Gb of memory, which is beyond reach of current computers. The left hand side of the normal equations preceding Eq. (3.3) would occupy more than 18 Gb for the above number of parameters, i. e. $6 \times 32 \times 32 \times 8$. Hence efficient algorithmic implementation and programming using sparsity and approximations is essential (see also the Appendix).

## 3.2 Successive smoothing with univariate basis functions

A simplification of the VCM version with tensor products can be derived from Dierckx (1982; 1993: p. 172ff) who proposes the so-called "new smoothing spline" for penalized 2d smoothing. The key point concerns the data arrangement which is left in multidimensional array structure, here $(r \times n_1 \times n_2 \times n_3)$-array $\mathbf{Y}$. Starting from the univariate B-spline matrices augmented by the corresponding difference penalties, the normal equations can be transformed employing linear algebra properties of the Kronecker product. Then the coefficient vector $\boldsymbol{\gamma}$ results from consecutive univariate smoothing with the observational data as input to the first smoothing cycle exclusively. The (preliminary) coefficients attained last serve each as input to further iterations.

The generalization to higher dimensions can be implemented on the basis of the function $\rho(\mathbf{V}, \mathbf{W}, i)$ published by Eilers et al. (2006) and described in detail by Currie et al. (2006). The argument $\mathbf{V}$ is a 2d matrix, $\mathbf{W}$ a higher-dimensional array, and $i$ a dimension index. This function performs the usual computations of the matrix product, along the rows of $\mathbf{V}$ and along dimension $i$ of $\mathbf{W}$. The core trick is to rotate dimension $i$ of $\mathbf{W}$ to the front by dimension permutation, reduce the array $\mathbf{W}$ to two dimensions, compute the standard matrix product with $\mathbf{V}$, transform $\mathbf{W}$ back to its original dimensionality and rotate the $i$th dimension back

to place. Repeated application of $\rho$ to the signal intensities $\mathbf{Y}$ yields the following $(p \times K \times L \times M)$-array estimate of $\mathbf{\Gamma} = \mathrm{vec}(\boldsymbol{\gamma})$, where the vec-operator performs column-wise stacking of its argument:

$$\hat{\mathbf{\Gamma}} = \rho\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',\ \rho\left((\mathbf{B}_1'\mathbf{B}_1 + \lambda_1\boldsymbol{\Delta}_1'\boldsymbol{\Delta}_1)^{-1}\mathbf{B}_1',\ \rho\left((\mathbf{B}_2'\mathbf{B}_2 + \lambda_2\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}\mathbf{B}_2',\right.\right.\right.$$
$$\left.\left.\left.\rho\left((\mathbf{B}_3'\mathbf{B}_3 + \lambda_3\boldsymbol{\Delta}_3'\boldsymbol{\Delta}_3)^{-1}\mathbf{B}_3',\ \mathbf{Y},\ 4\right),\ 3\right),\ 2\right),\ 1\right)\ . \tag{3.4}$$

Starting from the inner brackets, $\rho$ smoothes the fourth dimension of $\mathbf{Y}$, thereby transforming the orginal $(r \times n_1 \times n_2 \times n_3)$-dimensional $\mathbf{Y}$ to a $(r \times n_1 \times n_2 \times M)$-array. Analogous proceeding renders this $(r \times n_1 \times n_2 \times M)$-array $(r \times n_1 \times L \times M)$-dimensional, then $(r \times K \times L \times M)$- and finally $(p \times K \times L \times M)$-dimensional. Note that sequential univariate smoothing according to Eq. (3.4) is highly attractive in terms of working memory and computation time. This approach affords an entirely different opportunity to overcome implementational deficiencies. Due to the sequential character of the procedure, involved matrices are small a priori allowing to place a large number of basis functions. Moreover, the GCV can be computed very efficiently since the trace of Kronecker products equals the product of traces.

# 4 Simulation study

For the purpose of performance rating of the VCM approaches versus the standard estimation method, a simulation study is conducted with quasi-realistic parameter settings and a fiber tract following the geometry of a spiral. Figure 2 displays the stylized fiber bundle together with the auxillary grid of anisotropic voxel size typical to clinical experiments. Since each fiber-transit voxel is considered a fiber voxel (compare e. g. the projection on the XY-plane), the overall ratio of fiber to non-fiber voxels amounts to 0.2.

——— *figure 2 around here* ———

In the given simulation model the background tensors are spherically shaped corresponding to a totally isotropic diffusion process. Yet, in spiral voxels the tensor shape is cigar-like. This means that the underlying diffusion process is unambiguously one-directional. In other terms, the template eigenvalues determining the spiral tensors obey a ratio of 2:1:1 giving a fractional anisotropy value of FA $= 0.4$. We defer to Fig. 5 for a complete map of the true tensors. The imposed amount of Gaussian error is taken from 70 real data sets as the average background noise. Also the voxel size mimics real data situation generally occuring in DTI.

In the following, standard voxelwise regression is abbreviated ST1, ST2 if Gaussian kernel smoothing is appended, and ST3 for additional interpolation. Among

the VCM variants, realization with tensor product basis functions, as well as successive univariate smoothing according to Dierckx, are examined constrained to global and dimension-specific smoothing parameters, respectively. We will refer to these approaches as TPglob, TP3d, DXglob, and DX3d. To ensure sufficient reconstruction of fine, although essential features, one knot is required at each 1.25 voxel in both artificial and real data. Linear spline basis functions and first order difference penalties further promote texture preservation. Note that the analysis is restricted to the fitted coefficient surfaces $\{\hat{\boldsymbol{\beta}}_j, j = 1, \ldots, p\}$ constituting the tensor field, and to the derived 3d field of FA which represents the most widely used scalar measure of anisotropy. The total number of regression coefficients, i. e. of amplitudes of the basis functions, amounts to $12 \times 12 \times 4 \times 6 = 3456$ in all VCM approaches.

For each of the $N = 100$ simulation runs, we assessed the overall quality of the competing estimation procedures by the averaged mean squared error (AMSE) of the tensor fit, depending on the voxel type 'spiral':

$$\mathrm{AMSE}_{\mathrm{sp}}^{(i)} = \frac{1}{p\, n_{\mathrm{sp}}} \sum_{s \in \mathrm{spiral}} \left\| \boldsymbol{\beta}(s) - \hat{\boldsymbol{\beta}}^{(i)}(s) \right\|^2 , \quad i = 1, \ldots, N ,$$

or 'background':

$$\mathrm{AMSE}_{\mathrm{bg}}^{(i)} = \frac{1}{p\, n_{\mathrm{bg}}} \sum_{s \in \mathrm{background}} \left\| \boldsymbol{\beta}(s) - \hat{\boldsymbol{\beta}}^{(i)}(s) \right\|^2 , \quad i = 1, \ldots, N .$$

The $\boldsymbol{\beta}(s)$ and $\hat{\boldsymbol{\beta}}^{(i)}(s)$ denote the template coefficient vector and its counterpart fitted from the $i$th simulation run, respectively. Note that AMSE is a region-specific error measure per simulation run, averaged over all region-related voxels and all tensor elements. Analogous definitions holds for FA.

Thus, we can obtain empirical error distributions of the estimated coefficient surfaces and also of the FA values once confined to fiber tissue (spiral) and once confined to background tissue. From the comparison of the relevant boxplots, we are able to judge which estimation method performs best with regard to tissue types of different diffusivity properties.

With respect to the spiral voxels that are of particular importance for fiber tracking, Figure 3 confirms the superiority of the VCM approaches over the current state-of-the-art ST2 when fitting the tensor components. DX3d performs best, followed by DXglob, TP3d, TPglob and ST2 with medians at -18.98, -18.93, -18.85, -18.83, and -18.53 (left boxplots of top supfigure). This ranking changes to advantage of 3d tuning parameter when the resolution is duplicated as evident from the left boxplots of the bottom subfigure. On this last level of data processing prior to fiber tracking, the various VCM approaches outperform the ST3 procedure (median $= -18.24$) with the clearest improvement yielded by DX3d (-19.04), followed by TP3d (-18.91), DXglob (-18.87) and lastly by TPglob (-18.82). In the isotropic background, error distributions appear rather similar between the considered approaches both at original and refined resolution.

——— *figure 3 around here* ———

If the fitted tensor field is transformed to the 3d FA field, almost the same relationship of quality of fit holds for the different techniques: FA is fitted more accurately in the background than in the spiral voxels, where the various VCM approaches lead to clearly smaller errors than ST2. This latter distinction becomes more pronounced with respect to interpolated estimates though the ranking remains exactly the same, i. e. DX3d, TP3d, TPglob, DXglob, and ST2/ST3 in decreasing performance order. It is worth to mention that the VCM variants result in even smaller errors of spiral than background FA estimates, if the degree of anisotropy is further augmented to an eigenvalue ratio of 10:1:1 (FA = 0.89; data not shown). In the background compartment, all methods show a tendency of larger errors at increased resolution compared to original acquisition.

——— *figure 4 around here* ———

Concerning AMSE of both tensor and FA estimation, VCM approaches with a 3d penalty perform altogether better than those with a global tuning parameter. Based on these results and from a computational point of view, we clearly favor DX3d. For comparison, TP3d and ST2 are also examined in more detail, namely for the tensor estimation.

To glean intravoxel quality information, we compute for each coefficient surface $\boldsymbol{\beta}_j, j = 1, \ldots, r$, the voxelwise mean squared error (VMSE) defined as

$$\text{VMSE}_j^{(s)} \quad = \quad \frac{1}{N} \sum_{i=1}^{N} \left( \boldsymbol{\beta}_j(s) - \hat{\boldsymbol{\beta}}_j^{(i)}(s) \right)^2 , \quad s = 1, \ldots, n \, .$$

This measure reflects the error of each coefficient at each voxel $s$, averaged over the simulation runs.

Figure 5 displays the error ratio of two respective estimation procedures on a log scale. The VSME map corresponding to one method is set relative to the VMSE map resulting from the other method. Taking the logarithm leads to a symmetric scale of the ratio. For example, if $\log(\text{VMSE}_{\text{DX3d}}/\text{VMSE}_{\text{ST2}}) = 2$ this corresponds to an error ratio of $\exp(2) = 7.389 \approx 7 : 1$, meaning that method DX3d results in seven times larger errors than the standard ST2 and vice versa.

——— *figure 5 around here* ———

In general, this type of graphic allows to assess both the relative magnitude and the distributional structure of the errors. The emerging color-coded pattern suggests that the standard method mainly results in more biased tensor components *at the ridges* while the basis function approaches lead to larger errors *at the edges* (see bottom row of Fig. 5). This becomes obvious from comparison of the log error ratios (bottom row) with the template structure (top left). Whereas green spots occur above all at true edges and in a second instance in the background, rose

colored points tend to build up artificial ridges which can be associated with spiral echoes. For example, the second slice from below shows a u-shaped semicircle for all true tensor elements. Exactly at the edges of these semicircles, the VCM approach performs worse than the standard procedure (green points). Reversely, the standard procedure exhibits larger errors than the VCM approach not only at the true ridges, but also in the upper half of the slice (rose points). Yet compared to DX3d (bottom left), the TP3d variant (bottom right) seems to lack successful reproduction of the spiral structure: The upper half of the second slice contains slightly more rim artifacts on the basis of TP3d than of DX3d (see green spots). Note that errors of the dark-green range do hardly occur at all whereas we can detect a considerable number of points colored dark-rose. With respect to DX3d (bottom left), the impression arises that the same proportional shift holds for the medium saturated nuances. This can be understood as superiority of VCM approaches over the standard procedure.

Analogous comparison of the two best VCM approaches against each other is included top right in Fig. 5. The prevailing texture of green and rose colored voxels suggests that TP3d does not agglomerate as large errors in artificial clusters as the DX3d method (green), but loses in accuracy at *true* edges (rose).

To conclude, there is an obvious preferability of the VCM approaches compared to the current standard, in particular, if the advantage of inherent approximation and, thus, resolution refinement is considered.

# 5 Application to real data

Besides from being the preferential estimation procedure according to the simulation study, successive smoothing with dimension-specific penalties exhibits several computational advantages for application to real data (see Discussion and Appendix). The data set was shrunken from a recorded number of $128 \times 128$ within plane voxels to $90 \times 75$ due to non-informative background voxels. Furthermore a selection of six subsequent slices was considered sufficient, also if fiber tracking of, for instance, the visual system is intended. From the remaining volume of human brain images, the diffusion tensor field was derived using ST1, ST2 and DX3d. The same knot density as in the simulation study served to maintain important biological features. In total, $72 \times 60 \times 5 \times 6$ coefficients needed to be determined in the VCM; this are $113,400$ unknown parameters less than in the standard model of voxelwise regression.

Figure 6 contains the respective results of the first tensor element, i. e. $\hat{\boldsymbol{\beta}}_{1,\text{ST1}}$, $\hat{\boldsymbol{\beta}}_{1,\text{ST2}}$ and $\hat{\boldsymbol{\beta}}_{1,\text{DX3d}}$ with an inferior-superior ordering of the axial slices from left to right. When the focus is on fiber tracking, standard estimation on the basis of voxelwise regression (ST1, top row) is usually regularized to eliminate noise artifacts as in the present manner of Gössl et al. (2002) (ST2, middle row). The bottom slices depict the estimated surface achieved by the VCM application. Looking at the second and sixth column, in particular, reveals an obvious oversmoothing of the

Gaussian kernel causing ventricles (bright blobs) to be severely smeared across neighbouring slices.

——— *figure 6 around here* ———

The described combined algorithm for smoothing and interpolating can enhance typical image post-processing steps in brain mapping as image segmentation and coregistration. Especially non-linear spatial transformation steps needed for the matching of interindividual brain anatomy can benefit from a continuous vector field. Refinement of the observational grid - as also required by each tracking algorithm - is demonstrated for a smaller area superimposed as white sketch (Fig. 6). The box segment comprises three tissue types (cerebro-spinal fluid, gray and white matter) holding distinct diffusivity properties. In practice, some interpolation method leading to ST3 is appended to the standard 2-step procedure in order to allow for quasi-continuous fiber reconstruction. With respect to the alternative VCM, additional evaluation of the basis functions at intermediate points complies with this demand. Figure 7 delineates the close-up of the respective segment layers including two intermediate planes at double resolution: The VCM estimation (bottom) reveals by far more details and achieves a stronger contrast than the standard data preprocessing cascade ST3 (here with tri-linear interpolation, top row).

——— *figure 7 around here* ———

# 6 Discussion

In this paper, we proposed the varying coefficient model with 3d basis functions as alternative estimation method in diffusion tensor imaging. A simulation study proved the superiority of our VCM variants in terms of averaged mean squared error (AMSE) and voxelwise mean squared error (VMSE) to the current standard procedure of voxelwise regression with subsequent regularization. A salient property of a VCM approach with basis functions is the inherent possibility of going beyond the rigid grid resolution limited by the data acquisition process. For real data a higher contrast of the diffusion tensor field could be achieved than with tri-linear interpolation that completes the standard estimation method. We solely succeeded to handle the clinical example of $90 \times 75 \times 6$ selected voxels by applying the sequential and thus computationally advantageous DX3d VCM. Despite the approximative character, it had also turned out to be the 'candidate of choice' in the simulation study. In contrast to a global penalization, the optimization of a dimension-specific tuning parameter showed to account for the commonly anisotropic voxel size more appropriately as evident from slightly smaller mean squared errors.

Linear B-splines were used in all cases although cubic B-splines are conventionally preferred for their quasi-isotropy. For DT application however, the sustainment of important biological features is of primary concern, but likely to be impaired

by too large influence regions of adjacent voxels. Therefore, a lower spline degree seems more appropriate as confirmed from a pilot study. First order difference penalties and a knot at every 1.25 voxel were additionally chosen for both simulated and real data. Unfortunately, the demands on computer memory and computation time load increase sharply with knot density. We faced this challenge by the extensive use of the sparsity of the involved matrices, hence solving the penalized normal equations in an efficient way again. Yet, for 3d images of the original size it will essentially be impossible to explicitly form the necessary tensor products. The sequential methods DXglob and DX3d represent so far the one and only possibility to handle such big problems. See the Appendix for remarks on implementational matters.

Despite all efforts to preserve enough detail, the proposed VCM approach still suffers from so-called Gibbs phenomena as obvious from Fig. 5 of the VMSE quotient. This under- and overshoot around discontinuities or areas with high curvature appears when the reconstruction of a discontinuous or rapidly changing function is aimed for by a set of continuous ones. Similar phenomena are also known from the one- and two-dimensional case, i. e. in nonparametric curve and surface estimation, and various proposals have been made to improve local adaptivity of estimators. Two main concepts seem promising: The first one still relies on spline basis functions but introduces spatially adaptive penalties, see e. g. Ruppert and Carroll (2000), Lang et al. (2002), Brezger and Lang (2005) for

the 1d case, and Brezger et al. (2005) for the 2d case. The latter paper shows that computational demands already grow enormously when extending 1d-methods to two dimensions. A second possibility is to switch to different basis functions such as wavelets and radial basis functions, or to base the model on anisotropic random fields. Though being conceptually published in 2d, all these locally adaptive modifications are again extremely challenging in 3d from an algorithmic point of view.

## Acknowledgements

We gratefully acknowledge Thomas Kneib for valuable comments and Philipp Sämann for advice on neuroscientifical applications. The authors thank the Max-Planck-Institute of Psychiatry (Munich, Germany) for providing us with human DTI data. This study was kindly supported by the collaborative research center SFB386 (Statistical Analysis of Discrete Structures) as part of the German Science Foundation (DFG). The research of Brian Marx was partially supported by National Science Foundation grant: IIS-0326387 and AFOSR grants: F49620-03-1-0238, F49620-03-1-0239, and F49620-03-1-0241, FA9550-05-1-0454.

# Appendix

## Implementational Issues

As the straightforward approach to VCM (Eq. (3.3)) runs into difficulties, we can alleviate computation by taking advantage of the repeated measures structure, i. e. $\mathbf{X}$ does not vary across voxels. Since $\mathbf{X}$ is on grid, the expectation of $\mathbf{Y}$ can be expressed as a tensor product involving $\mathbf{X}$. Hence, we can apply the fast, compact smoothing algorithm of Eilers et al. (2006). Such an approach avoids the computation of large Kronecker products of B-spline bases, and with a trick, this algorithm turns the 3d VCM into smoothing with 4d tensor products. The limiting bottleneck remains the size of the equation system to be solved, which consists of several tens of thousands normal equations in a realistic scenario. The relevant left hand side, namely $\mathbf{U}'\mathbf{U} + \mathbf{P}$, would occupy more than 18 Gb for a realistic number of parameters, i. e. $6 \cdot 32 \cdot 32 \cdot 8 = 49,152$.

We achieve a crucial storage gain by exploiting the sparsity of the involved matrices, making the brain VCM model tractable again. For illustration purpose, Figure 8 includes a spy diagram of row, column, slice, and complete penalty matrices with non-zero entries colored black.

——— *figure 8 around here* ———

Beyond the conspicuously large amount of zeros, the banded structure strikes at

first glance. Since the matrix $\mathbf{B}$ of tensor product basis functions is also sparse and exhibits a band pattern, $\mathbf{U} = \mathbf{B} \otimes \mathbf{X}$ and $\mathbf{U}'\mathbf{U}$ inherit these two properties being unaffected by the addition of the complete penalty. At least, the usage of suitable sparse matrix libraries is therefore indispensable. For example, the aforementioned scenario then becomes feasible again with a required amount of 53 Mb working memory for storing the left hand side of the corresponding equation system. Note that combination with array regression as decsribed in Eilers et al. (2006) and Currie et al. (2006) would require a compressed format for sparse 4d arrays which is, to the best of our knowledge, not available with current software packages.

Concerning the optimization of the smoothing parameter, a greedy grid search with look-up table was performed in the dimension-specific case while a combination of golden section search and successive parabolic interpolation (as available in R; R Development Core Team, 2004) was applied in the one-dimensional case. The greedy grid search finds the minimum GCV correctly on a refined $10 \times 10 \times 10$ log-scaled grid with equidistant exponents, $\lambda_1$ and $\lambda_2$ varying from $10^{-3}$ to $10^{0.5}$, and $\lambda_3$ within the range of $10^{-8}$ to $10^{-3}$. It has to be pointed out that fixing the tuning parameter to the average over the optimal parameter values corresponding to data sets from a small sample of representative healthy subjects, is expected to satisfy the application of our proposed VCM in daily use. Hence, the time-expensive optimization is dispensable in practice.

All routines are implemented in R (R Development Core Team, 2004) and available

on the first author's website (URL http://www.stat.uni-muenchen.de/∼heim). A respective R-package will be posted. To this end, the library Taucs (Toledo, 2003) is incorporated to assess the GCV.

# References


Assaf Y., Freidlin R. Z., Rohde G. K. and Basser P. J. (2002). A new modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine* **52**, 965-978.

Assunção R. (2003). Space varying coefficient models for small area data. *Environmetrics* **14**, 453-473.

Basser P. J. and Jones D. K. (2002). Diffusion-tensor MRI: Theory, experimental design and data analysis – a technical review. *NMR in Biomedicine* **15**, 456-467.

Basser P. J., Matthiello J. and LeBihan D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance* B **103**, 247-254.

Basser P. J. and Pierpaoli C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of*

*Magnetic Resonance* B **111**, 209-219.

Brezger A., Fahrmeir L. and Hennerfeind A. (2005). Locally adaptive Gaussian Markov random fields in human brain mapping. *SBF386 discussion paper* **456**, University of Munich.

Brezger A. and Lang S. (17 Nov 2004). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, doi:10.1016/j.csda.2004.10.011.

Chang L. C., Jones D. K. and Pierpaoli C. (2005). RESTORE: Robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine* **53**, 1088-1095.

Currie I., Durbán M. and Eilers P. (2006). Generalized linear array models with applications to multidimendional smoothing. *Journal of the Royal Statistical Society* B, to appear.

Dierckx P. (1982). A fast algorithm for smoothing data on a rectangular grid while using spline functions. *SIAM Journal on Numerical Analysis* **19**(6), 1286-1304.

Dierckx P. (1993). *Curve and surface fitting with splines.* Oxford: Monographs on numerical analysis, Oxford University Press.

Eilers P. H. C., Currie I. and Durbán M. (2006). Fast and compact smooth-

ing on large multidimensional grids.*Computational Statistics & Data Analysis* **50**, 61-76. doi:10.1016/j.csda.2004.07.008

Eilers P. H. C., Heim S. and Marx B. D. (2005). Varying coefficient tensor models for brain imaging. In: Statistical Solutions to Modern Problems. *Proceedings of the 20th IWSM*, Sydney, Australia; 157-164.

Eilers P. H. C. and Marx B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems* **66**, 159-174.

Fahrmeir L., Kneib T. and Lang S. (2004). Penalized structured additive regression for space-time data: A Bayesian perspective. *Statist. Sinica* **14**, 731-761.

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

Gössl C., Fahrmeir L., Pütz B., Auer L. M. and Auer D. P. (2002). Fiber tracking from DTI using linear state space models: detectability of the pyramidal tract. *NeuroImage* **16**, 378-388.

Hastie T. and Tibshirani R. (1993). Varying-Coefficient Models (with discussion). *Journal of the Royal Statistical Society* B **55**, 757-796.

Heim S., Hahn K., Auer D. P. and Fahrmeir L. (2004). Optimization of fiber tracking in human brain mapping: Statistical challenges. In: Statistical Modelling. *Proceedings of the 19th IWSM*, Florence, Italy; 296-300.

Lang S., Fronk E. M. and Fahrmeir L. (2002). Function estimation with locally adaptive dynamic models. *Computational Statistics* **17**, 479-499.

Mangin J. F., Poupon C., Clark C., Le Bihan D. and Bloch I. (2002). Distortion correction and robust tensor estimation for MR diffusion imaging. *Medical Image Analysis* **6**, 191-198.

Mori S. and Barker P. B. (1999). Diffusion magnetic resonance imaging: Its principle and applications. *The Anatomical Record* **257**, 102-109.

Mori S. and van Zijl P. C. M. (2002). Fiber tracking: principles and strategies - a technical review. *NMR in Biomedicine* **15**, 468-480.

Papadakis N. G., Xing D., Huang C. L.-H., Hall L. and Carpenter T. A. (1999). A comparative study of acquisition schemes for diffusion tensor imaging using MRI. *Journal of Magnetic Resonance* **137**, 67-82.

R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ruppert D. and Carroll R. J. (2000). Spatially-adaptive penalties for spline

fitting. *Australian and New Zealand Journal of Statistics* **42**, 205-223.

Stejskal E. O. and Tanner J. E. (1965). Spin diffusion measurements: Spin echoes in the presence of time-dependent field gradient. *The Journal of Chemical Physics* **42**, 288-292.

Toledo S., Rotkin V. and Chen D. (2003). TAUCS: A library of sparse linear solvers. Tel-Aviv University, URL http://www.tau.ac.il/∼stoledo/taucs/

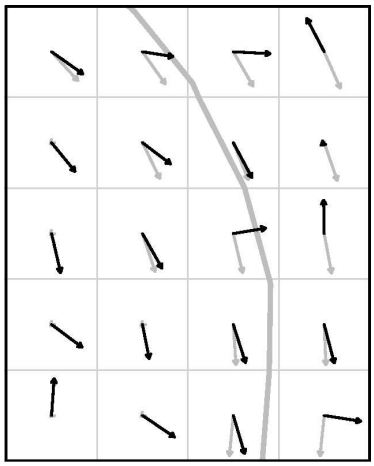
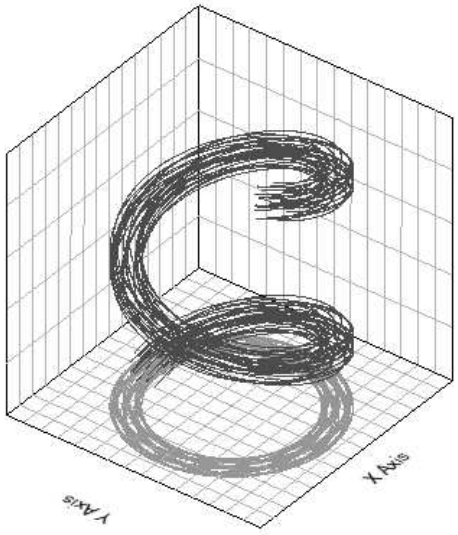


**Figure 1:** Vector map of template (gray) and recorded (black) principal diffusivity directions; the underlying fiber tract is indicated as gray curve. Note, at the middle fiber-transit voxel, the severe distortion of the main eigenvector due to noise.




- 3d data grid of $\{15 \times 15 \times 5\} \subset \mathbb{R}^3$ voxels
- $2 \times 2 \times 4$ mm$^3$ voxel size
- six 3d varying coefficient surfaces
- spiral tensors are anisotropic;
  background tensors are isotropic
- simulated Gaussian error with $\sigma = 10$

**Figure 2:** Design of the simulation study and geometry of the underlying fiber bundle.

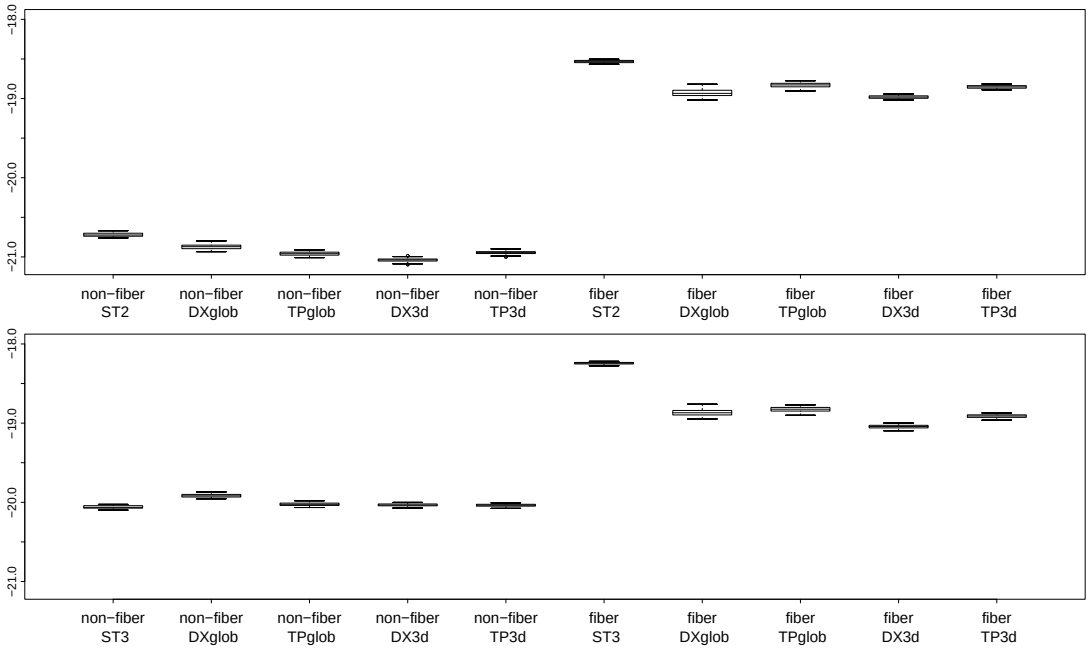



**Figure 3:** Log AMSE of all six tensor elements estimated by different approaches and grouped according to background (left) and spiral voxels (right). Boxplots comprise the voxel volume at original (top) and duplicated (bottom) resolution.

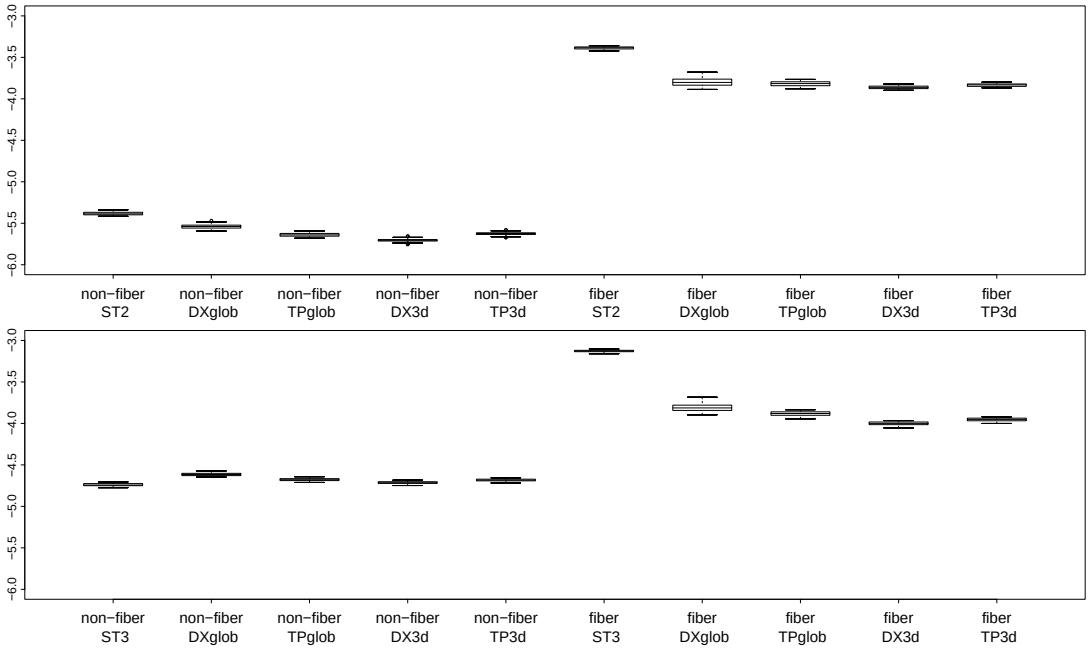



**Figure 4:** Log AMSE of FA values for different estimation approaches based on the original (top) and the duplicated resolution (bottom). Distinction refers to background (left) and spiral voxels (right).

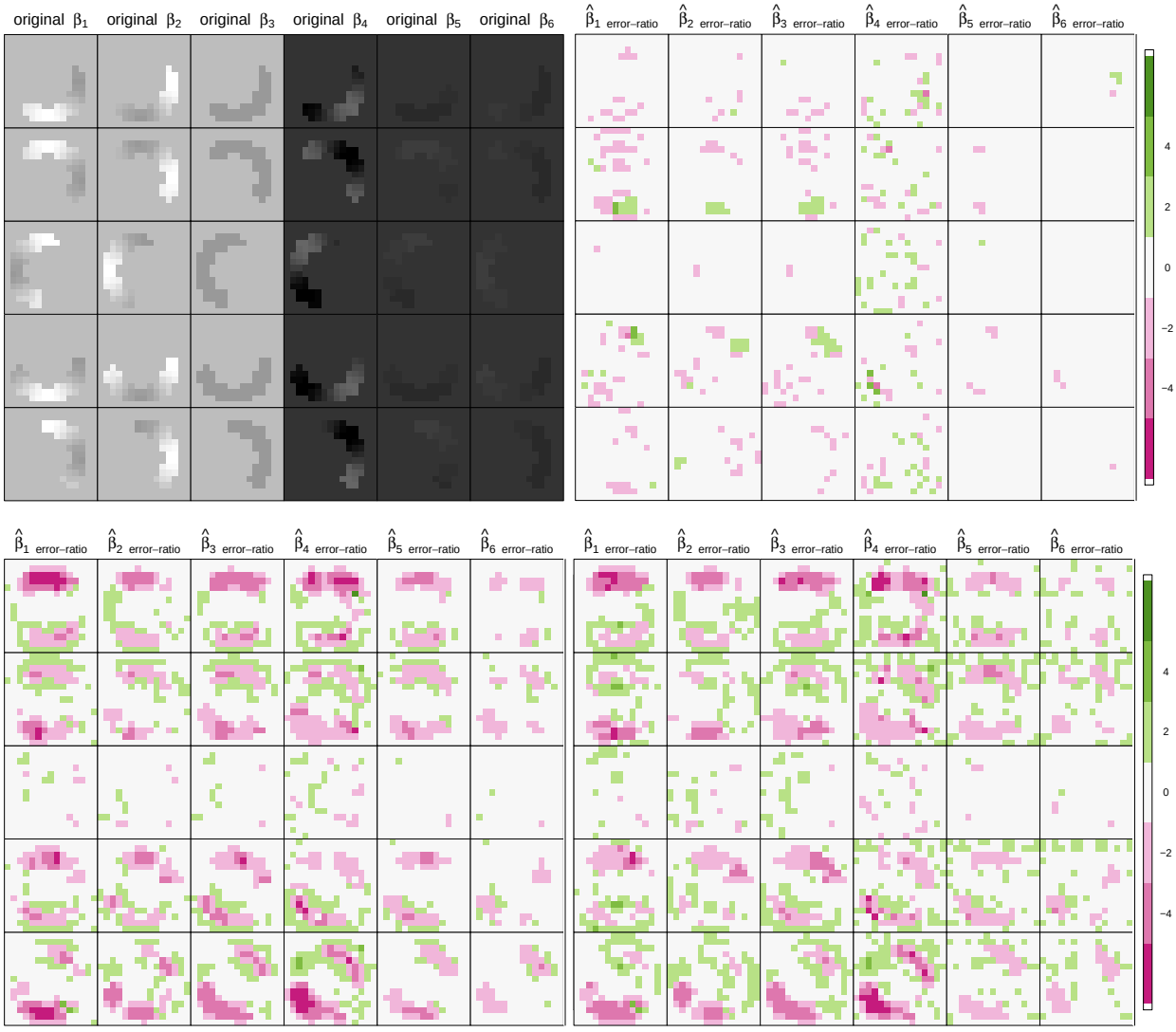



**Figure 5:** Template coefficient surfaces (top left) correspond to the six elements of the 3d tensor field. Log ratio of VMSE is given for both DX3d (bottom left) and TP3d approach (bottom right) relative to ST2. Top right shows the log ratio of VMSE from DX3d versus TP3d.

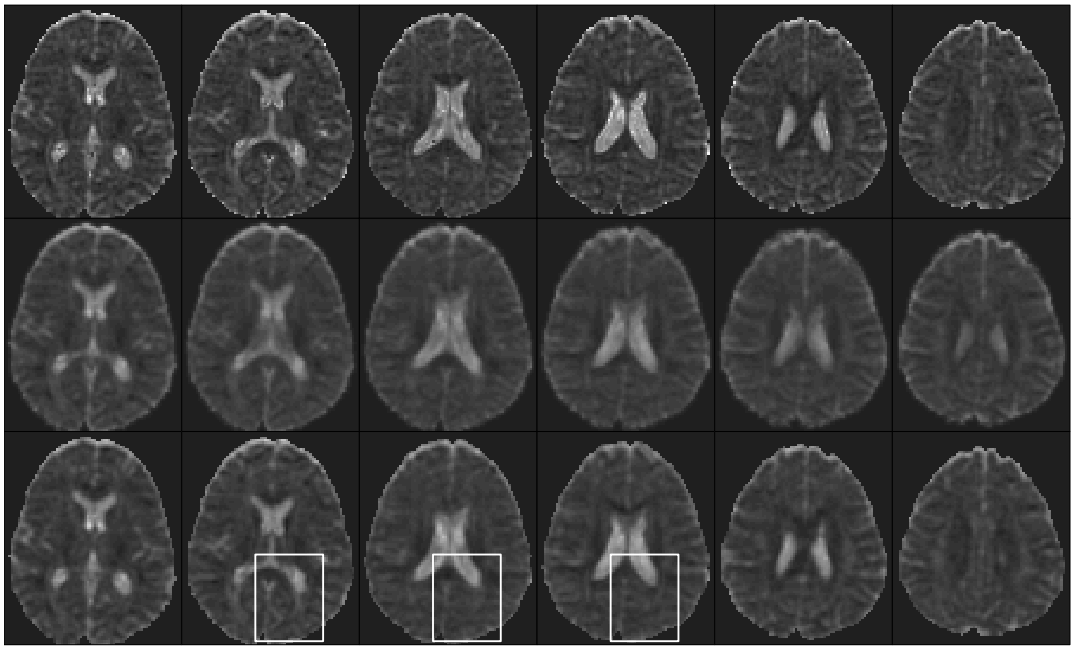


**Figure 6:** Second diagonal element of the diffusion tensor estimated by mere voxelwise regression (top row), plus subsequent regularization with a Gaussian kernel (middle row) and by applying the VCM with DX3d (bottom row). The white rectangulars indicate the segment that was interpolated to double resolution as shown in Fig. 7.

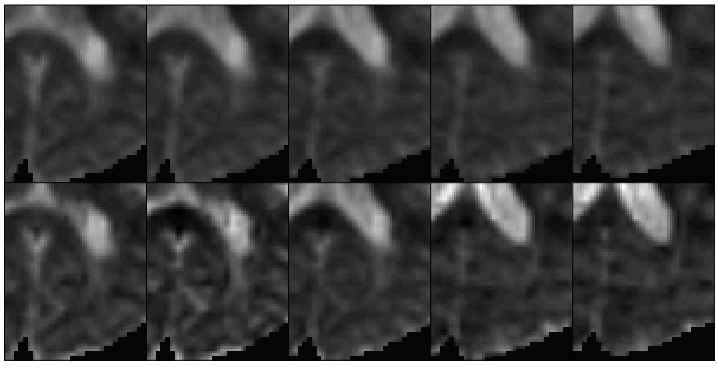


**Figure 7:** Top row represents the second diagonal element of the diffusion tensor at duplicated resolution when the current 3-step data processing is applied. Juxtaposed are the corresponding results yielded by the sophisticated VCM model (bottom row).

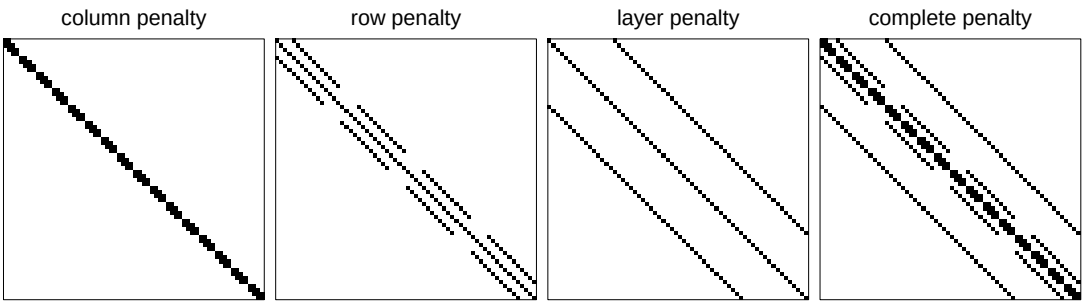



**Figure 8:** Penalty matrices corresponding to $3 \times 3 \times 3$ knots, quadratic B-splines, first order of difference penalties.