Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Image Analysis and Statistical Inference in Neuroimaging with R

Karsten Tabelow[1], Jon D. Clayden[2], Pierre Lafaye de Micheaux[3], Jörg Polzehl[1], Volker J. Schmid[4], Brandon Whitcher[5]

submitted: December 15, 2010

[1] Weierstrass Institute
Mohrenstr. 39, 10117 Berlin
Germany
E-Mail: karsten.tabelow@wias-berlin.de
E-Mail: joerg.polzehl@wias-berlin.de

[2] UCL Institute of Child Health
30 Guilford Street London WC1N 1EH
United Kingdom
E-Mail: jon.clayden@gmail.com

[3] Mathematics and Statistics Department
Université de Montréal
CP 6128 succ. Centre-ville
Montréal Québec H3C 3J7
Canada
E-Mail: lafaye@dms.umontreal.ca

[4] Department of Statistics
Ludwig-Maximilians-Universität
Ludwigstr. 33, 80539 München
Germany
E-Mail:
volker.schmid@stat.uni-muenchen.de

[5] GlaxoSmithKline
Clinical Imaging Centre
London, United Kingdom
E-Mail: Brandon.J.Whitcher@gsk.com

No. 1578
Berlin 2010

2010 *Mathematics Subject Classification.* 62P10, 92C55, 62H35, 62G05 .

Key words and phrases. diffusion weighted imaging, functional magnetic resonance imaging, dynamic contrast-enhanced magnetic resonance imaging, computational neuroscience.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

R is a language and environment for statistical computing and graphics. It can be considered an alternative implementation of the S language developed in the 1970s and 1980s for data analysis and graphics (Becker and Chambers, 1984; Becker et al., 1988). The R language is part of the GNU project and offers versions that compile and run on almost every major operating system currently available. We highlight several R packages built specifically for the analysis of neuroimaging data in the context of functional MRI, diffusion tensor imaging, and dynamic contrast-enhanced MRI. We review their methodology and give an overview of their capabilities for neuroimaging. In addition we summarize some of the current activities in the area of neuroimaging software development in R.

Introduction

The rapid progress of research in the neuroscience and neuroimaging fields has been accompanied by the development of many excellent analysis software tools. These are implemented in a variety of computer languages and programming environments, such as Matlab, IDL, Python, C/C++ and others. This diversity has developed over time through a combination of user preferences and the strengths/weaknesses of the computing environments. Many of these software tools are freely available, like SPM (Ashburner et al., 2008), AFNI (Cox, 1996), FSL (Smith et al., 2004), Freesurfer, or BrainVISA, some are commercial like BrainVoyager and Analyze. Typically, the software tools can be extended by the user to fit their needs in the data analysis. For example, the NIPY project is written in Python and explicitly allows for the integration, modification and extension of code (Millman and Brett, 2007).

R is a free software environment for statistical computing and graphics (Ihaka and Gentleman, 1996; R Development Core Team, 2010). It compiles and runs on almost every UNIX platform, Windows, and Mac OS. Access to R, and its packages/extensions, is provided via the Comprehensive R Archive Network (CRAN, `http://cran.r-project.org`) and R-Forge (`https://r-forge.r-project.org`). R provides a wide range of statistical (linear and nonlinear regression modeling, classical statistical tests, time-series analysis, classification, clustering, etc.) and graphical techniques, and is highly extensible. As of November 2010, the CRAN package repository features over 2600 separate packages contributed by R users. A recent community website (`http://crantastic.org`) provides the facilities to search for, review and tag CRAN packages. Several mailing lists are maintained in order to provide updates and access to literally thousands of R users. This is in addition to a complete set of open-access manuals about the R language.

Why is it worthwhile to consider another programming environment for neuroimaging? R is the free and platform-independent quasi-standard computational environment within the statistics

community. R grants access to many well-developed statistical tools needed for the analysis of neuroimaging data. R offers the use of other software from within R and R can be used within a more general workflow. Finally, special R packages for neuroimaging provide enhanced functionality which is not available elsewhere.

In this paper, we report on several R packages built for the analysis of neuroimaging data in various contexts. We will shortly review the methodology of each package to give an impression of their capabilities for neuroimaging. We will present results created using the packages. The paper is organized by the type of data which is to be analyzed and covers functional Magnetic Resonance Imaging (fMRI), Diffusion Tensor Imaging (DTI), and Dynamic Contrast-Enhanced MRI (DCE-MRI). In the discussion we want to summarize some of the activities in the area of neuroimaging software in R in general.

On features of R

R is a high-level programming environment which provides both command-line and graphical interfaces. It allows for interactive data analysis as well as scripting in order to process large amounts of data offline. R is the quasi-standard environment for the development of new statistical methods, often overtaking commercial statistical packages, such as, SAS, SPSS, etc. Yet its matrix computational capabilities also compare well with those in, for example, Matlab.

The concept of packages as extensions to the R base system is one of its greatest strengths. R easily integrates compiled code written in low-level languages, like C/C++ or FORTRAN, providing the basis for efficient programming of computationally expensive algorithms combined with an easy-to-use interface at the scripting level.

Memory usage in R can be a source of difficulty when handling neuroimaging data sets, since the multi-dimensional data arrays are generally stored in double precision, and in computer main memory. However, there are several projects and packages for R to overcome the memory limitations for very large data sets. One of them is the **ff** (Adler et al., 2010) which provides flexible data structures that are stored on disk but behave (almost) as if they were in RAM by transparently mapping only a section (page-size) in main memory. Another example is the BigMemory project (`http://www.bigmemory.org`) which supports the creation, storage, access, and manipulation of massive matrices. Work is ongoing to incorporate these capabilities into the packages described in this paper. A reference type has recently been introduced in R (2.12.0) which will minimize the need to generate a copy of the data every time a function is called.

R supports modern concepts like object-oriented programming, which is provided as implementation of so-called S4-classes and methods (Chambers, 2008). While not as strict as in low-level languages like C/C++ it allows for class/method definition, data encapsulation, polymorphism, etc.

The analysis of huge data sets with sophisticated methodology comes at the cost of large computational expenses. R provides the possibility to parallelize code natively in R, or within the C/C++ code. It is also possible to include computation on a graphic card (GPU) instead of CPU,

which in appropriate cases may benefit by orders of magnitude from the speed of modern GPUs.

Data Input/Output

The industry standard format, for data coming off a clinical imaging device, is DICOM (Digital Imaging and Communications in Medicine, `http://medical.nema.org`). The DICOM "standard" is very broad and very complicated. Roughly speaking each DICOM-compliant file is a collection of fields organized into two four-byte sequences (group,element) that are represented as hexadecimal numbers and form a tag. The (group,element) combination announces what type of information is coming next. There is no fixed number of bytes for a DICOM header. The packages **oro.dicom** (Whitcher, 2010), **fmri** (Tabelow and Polzehl, 2010c) and **tractor.base** (Clayden, 2010b) provide R functions that read DICOM files and facilitate their conversion to ANALYZE or NIfTI formats.

Although the industry standard for medical imaging data is DICOM, another format has come to be heavily used in the image analysis community. The ANALYZE format was originally developed in conjunction with an image processing system (of the same name) at the Mayo Foundation (Biomedical Imaging Resource, 2001). An Analyze (7.5) format image is comprised of two files, the "img" and "hdr" files, that contain the data and information about the acquisition itself. A more recent adaption of this format is known as NIfTI-1 and is a product of the Data Format Working Group (DFWG) from the Neuroimaging Informatics Technology Initiative (NIfTI). The NIfTI-1 data format is almost identical to the ANALYZE format, but offers a few improvements: merging of the header and image information into one file (.nii), re-organization of the 348-byte fixed header into more relevant categories and the possibility of extending the header information. The packages **AnalyzeFMRI** (Marchini and Lafaye de Micheaux, 2010), **fmri**, **tractor.base**, and **oro.nifti** (Whitcher et al., 2010) all provide functions that read/write ANALYZE and NIfTI files.

Additionally, the **fmri** package provides capabilities to read and write AFNI's HEAD/BRIK files.

Functional MRI

Functional magnetic resonance imaging (fMRI) has become the most informative tool for *in vivo* examination of human brain function on small spatial scales. It is nowadays utilized both in research as well as in clinical applications such as diagnosis, surgical planning and the treatment of brain lesions.

Package AnalyzeFMRI

AnalyzeFMRI is a package originally written by Marchini (2002) for the processing and analysis of large structural and functional MRI data sets under the ANALYZE format. It has been updated since 2007 to include new functionalities: conversion from ANALYZE to NIfTI, complete

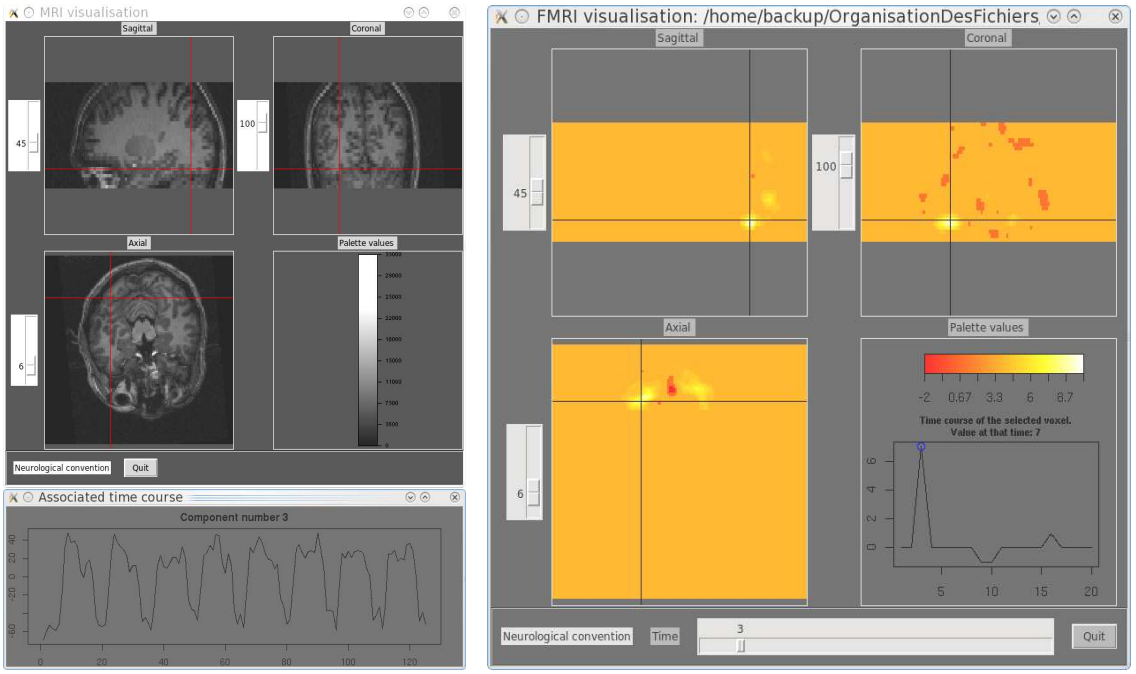

Figure 1 – *Visualization of temporal ICA results. Bottom left: One of the extracted temporal components. Right: Its associated spatial map of activations. Top left: corresponding anatomical image*

NIfTI input/output, functions to obtain spatial coordinates from voxel indices and vice-versa, various geometrical utilities, cross-platform visualization based on Tcl/Tk components, and spatial/temporal independent component analysis (ICA) via a graphical user interface (GUI).

Independent component analysis is a statistical technique that can recover hidden underlying source signals from an observed mixture of these sources (Hyvärinen et al., 2001). The only hypothesis made to solve this problem, known as the blind source separation problem, is that the sources are statistically mutually independent and not Gaussian. ICA is now used to analyze fMRI data since the late 1990's (McKeown et al., 1998) and has been detailed in many papers (see e.g. Stone, 2002; Thomas et al., 2002). Note that ICA can be defined in two dual approaches: spatial ICA (sICA) and temporal ICA (tICA), the latter is not really used in fMRI studies due to the computational difficulties in diagonalizing the large correlation matrix. This is why as far as we know tICA has never been applied to an entire fMRI acquisition, but only on a few slices of the brain or on a very reduced portion of it (Calhoun et al., 2001; Seifritz et al., 2002; Hu et al., 2005). Yet, supposing that we have temporally independent source signals can be seen as a natural assumption in several fMRI studies. The R package **AnalyzeFMRI** uses a property of the singular value decomposition that permits one to obtain the non-zero eigenvalues of the aforementioned correlation matrix and their associated eigenvectors. It then becomes feasible to perform tICA for fMRI data on the whole brain volume. The theoretical details, as well as a complete analysis of simulated and real data, can be found in Bordier et al. (2010).

Results from a tICA are presented in Figure 1, and have been obtained using the GUI in **AnalyzeFMRI**. Note that a sICA also could have been performed. The data used here are a $128\times128\times30$ array of voxels, with a spatial resolution of $1.6875\times1.6875\times3\text{mm}^3$, coming

from a typical fMRI block design visual experiment by alternating the presentation of a set of chromatic rectangles (Mondrian like patterns) and the same patterns in an achromatic version. A mask file ($128\times33\times30$) has been used to limit the search of the associated spatial maps of temporal components in interesting areas, although this is not a required step for the tICA method. The number of temporal components to be extracted has been estimated to be 20 using PCA. Then a cross-correlation between each extracted component and the temporal signal of the stimulus was performed: the highest correlation was associated with the third component, which is displayed in Figure 1. Its associated spatial activation map is also shown on the right of Figure 1. Note that the GUI visualization tool also displays the anatomical images, thus enabling the anatomical localization of the most active voxels via a cross-clicking link with ICA spatial maps. These are located in the ventral cortical region V4-V8, known to be sensitive to color perception.

Package fmri

The R package **fmri** adopts the common view (Friston et al., 1995; Worsley et al., 2002) of a linear model for the time series Y_i in each voxel i

$$Y_i = X\beta_i + \varepsilon_i, \tag{1}$$

where X denotes the design matrix and ε_i the error vector (Polzehl and Tabelow, 2007). The package requires pre-processing steps such as motion correction, registration, and normalization to be performed by third-party tools. Note, that within the workflow using the package smoothing is not considered to be pre-processing. Smoothing as a pre-processing step is prone to a loss in detailed information that is needed in the structural adaptation approaches employed within **fmri**. The fMRI data should therefore not be smoothed in advance.

The package includes functions for input/output of data in standard imaging formats (ANALYZE, NIfTI, AFNI, DICOM). Linear modeling of the data according to (1) includes the description of temporal correlations with an autoregressive AR(1) model. The estimated correlation parameters are bias corrected (Worsley et al., 2002) and can be smoothed (Worsley, 2005). The linear model of the pre-whitened data results in a statistical parametric map (SPM) which is a voxel-wise array of the estimated parameter β and its estimated variance. This information is then used to perform a structural adaptive smoothing (Tabelow et al., 2006). The result of this algorithm is twofold: it improves the estimates of β by reducing their variance and it simplifies the inherent multiple testing problem by introducing a specified smoothness under the null hypothesis. The **fmri** package uses random field theory to determine thresholds for the test statistics (Worsley et al., 1996). The method allows for a significant signal enhancement and reduction of false positive detections without reducing the effective spatial resolution, in contrast to traditional non-adaptive smoothing methods. Note, that the smoothing method accounts for the intrinsic spatial correlation of the data (Tabelow et al., 2006). This has been demonstrated in a series of papers (Voss et al., 2007; Tabelow et al., 2008b), especially for the analysis of high-resolution functional MRI (Tabelow et al., 2009).

The structural adaptive smoothing algorithm has been refined and now integrates both the smoothing and the signal detection step (Polzehl et al., 2010). It is based on a multi-scale

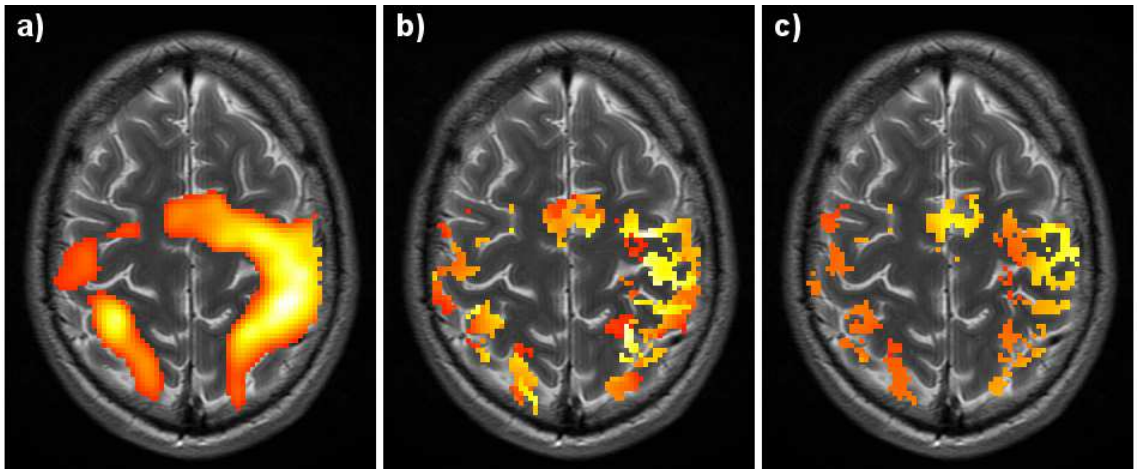

Figure 2 – *Signal detection in the package* **fmri** *using different smoothing methods (multiple test corrected $p = 0.05$): a) signal detection using Gaussian smoothing, b) signal detection using structural adaptive smoothing with signal detection by random field theory, c) results using structural adaptive segmentation. Since the algorithm provides only two segments (activation/no-activation) additional information can be overlayed in color. Here the estimated β is shown. See Polzehl et al. (2010) for more details.*

test performed in the iterative smoothing procedure. The algorithm has been named structural adaptive segmentation since it divides the region of interest into a segment of no activation and a segment where the null hypotheses has been rejected. Figure 2 shows the signal detection results using different smoothing methods. The images were produced using the packages **fmri** and **adimpro** (Tabelow and Polzehl, 2010a).

It is worth noting that the computational time for a complete single-subject fMRI analysis including linear modeling, smoothing signal detection and graphical output is usually in the order of one minute. Although the structural adaptive smoothing algorithm is in principle computationally expensive through its iterative nature, the fact that it operates on the estimated SPM rather than the individual volumes makes it very efficient.

Diffusion Tensor Imaging (DTI)

While functional MRI focuses on the brain's gray-matter functionality in cortical regions, diffusion weighted imaging (DWI) measures directional water diffusion which is highly anisotropic in the brain's white matter. Among the models for DWI data is the widely used diffusion tensor model, where the directional dependence is described by a local diffusion tensor. The estimated anisotropy can be directly associated with the anatomical structure in the brain; mainly (but not solely) with white matter fiber structure.

Package dti

The R package **dti** has been written for the analysis of diffusion-weighted MRI data (Tabelow and Polzehl, 2010b). Using a Gaussian model of diffusion, the data can be described by a rank-2 diffusion tensor $\mathcal{D}$, which is represented by a symmetric positive definite 3×3 matrix

$$\mathcal{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} . \tag{2}$$

Assuming tissue homogeneity within a voxel of the diffusion-weighted image, signal intensities S_b associated with a gradient direction $\vec{g}$ and a b-value b are related to unweighted signal intensities S_0 by

$$S_b = S_0 e^{-b\vec{g}^\top \mathcal{D}\vec{g}}. \tag{3}$$

This model is known as diffusion tensor imaging (DTI, Basser et al., 1994a,b).

The package **dti** uses the package **fmri** for reading diffusion-weighted MRI data from DICOM or NIfTI files. The package provides estimates of diffusion tensors using non-linear regression (Polzehl and Tabelow, 2009) or a linearization of (3). Inference on the diffusion tensor is provided by estimating rotationally invariant tensor characteristics like mean diffusivity, fractional anisotropy (FA), main diffusion direction, etc. The package can be used (in combination with the R package **adimpro** and ImageMagick) to create publication ready images of color-coded directional FA maps or 3D tensor visualizations in common image formats like JPEG, PNG and many others.

One key feature of the **dti** package is the implementation of a structural adaptive smoothing method for the analysis of diffusion-weighted data in the context of the DTI model. Due to its edge-preserving properties these smoothing methods are capable of reducing noise without compromising significant structures like fiber tracts (Tabelow et al., 2008a). Smoothing is performed directly on the diffusion-weighted images using information from the low-dimensional space of the diffusion tensors for adaptation of weights. At each iteration the diffusion tensor is re-estimated from the smoothed diffusion-weighted images. The iterative process is performed from small to larger scales. In contrast to other smoothing methods based on Partial Differential Equations (Ding et al., 2005) the method exhibits an intrinsic stopping criterion. See Figure 3 for color-coded FA images of one slice before and after smoothing.

The **dti** package also provides methods for the analysis of high angular resolution diffusion imaging (HARDI) data (Tabelow et al., 2010). The reconstruction of the orientation distribution function (ODF) from a spherical harmonic expansion of the diffusion-weighted data can be performed as well as the estimation of mixed-tensor models and the expansion of the weighted ODF into central angular Gaussian distribution functions. The package also implements a streamline fiber tracking algorithm for single- and mixed-tensor models. Three-dimensional visualization is provided using the package **rgl** (Adler and Murdoch, 2010) that provides an interface to the OpenGL specification (`http://www.opengl.org`).

The package is completely written using the S4 object-oriented model in R. As for the package **fmri**, the implementation of structural adaptive smoothing methods is potentially computationally expensive through its iterative nature. A typical DTI analysis, including structural adaptive

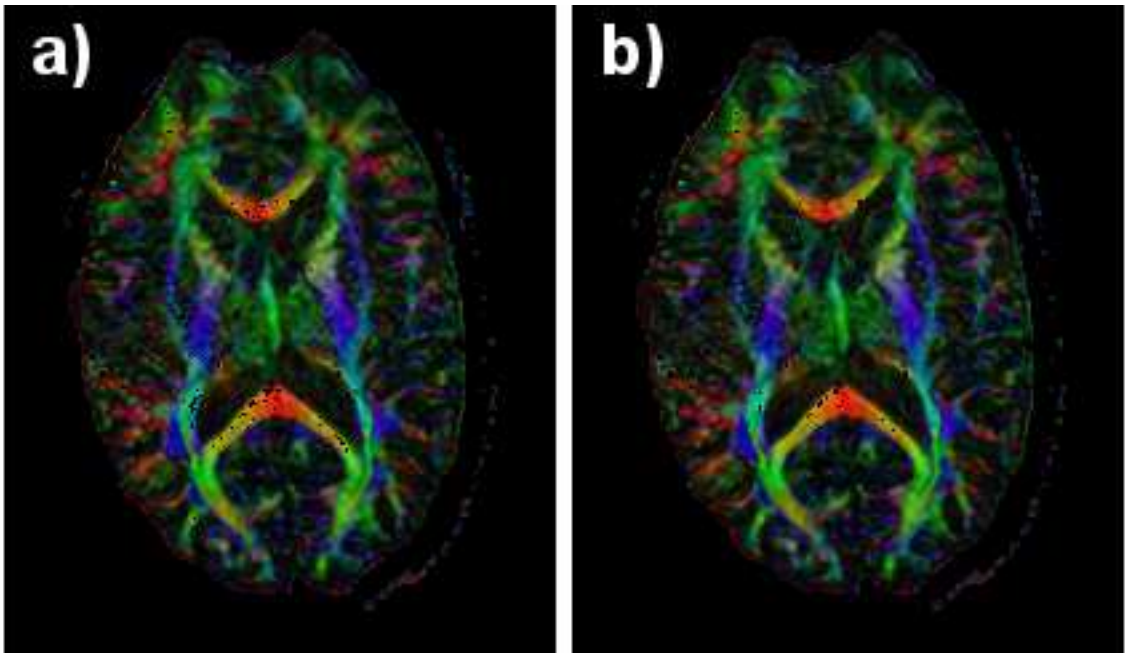

Figure 3 – *Application of the structural adaptive smoothing algorithm in (Tabelow et al., 2008a) to a brain scan: a) color coded directional map weighted with FA of an axial slice obtained by voxel-wise analysis of the DWI data consisting of 55 diffusion weighted images and b) directional map resulting from structural adaptive smoothing. In all images, black regions inside the brain denote areas in which at least one of the eigenvalues was negative. The color coding is red for right-left, green for anterior-posterior and blue for inferior-superior. See Tabelow et al. (2008a) for more details.*

smoothing, takes approximately 30 minutes on common hardware. However, as the amount of data for diffusion-weighted imaging is rather large, a large amount of memory is advisable. This issue will be solved in future versions of the package by using efficient memory management via **ff** (Adler et al., 2010) or **bigmemory** (Kane and Emerson, 2010). For a more complete survey on algorithmic and computational details we refer the reader to Polzehl and Tabelow (2009).

The TractoR project

The TractoR (Tractography with R) project provides tools for working with diffusion MRI and fiber tractography, with a strong focus on group-wise analysis. The project is currently built upon four R packages, but also provides an interface for performing common tasks without direct interaction with R. Full source code is available at the project web site (`http://code.google.com/p/tractor`).

The **tractor.base** package provides data structures and functions for reading images from DICOM, ANALYZE or NIfTI storage formats, visualizing and manipulating images—for example, by thresholding or masking—and writing images back to file. The second package **tractor.utils** provides a variety of utility functions, primarily for use by the project interface. The **tractor.session** package provides a file hierarchy abstraction, designed to facilitate working with a large number of data sets; and also provides R interfaces to third-party image analysis software pack-

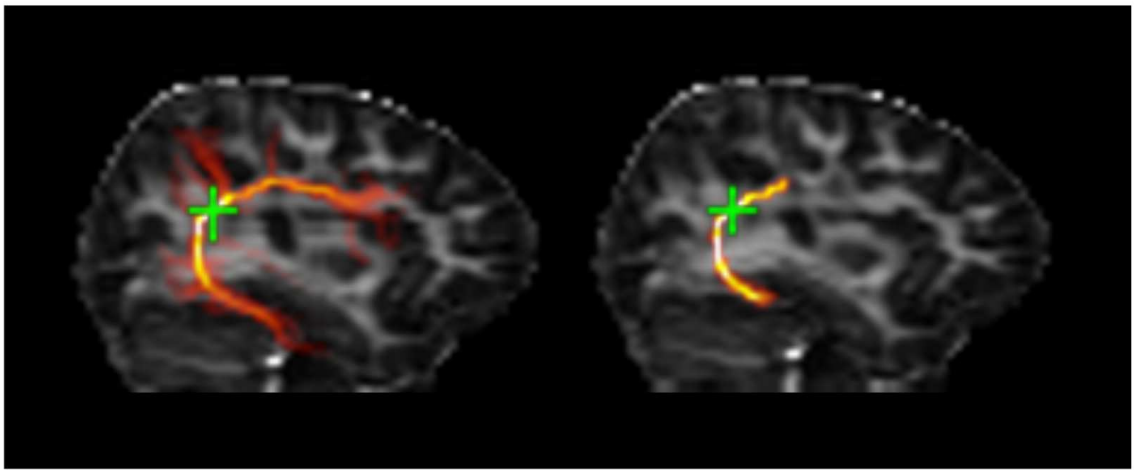

Figure 4 – *Rejection of false positive streamlines using a predefined reference tract and statistical shape model (Clayden et al., 2009a). After "pruning" (right) the main curve of the arcuate fasciculus appears without any superfluous branching structures.*

ages, including FSL (Smith et al., 2004) and Camino (Cook et al., 2006). Finally, the **tractor.nt** package provides reference implementations of "neighborhood tractography" methods, which use anatomical prior information and probabilistic models to segment white-matter structures in groups with high robustness and consistency.

Neighborhood tractography overcomes a number of problems with standard diffusion tractography methods, particularly the difficulty of choosing suitable seed points to initialize tracking, by introducing anatomical prior information in the form of reference tracts (Clayden et al., 2006, 2007; Muñoz Maniega et al., 2008). In this way, high levels of reproducibility and consistency can be achieved for diffusion-based measurements within white matter tracts, without time-consuming and error-prone manual intervention (Clayden et al., 2009b). Finally, the approach of explicitly modeling tract shape variability across individuals offers a natural solution to the common problem of false positive pathways generated by tractography methods (Fig. 4; Clayden et al., 2009a). The statistical pedigree of R makes it an ideal environment in which to develop and apply machine learning techniques for neuroimaging.

Dynamic Contrast-Enhanced MRI (DCE-MRI)

The **dcemriS4** package contains a collection of data classes and functions to perform quantitative analysis from a dynamic contrast-enhanced MRI (DCE-MRI) acquisition on a voxel-by-voxel basis. Patients undergoing a DCE-MRI acquisition have several minutes of T1-weighted scans performed, with a typical temporal resolution between 3-15 seconds, where a bolus of gadolinium is injected after a sufficient number of pre-contrast scans have been acquired. Assuming that the biology is explained by a system of linear differential equations, the model of contrast agent concentration over time is given by a sum of exponentials convolved with an arterial input function.

The workflow may be defined by the following steps: motion correction and/or co-registration, T1 estimation, conversion of signal intensity to gadolinium contrast-agent concentration and kinetic parameter estimation. The S4 object classes for common medical image formats, specifically ANALYZE and NIfTI, are provided by the **oro.nifti** package (Whitcher et al., 2010) along with the ability to extend the NIfTI data format header via extensions. Users are allowed to add extensions to newly-created NIfTI S4 objects using various functions and the **XML** package (Temple Lang, 2010). All operations that are performed on a NIfTI object will generate a so-called *audit trail* that consists of an XML-based log. Each log entry contains information not only about the function applied to the NIfTI object, but also various system-level information; e.g., version of R, user name, date, time, etc. When writing NIfTI-class objects to disk, the XML-based NIfTI extension is converted into plain text and saved appropriately (ecode = 6).

The estimation of voxel-wise T1 relaxation, and subsequent conversion of the signal intensity to contrast agent concentration for the dynamic acquisition, has been implemented using the relationship between signal intensity and flip angle for spoiled gradient echo (SPGR) sequences and fitting the non-linear curve to all available flip-angle acquisitions (Buckley and Parker, 2005). The estimated T1 values are then used to convert signal intensity into contrast agent concentration $C_t(t)$ for the dynamic acquisition using

$$[\mathsf{Gd}] = \frac{1}{r_1}\left(\frac{1}{T_1} - \frac{1}{T_{10}}\right), \tag{4}$$

where r_1 is the spin-lattice relaxivity constant and T_{10} is the spin-lattice relaxation time in the absence of contrast media (Buckley and Parker, 2005). For computational reasons, we follow the method of Li et al. (2000).

Whereas quantitative PET studies employ arterial cannulation on the subject to characterize the arterial input function (AIF) directly, it has been common to use literature-based AIFs in the DCE-MRI literature. Examples include

$$C_p(t) = D\left(a_1 e^{-m_1 t} + a_2 e^{-m_2 t}\right), \tag{5}$$

where D is the dose of the contrast agent and $\theta = (a_1, m_1, a_2, m_2)$ are parameters taken from the literature (Weinmann et al., 1984; Tofts and Kermode, 1984; Fritz-Hansen et al., 1996). There has been progress in measuring the AIF using the dynamic acquisition and fitting a parametric model to the observed data. The **dcemriS4** package has incorporated these literature-based models and a data-driven model given by

$$C_p(t) = A_B t e^{-\mu_B t} + A_G\left(e^{-\mu_G t} + e^{-\mu_B t}\right) \tag{6}$$

(Orton et al., 2008), which is applied to the observed data using nonlinear regression using the Levenberg-Marquardt algorithm.

A common parametric model for DCE-MRI data is the "extended Kety model" given by

$$C_t(t) = v_p C_p(t) + K^{\mathsf{trans}}\left[C_p(t) \otimes \exp(-k_{\mathsf{ep}} t)\right], \tag{7}$$

where $C_t(t)$ is the concentration of the contrast agent in tissue as a function of time t, v_p is the volume of contrast agent in the plasma, K^{trans} is the transfer rate constant from plasma to EES

(extravascular extracellular space) and k_{ep} is the rate parameter for transport from the EES to plasma (Parker and Buckley, 2005). The parametric model (7) may be applied to all voxels in a pre-specified region of interest (ROI), independently, using two distinct procedures: nonlinear regression with the Levenberg-Marquardt algorithm, using the **minpack.lm** package (Elzhov and Mullen, 2009), and Bayesian estimation (Schmid et al., 2006). Voxel-wise analysis of DCE-MRI can be very time consuming and **dcemriS4** supports parallel computing by incorporating the **multicore** package (Urbanek, 2010). Computations are easily spread across multiple CPUs by setting `multicore=TRUE` in all functions that provide parameter estimates.

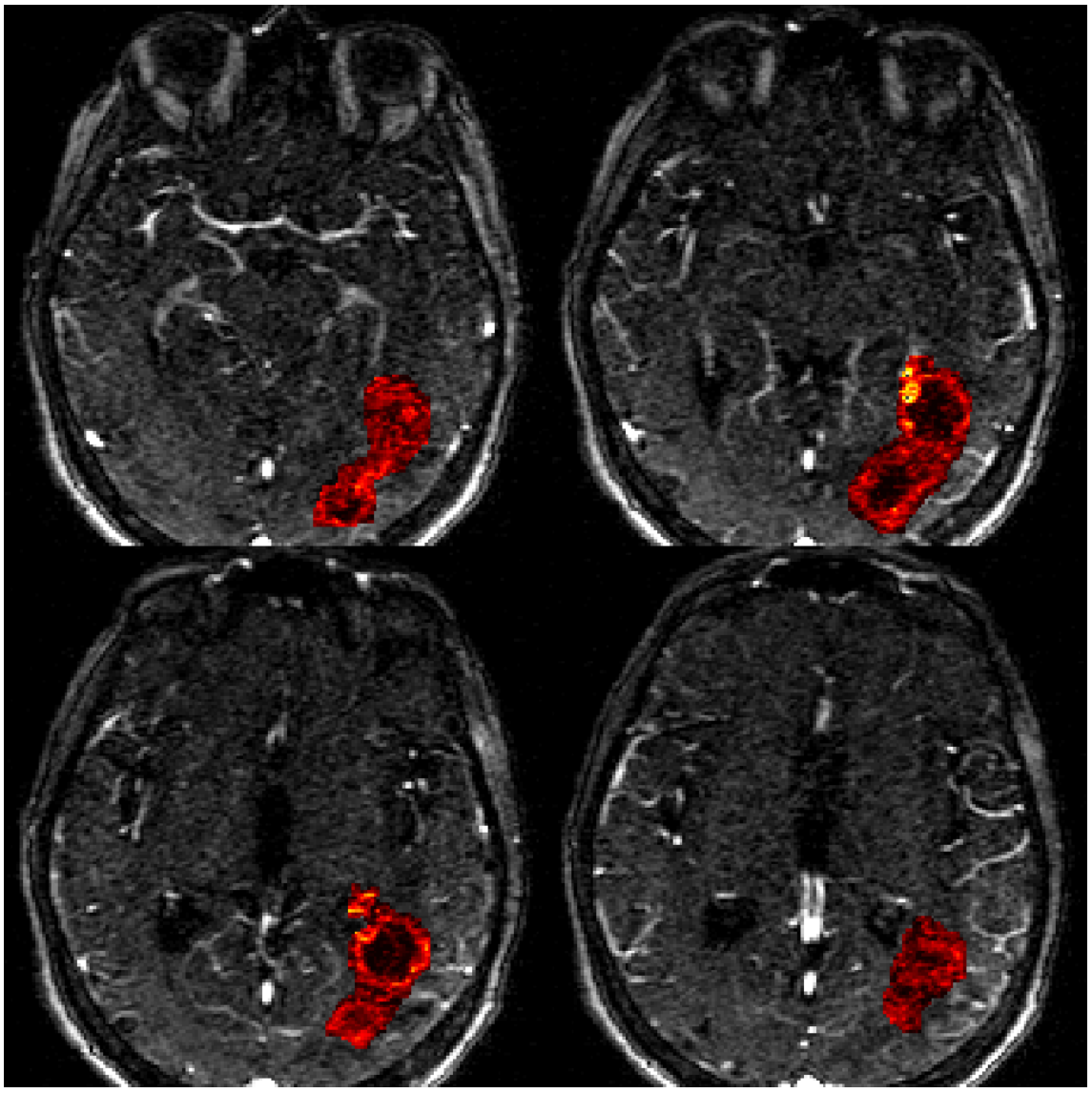

Figure 5 – *Perfusion characteristic K^{trans}, the transfer rate constant from plasma to extravascular extracellular space, estimated via non-linear regression in a dynamic contrast-enhanced MRI acquisition. Displayed values are in the range $[0, 0.15]$ min^{-1}.*

An illustration of the parametric model is provided via the National Biomedical Imaging Archive (NBIA, `http://cabig.nci.nih.gov/tools/NCIA`). Figure 5 displays the estimated

K^{trans} values for a region-of-interest (ROI) approximately covering a brain tumor. The uptake of the contrast agent varies drastically across the tissue in the ROI, exhibiting a hypovascular response in the core of the tumor and a potentially hypervascular response in the tumor rim. While statistical images provide an invaluable tool for exploratory data analysis, longitudinal assessment of disease progression or treatment response would have to be evaluated using a suitable scalar summary of the tumor ROI; e.g., using hierarchical models (Schmid et al., 2009).

Additional R Packages for Neuroimaging

The following paragraphs provide brief descriptions to additional R packages that are relevant to neuroimaging.

The R package **neuroim** provides an object-oriented implementation to manipulate volumetric brain imaging data (Buchsbaum, 2010). The **Rniftilib** package provides read/write capabilities for the NIfTI-1 format (Granert, 2010) using the C reference library provided by the Neuroimaging Informatics Technology Initiative. The aim of this package is to serve as a common basis for working with multi-dimensional volumetric (neuro)imaging data.

PTAk (Principal Tensor Analysis on k modes) is an R package that uses a multi-way method to decompose a tensor (array) of any order, as a generalization of the singular value decomposition (SVD), and supports non-identity metrics and penalizations (Leibovici, 2010). A two-way SVD with these extensions is also available. The package includes additional multi-way methods: PCAn (Tucker-n), PARAFAC/CANDECOMP and FCAk (multi-way correspondence analysis).

An important issue in the analysis of fMRI is how to account for the spatial smoothness of activated regions. The R-Forge project **arf** uses a method to accomplish this by modeling activated regions with Gaussian shapes. Hypothesis tests on the location, spatial extent, and amplitude of these regions are performed instead of hypothesis tests of individual voxels (Weeda et al., 2009).

As previously mentioned R is capable of making use of parallel computing techniques, CPU as well as GPU based. CUDA (Compute Unified Device Architecture) is a software platform for massively parallel high-performance computing on NVIDIA GPUs. The **cudaBayesreg** package provides a CUDA implementation of a Bayesian multilevel model for the analysis of fMRI data (Ferreira da Silva, 2010). The CUDA programming model uses a separate thread for fitting a linear regression model at each voxel in parallel. The global statistical model implements a Gibbs Sampler for hierarchical linear models with a Gaussian prior. This model has been proposed by Rossi et al. (2005, Chapter 3) and is referred to as "rhierLinearModel" in the R package **bayesm** (Rossi, 2010).

General-purpose image registration tools have recently been made available in R through the **RNiftyReg** package (Clayden, 2010a), which provides an interface to the NiftyReg image registration tools (`http://sourceforge.net/projects/niftyreg/`).

Discussion

A number of packages specific to medical imaging, and in particular neuroimaging, have been developed within the R community. In this paper we have concentrated on packages that have reached a level of maturity that guarantees the stability of the package's capabilities. New functionality is expected to be added and the underlying code is expected to improve, especially concerning memory management and computational speed, over time. The packages described in the present paper cover functional magnetic resonance imaging (fMRI), diffusion tensor imaging (DTI), and dynamic contrast-enhanced MRI (DCE-MRI). However, there is much more activity concerning structural and functional neuroimaging in R, with up-to-date information presented in the CRAN Medical Imaging task view (`http://cran.r-project.org/web/views/MedicalImaging.html`). This website provides a brief overview of existing packages and a categorization with respect to their main area of application, without going into the details of methodology or implementation.

Package	Data type	Main features	License
oro.nifti, **oro.dicom**	general	Reading/writing DICOM, NIfTI and ANALYZE files	BSD
AnalyzeFMRI	fMRI	spatial/temporal ICA analysis for fMRI data	GPL ≥ 2
fmri	fMRI	Linear modeling, structural adaptive smoothing, signal detection for single subject fMRI data	GPL ≥ 2
dti	DWI	Diffusion tensor analysis for diffusion-weighted MR data, structural adaptive smoothing, Modeling of HARDI data	GPL ≥ 2
tractor	DWI	Diffusion tensor analysis for diffusion-weighted MR data, probabilistic tractography, segmentation of specific tracts	GPL
dcemriS4	DCE-MRI	Voxel-wise quantitative analysis of dynamic contrast-enhanced or diffusion-weighted MRI data	BSD

Table 1 – *Main features and licenses of the* R *packages described in this paper.*

In Table 1 we summarize the main features of the packages described in the present paper and give the applicable license.

Conclusions

The R language provides an excellent environment for all levels of analysis with neuroimaging data, from basic image processing to advanced statistical techniques via the current list of contributed packages in the Medical Imaging task view. These packages can assist user-guided data analysis for fMRI, DCE-MRI, and DWI data as well as automated bulk analysis of imaging data. The user is free to create additional data structures or analysis routines using the programming environment in R—making it easily customizable. It is very easy to link R to compiled C/C++/Fortran code for computational efficiency. R may be run in either interactive or batch-processing modes in order to scale with the application, and may be combined with other computing environments (e.g., Matlab or NIPY) to allow even greater flexibility.

Acknowledgments

K.T. is supported by the DFG Research Center MATHEON. K.T. and J.P. thank H.U. Voss at the Citigroup Biomedical Imaging Center, Weill Cornell Medical College for providing diffusion-weighted and functional MRI data sets and for numerous intense and helpful discussions on magnetic resonance imaging and related issues. P.LdM. is supported by NSERC and would like to thank M. Dojat from Grenoble Institut des Neurosciences INSERM 836 for providing fMRI data related to color vision experiments and for fruitful discussions on spatial and temporal independent component analysis. V.S. is supported by the LMU innovative project BioMed-S.

References

References

Adler, D., Gläser, C., Nenadic, O., Oehlschlägel, J., Zucchini, W., 2010. ff: memory-efficient storage of large data on disk and fast access functions. R package version 2.2-1.
URL http://cran.r-project.org/package=ff

Adler, D., Murdoch, D., 2010. rgl: 3D visualization device (OpenGL). R package version 0.92.794.
URL http://cran.r-project.org/package=rgl

Ashburner, J., Chen, C. C., Flandin, G., Henson, R., Kiebel, S., Kilner, J., Litvak, V., Moran, R., Penny, W., Stephan, K., Hutton, C., Glauche, V., Mattout, J., Phillips, C., oct. 2008. The SPM8 Manual. Functional Imaging Laboratory, Wellcome Trust Centre for Neuroimaging, Institute of

Neurology, UCL, London.
URL http://www.fil.ion.ucl.ac.uk/spm/

Basser, P. J., Mattiello, J., LeBihan, D., 1994a. Estimation of the effective self-diffusion *tensor* from the NMR spin echo. Journal of Magnetic Resonance 103, 247–254.

Basser, P. J., Mattiello, J., LeBihan, D., 1994b. MR diffusion tensor spectroscopy and imaging. Biophysical Journal 66, 259–267.

Becker, R., Chambers, J., , Wilks, A., 1988. The New S Language: A Programming Environment for Data Analysis and Graphics. Wadsworth & Brooks/Cole, Pacific Grove, California.

Becker, R., Chambers, J., 1984. S: An Interactive Environment for Data Analysis and Graphics. Wadsworth & Brooks/Cole, Pacific Grove, California.

Biomedical Imaging Resource, 2001. Analyze Program. Mayo Foundation.

Bordier, C., Dojat, M., Lafaye de Micheaux, P., 2010. Temporal and spatial independent component analysis for fMRI data sets embedded in a R package. Tech. Rep. 1012.0269v1, arXiv.

Buchsbaum, B., 2010. neuroim: R software for reading, writing and representing brain imaging data. R package version 0.0.1.
URL https://r-forge.r-project.org/projects/neuroim

Buckley, D. L., Parker, G. J. M., 2005. Measuring contrast agent concentration in T_1-weighted dynamic contrast-enhanced MRI. In: Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology. Springer, pp. 69–80.

Calhoun, V. D., Adali, T., Pearlson, G. D., Pekar, J. J., 2001. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Human Brain Mapping 13 (1), 43–53.

Chambers, J. M., 2008. Software for Data Analysis: Programming with R. Springer, New York.

Clayden, J., 2010a. RNiftyReg: Medical image registration using the NiftyReg library. R package version 0.2.0.
URL http://cran.r-project.org/package=RNiftyReg

Clayden, J., 2010b. tractor: Magnetic resonance imaging and tractography with R. Version 1.6.0.
URL http://code.google.com/p/tractor

Clayden, J. D., Bastin, M. E., Storkey, A. J., November 2006. Improved segmentation reproducibility in group tractography using a quantitative tract similarity measure. NeuroImage 33 (2), 482–492.

Clayden, J. D., King, M. D., Clark, C. A., 2009a. Shape modelling for tract selection. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (Eds.), Medical Image Computing and Computer-Assisted Intervention. Vol. 5762 of Lecture Notes in Computer Science. Springer-Verlag, pp. 150–157.

Clayden, J. D., Storkey, A. J., Bastin, M. E., November 2007. A probabilistic model-based approach to consistent white matter tract segmentation. IEEE Transactions on Medical Imaging 26 (11), 1555–1561.

Clayden, J. D., Storkey, A. J., Muñoz Maniega, S., Bastin, M. E., April 2009b. Reproducibility of tract segmentation between sessions using an unsupervised modelling-based approach. NeuroImage 45 (2), 377–385.

Cook, P. A., Bai, Y., Nedjati-Gilani, S., Seunarine, K. K., Hall, M. G., Parker, G. J. M., Alexander, D. C., 2006. Camino: Open-source diffusion-MRI reconstruction and processing. In: Proceedings of the ISMRM 14th Scientific Meeting & Exhibition. International Society for Magnetic Resonance in Medicine, p. 2759.

Cox, R. W., 1996. Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and biomedical research 29, 162–173.

Ding, Z., Gore, J. C., Anderson, A. W., 2005. Reduction of noise in diffusion tensor images using anisotropic smoothing. Magnetic Resonance in Medicine 53 (2), 485–490.

Elzhov, T. V., Mullen, K. M., 2009. minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK. R package version 1.1-4.
URL `http://cran.r-project.org/package=minpack.lm`

Ferreira da Silva, A., 2010. cudaBayesreg: CUDA Parallel Implementation of a Bayesian Multilevel Model for fMRI Data Analysis. R package version 0.3-9.
URL `http://cran.r-project.org/package=cudaBayesreg`

Friston, K., Holmes, A., Worsley, K., Poline, J.-B., Frith, C., Frackowiak, R., 1995. Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2, 189–210.

Fritz-Hansen, T., Rostrup, E., Larsson, H. B. W., Sø ndergaard, L., Ring, P., Henriksen, O., 1996. Measurement of the arterial concentration of Gd-DTPA using MRI: A step toward quantitative perfusion imaging. Magnetic Resonance in Medicine 36 (2), 225–231.

Granert, O., 2010. Rniftilib: Rniftilib - R Interface to NIFTICLIB (V1.1.0). R package version 0.0-29.
URL `http://cran.r-project.org/package=Rniftilib`

Hu, D., Yan, L., Liu, Y., Zhou, Z., Friston, K. J., Tan, C., Wu, D., 2005. Unified SPM-ICA for fMRI analysis. NeuroImage 25 (3), 746–755.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5, 299–314.

Kane, M. J., Emerson, J. W., 2010. bigmemory: Manage massive matrices with shared memory and memory-mapped files. R package version 4.2.3.
URL http://CRAN.R-project.org/package=bigmemory

Leibovici, D., 2010. Spatio-temporal multiway data decomposition using principal tensor analysis on k-modes: The R package PTAk. Journal of Statistical Software 34 (10), 1–34.

Li, K.-L., Zhu, X. P., Waterton, J., Jackson, A., 2000. Improved 3D quantitative mapping of blood volume and endothelial permeability in brain tumors. Journal of Magnetic Resonance Imaging 12, 347–357.

Marchini, J., 2002. Analyzefmri: An R package for the exploration and analysis of MRI and fMRI datasets. R News 2 (1), 17–23.

Marchini, J. L., Lafaye de Micheaux, P., 2010. AnalyzeFMRI: Functions for analysis of fMRI datasets stored in the ANALYZE or NIFTI format. R package version 1.1-12.
URL http://cran.r-project.org/package=AnalyzeFMRI

McKeown, M., Makeig, S., Brown, G., Jung, T., Kindermann, S., Bell, A., Sejnowski, T., 1998. Analysis of fmri data by blind separation into independent spatial components. Human Brain Mapping 6, 160–188.

Millman, K. J., Brett, M., 2007. Analysis of functional magnetic resonance imaging in python. Computing in Science & Engineering 9 (3), 52–55.

Muñoz Maniega, S., Bastin, M. E., McIntosh, A. M., Lawrie, S. M., Clayden, J. D., 2008. Atlas-based reference tracts improve automatic white matter segmentation with neighbourhood tractography. In: Proceedings of the ISMRM 16th Scientific Meeting & Exhibition. International Society for Magnetic Resonance in Medicine, p. 3318.

Orton, M. R., d'Arcy, J. A., Walker-Samuel, S., Hawkes, D. J., Atkinson, D., Collins, D. J., Leach, M. O., 2008. Computationally efficient vascular input function models for quantitative kinetic modelling using DCE-MRI. Physics in Medicine and Biology 53, 1225–1239.

Parker, G. J. M., Buckley, D. L., 2005. Tracer kinetic modelling for T_1-weighted DCE-MRI. In: Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology. Springer, pp. 81–92.

Polzehl, J., Tabelow, K., 2007. fmri: A package for analyzing fmri data. R News 7, 13–17.

Polzehl, J., Tabelow, K., 2009. Structural adaptive smoothing in Diffusion Tensor Imaging: The R package dti. Journal of Statistical Software 31, 1–24.

Polzehl, J., Voss, H., Tabelow, K., 2010. Structural adaptive segmentation for statistical parametric mapping. NeuroImage 52, 515–523.

R Development Core Team, 2010. 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

Rossi, P., 2010. bayesm: Bayesian Inference for Marketing/Micro-econometrics. R package version 2.2-3.
URL http://cran.r-project.org/package=bayesm

Rossi, P., Allenby, G., McCulloch, R., 2005. Bayesian Statistics and Marketing. John Wiley and Sons.

Schmid, V., Whitcher, B., Padhani, A. R., Taylor, N. J., Yang, G.-Z., 2006. Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. IEEE Transactions on Medical Imaging 25 (12), 1627–1636.

Schmid, V. J., Whitcher, B., Padhani, A. R., Taylor, N. J., Yang, G.-Z., 2009. A Bayesian hierarchical model for the analysis of a longitudinal dynamic contrast-enhanced MRI oncology study. Magnetic Resonance in Medicine 61 (1), 163–174.

Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoff, J. G., Bilecen, D., Tedeschi, G., Scheffler, K., Di Salle, F., 2002. Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297 (5587), 1706–1708.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., Matthews, P. M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 Suppl 1, S208–S219.

Stone, J., 2002. Independent component analysis: an introduction. Trends in Cognitive Sciences 6 (2), 59–64.

Tabelow, K., Piëch, V., Polzehl, J., Voss, H. U., 2009. High-resolution fmri: Overcoming the signal-to-noise problem. Journal of Neuroscience Methods 178, 357–365.

Tabelow, K., Polzehl, J., 2010a. adimpro: Adaptive Smoothing of Digital Images. R package version 0.7.5.
URL http://cran.r-project.org/package=adimpro

Tabelow, K., Polzehl, J., 2010b. dti: Analysis of diffusion weighted imaging (DWI) data. R package version 0.9-3.1.
URL http://cran.r-project.org/package=dti

Tabelow, K., Polzehl, J., 2010c. fmri: Analysis of fMRI experiments. R package version 1.4.2.
URL http://cran.r-project.org/package=fmri

Tabelow, K., Polzehl, J., Spokoiny, V., Voss, H. U., 2008a. Diffusion tensor imaging: Structural adaptive smoothing. NeuroImage 39, 1763–1773.

Tabelow, K., Polzehl, J., Ulug, A. M., Dyke, J. P., Watts, R., Heier, L. A., Voss, H. U., 2008b. Accurate localization of brain activity in presurgical fmri by structure adaptive smoothing. IEEE Transactions on Medical Imaging 27, 531–537.

Tabelow, K., Polzehl, J., Voss, H. U., Spokoiny, V., 2006. Analyzing fmri experiments with structural adaptive smoothing procedures. NeuroImage 33, 55–62.

Tabelow, K., Voss, H. U., Polzehl, J., 2010. Modeling the orientation distribution function by mixtures of angular central gaussian distributions. Tech. Rep. 1559, WIAS, Berlin.

Temple Lang, D., 2010. XML: Tools for parsing and generating XML within R and S-Plus. R package version 3.1-0.
URL http://cran.r-project.org/package=XML

Thomas, C., Harshman, R., Menon, R., 2002. Noise reduction in BOLD-based fMRI using component analysis. NeuroImage 17, 1521–1537.

Tofts, P. S., Kermode, A. G., 1984. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. Magnetic Resonance in Medicine 17 (2), 357–367.

Urbanek, S., 2010. multicore: Parallel processing of R code on machines with multiple cores or CPUs. R package version 0.1-3.
URL http://www.rforge.net/multicore

Voss, H. U., Tabelow, K., Polzehl, J., Tchernichovski, O., Maul, K. K., Salgado-Commissariat, D., Ballon, D., Helekar., S. A., 2007. Functional MRI of the zebra finch brain during song stimulation suggests a lateralized response topography. Proceedings of the National Academy of Sciences USA 104, 10667–10672.

Weeda, W., Waldorp, L., Christoffels, I., Huizenga, H., 2009. Activated region fitting: A robust high-power method for fmri analysis using parameterized regions of activation. Human Brain Mapping 30, 2595–2605.

Weinmann, H. J., Laniado, M., Mutzel, W., 1984. Pharmacokinetics of Gd-DTPA/dimeglumine after intraveneous injection into healthy volunteers. Physiological Chemistry and Physics and Medical NMR 16, 167–172.

Whitcher, B., 2010. oro.dicom: Rigorous - DICOM Input / Output. R package version 0.2.7.
URL http://cran.r-project.org/package=oro.dicom

Whitcher, B., Schmid, V., Thornton, A., 2010. oro.nifti: Rigorous - NIfTI Input / Output. R package version 0.2.1.
URL http://cran.r-project.org/package=oro.nifti

Worsley, K., 2005. Spatial smoothing of autocorrelations to control the degrees of freedom in fMRI analysis. NeuroImage 26, 635–641.

Worsley, K., Liao, C., Aston, J. A. D., Petre, V., Duncan, G., Morales, F., Evans, A., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.

Worsley, K., Marrett, S., Neelin, P., Friston, K., Evans, A., 1996. A unified statistical approach for determing significant signals in images of cerebral activation. Human Brain Mapping 4, 58–73.