On the Reliability of Quantitative Volumetric and Structural Neuroimaging

Jan Klein[1], Frederik Laun[2], Peter Erhard[3], Volker Diehl[1,4], Horst K. Hahn[1]

[1]MeVis Research, Bremen, Germany
[2]German Cancer Research Center, Department of Medical Physics in Radiology, Heidelberg, Germany
[3]FB 2 (Chemistry) and Center of Advanced Imaging (CAI), Bremen, Germany
[4]MR- and PET/CT-Center Bremen Mitte, Bremen, Germany

Abstract

Quantitative neuroimaging techniques have become emerging technologies within clinical practice. In this paper, we survey a few clinical applications where quantification methods have received increasing attention, namely in the area of brain atrophy, lesion load computation, and quantification of diffusion processes. We focus on the reliability and reproducibility of such methods and will use the example of quantitative diffusion tensor imaging to discuss methodological details. There, we show possible avenues for evaluating correctness and reliability. On the one hand, we show results from our novel hardware phantom experiments, where axonal fibers are emulated by synthetic industry fibers. On the other hand, we present a new framework for constructing software phantoms which can be used to evaluate, for example, the impact of partial volume effects in case of axonal loss as to be found in multiple sclerosis. Advantages and disadvantages as well as pitfalls of quantification and evaluation techniques are illustrated throughout the paper.

Keywords: quantification, brain atrophy, lesion load, multiple sclerosis, diffusion tensor imaging, phantoms

1 Introduction

Quantitative neuroimaging techniques allow for comparisons between subjects and over time. Cerebrovascular diseases, neoplasia, epilepsy, infection, inflammation, demyelination, psychiatric diseases, neurodegenerative diseases, traumatic injuries and pediatric diseases are only a small subset of the whole range of diseases where quantitative neuroimaging techniques have already proven useful [27, 64]. In most cases, however, these methods severely depend on parameters and physical properties of the image acquisition process, on assumptions and models, as well as on user-dependent inputs. As a consequence, the reliability of a given quantitative method is of paramount importance within clinical settings, where therapy decisions might be drawn from individual measurements.

Before discussing their reliability, we should note that quantification algorithms compare and analyze quite different properties such as morphology or anatomy (MR T1/T2 relax-

ation etc.), local tissue structure (diffusion, magnetization transfer), functional properties (perfusion, functional MRI), as well as atrophy and other volumetric information (typically based on anatomical imaging). Also, the precision required for a specific clinical question and influencing factors of the acquisition and quantification process have to be carefully taken into account. To examine the reliability, several different methods and techniques exist:

- Phantom data: Image data acquired from phantoms are very useful in medical image analysis and quantification as they typically incorporate a precisely known ground truth. While software phantoms allow for modeling arbitrarily shaped complex structures within software, hardware phantoms are limited by the manageability of the used materials and their physical properties. Conversely, realistic imaging artifacts such as noise, spatial distortions, intensity nonuniformity, metal artifacts, etc. can easily be acquired by placing hardware phantoms in real scanners, and hard to simulate on the basis of software phantoms. We will describe related literature as well as our own novel results of software as well as hardware phantoms in Section 3.

- Bootstrapping techniques: The idea of this approach is to estimate the error of some variable by computing the probability distribution from the data itself. Having an original object computed from a list of data (e.g., voxel values, numbers, sequences), bootstrapping methods construct a new list with the same number of elements from the original list by randomly picking elements from the list. Any one element from the list can be picked any number of times. Bootstrapping has been used, e.g., in DTI quantification [6, 33, 36], segmentation [18] as well as in tumor diagnostics [15].

- Patient vs. volunteer studies: Although quantification algorithms may achieve reliable results for test data sets from healthy volunteers, algorithms can fail for real patient data. Thus, it is often crucial to perform the quantification and comparisons for both groups.

- Scan-rescan evaluation: This very well-known and often used technique has been utilized for testing the reliability of quantification tasks, mainly in the area of volumetry and lesion quantification. Furthermore, the reliability of some DTI quantification studies have been examined with respect to scan-rescan reliability where subjects are scanned at least two times within a certain period of time. Initial work has been done where inter-sequence and inter-scanner variability are examined by histogram analysis, but only based on magnetic resonance (MR) images of healthy volunteers [11, 32]. Disease-induced variations, if they are not the target of a measurement procedure, still pose open questions and have to be examined in the future.

- Inter- and intra-observer studies: While inter-observer studies are mainly used to detect user-induced systematic errors in a quantification process, intra-observer studies might identify statistical errors. However, many authors rely on the inter- and intra-observer variability when assessing the practical usefulness of a presented method and oversee that the scan-rescan variability typically is greater by some large factor.

For all evaluation techniques and corresponding quantification parameters it is important to be aware of possible pitfalls in order to set up appropriate testing environments. The

most important pitfall is the presence of partial volume effects (PVE) as virtually all quantitative imaging methods are influenced by these. Thus, PVE have to be taken into account not only during the quantification, but also during possible preprocessing steps of data like fMRI, DTI, or perfusion images. Inter-observer studies are often insufficient as the effective precision of a method typically goes far beyond its inter-observer variability. Moreover, if measuring the progression of a disease by the differences between single time points, it is important to know that the relative error of the difference is significantly higher compared to the relative error of the single measurements.

This paper will survey three major clinical application fields of quantitative neuroimaging where reliability plays a crucial role (Section 2) and will go into more detail at the example of quantitative diffusion tensor imaging (Section 3), followed by a concluding remarks on remaining methodological issues.

2 Selected Applications of Quantitative Neuroimaging

This section outlines three different areas of quantitative neuroimaging, namely atrophy quantification, lesion volumetry, and the quantification of white matter fiber structure or integrity. While brain atrophy quantification and the volumetry of brain lesions, most prominently in multiple sclerosis, are already widely-used, the quantification of white matter fiber structure utilizing diffusion tensor imaging (DTI) is still in an early stage and will have to reveal its full potential. This is not only because of the fact that diffusion tensor imaging is a relatively new imaging technique, but also because of its complexity and the large number of influencing factors that may affect the quantification results. We have chosen these three areas as they are complementary with respect to their complexity, their underlying algorithmic ideas, their areas of application, and, not least, their specific methodological problems.

2.1 Quantification of Brain Atrophy

Various indications exist for global and regional brain volume measurements. Major fields of application are diagnosis, disease monitoring, and evaluation of potential treatments in MS [14, 13, 22, 41, 46, 51, 53] and neurodegenerative diseases, most importantly AD [9, 25]. Rudick et al. [51] propose the BPF, which they define as the ratio of brain parenchymal volume to the total volume within the brain surface contour, as a marker for destructive pathologic processes in relapsing MS patients. De Stefano et al. [14] found substantial cortical GM volume loss in MS. They propose that neocortical GM pathology may occur early in the course of both relapsing-remitting and primary progressive forms of the disease and contribute significantly to neurologic impairment. In addition to a process that is secondary to WM inflammation, they also assume an independent neurodegenerative process, which mainly affects GM and raises the need for robust measures to independently quantify WM and GM volumes. Guttmann et al. [31] described the significance of white matter volume loss even in normal aging. They used the popular EM segmentation algorithm to perform a weighted voxel count over tissue classes for volumetry. These statistical and fully automatic methods work best on multi-spectral data, in this case T2 and PD. Many other whole brain atrophy quantification approaches rely on T1-weighted 3D sequences.

Therein, the importance of brain extraction accuracy has been stressed by Battaglini et al. [7]. Hahn et al. [32] have compared three different T1 based atrophy quantification techniques with respect to scan-rescan and inter-scanner variability as well as resolution, nonuniformity, and noise effects.

2.2 Quantification of Lesion Load in MS

Volumetric analysis of focal brain lesions in multiple sclerosis (MS) is an important issue mainly in therapy monitoring, but also for differential diagnosis [19]. Frequently, the lesion volume change has to be determined, either by independently segmenting serial MR imaging examinations or by subtraction imaging. Duan et al. [17] have compared those two techniques by assessing the relationship to brain atrophy and disease duration and have examined their scan-rescan reproducibility. Ding et al. [16] have developed a method for volumetric quantification of brain tissue studies based on fuzzy clustering of multiparameter MR images.

In order to show the clinical relevance of a proposed quantification method, its reproducibility and accuracy have to be validated. Rexilius et al. [50] have tested the reliability of lesion volumetry analysis using lesion phantoms. Artificial lesions differing in shape, size and orientation are placed in real MR data so that several new artificial data sets can be created. Based on these software phantoms, an evaluation is performed including manual contouring by three human experts and two different semiautomatic approaches (with and without explicit modeling of PVE). All experts overestimated the true lesion volume where the median overestimation was about 50%. Only the quantification approach including explicit PVE modeling leads to good results with low error rates. Their results clearly show the importance of an improved gold standard in lesion volumetry beyond voxel counting. Tofts et al. [58] try to circumvent the partial volume problem in lesion volumetry by a measure of object strength.

2.3 Quantification of White Matter Fiber Structure

The possibility of quantifying DTI parameters has established a whole range of new clinically useful applications and research studies with focus on monitoring disease progression like MS or ALS [1, 10, 29], establishing surrogate markers in assessing the grade of brain tumors [3, 12, 61] or initiating therapies to ensures the best possible development of children [49]. In several studies it has been shown that modified values of fractional anisotropy (FA), relative anisotropy, or diffusion strength are an indicator of diseases affecting white matter tissue [10, 29, 24, 55, 57, 61]. Multiple sclerosis lesions have been investigated by ROI-based analysis [29] and voxel-wise FA comparisons where it has been found that FA changes occur in areas containing lesions and in areas of normal appearing white matter [10]. Moreover, methods for tract-based quantification have been developed [1, 23, 39, 40, 54] where parameters are computed depending on the local curvature or depending on the geodesic distance from a user-defined origin. They allow the clinicians to automatically determine DTI derived parameters along fiber bundles and have already been used for a reproducible quantification of fiber integrity profiles in small structures like the cingulum and the fornix [56], and for mirroring disease progression and executive functioning in MS

Figure 1: DTI data of our hardware phantoms have been acquired on two scanners (left: Siemens Magnetom Verio 3T, right: Siemens Allegra 3T head scanner). In the right image, a cylindrical fiber phantom can be seen.

patients [24]. Probabilistic tracking methods [26, 62] allow for tracking into regions of low anisotropy and are also used to provide a quantitative measure on the probability of a connection between two regions. Session reproducibility and subject variability on fiber tracking algorithms have been examined by Heiervang et al. [34]. They found that data sets acquired by only 12 gradient directions are sufficient for reproducibly defining large fiber bundles like the corpus callosum or the pyramidal tracts, but may not be used for quantifying smaller fibers tracts like the optic radiations. For those fiber tracts, more gradient directions lead to remarkably better results.

3 Learning by Example: Evaluation of Quantitative DTI

Fiber tracking and quantification algorithms are obviously approximations of the reality not only because of the above mentioned pitfalls, but also due to limited spatial resolution (typically a few millimeters) model assumptions (e.g., diffusion assumed to be Gaussian distributed), user-defined parameter settings, and physical imaging artifacts resulting from diffusion sequences. Quantitative DTI, probably representing the most complex of the various quantitative neuroimaging techniques, provides a compelling example for the evaluation of statistical and systematic errors.

Correctness, plausibility, and reliability of both fiber tracking and quantification techniques have mainly been verified using histologic knowledge [35, 63, 65]. In some few animal studies, manganese has already been used as tracer to directly examine the diffusion process [44]. First quantitative results with respect to precision, uncertainty and reproducibility have also been published [6, 8, 36]. Behrens et al. [8] estimate the local probability density using a model describing the diffusion process. The model is used to determine the probability of a connection between two points and, therefore, is used as a quantitative measure for the correctness of the fiber tracking results. Jones [36] makes use of the bootstrapping method in order to compute cones of uncertainties showing a 95% confidence angle. Basser et al. [6] propose a Gaussian distribution that describes the

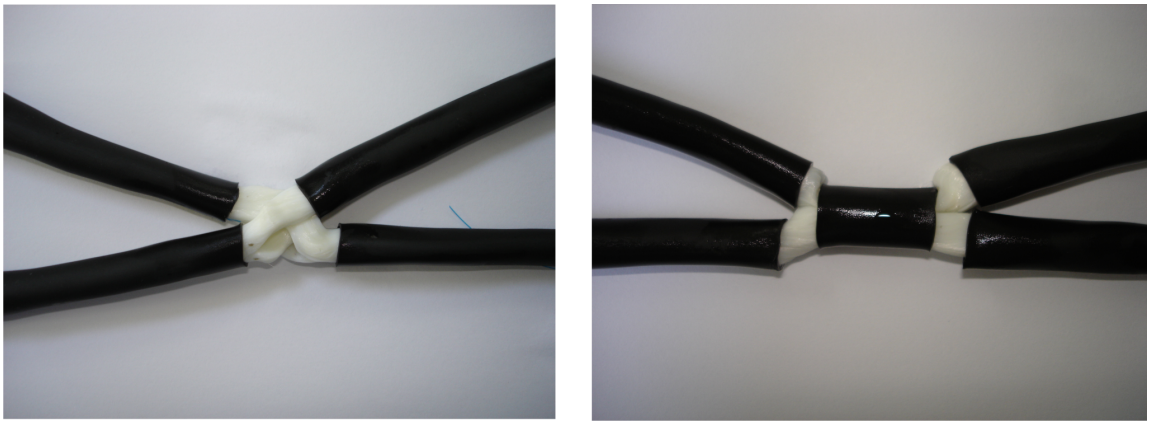


Figure 2: Two hardware phantoms constructed from $8\mu m$ polyfil industry fibers. The fibers are enclosed by a black heat shrink tube to avoid that larger spaces between single fibers occur.

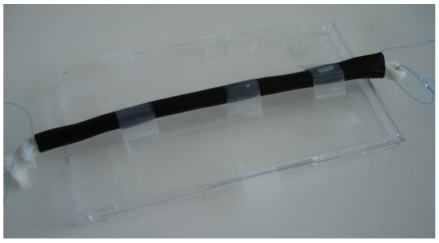
variability of the tensors in the ideal case where the image is only disturbed by radio frequency background noise. In combination with bootstrapping, where the real variability is measured, they are able to benchmark the quality of DTI data. Thereby, wavelet-based methods help them to reduce noise and to preserve borders between different tissue classes.

Phantoms, modeling physically plausible fiber bundles that conform (partially) with human anatomy are important in order to examine different quantification algorithms with respect to the points mentioned above. A phantom must allow to steer the respective DTI data generation under controlled conditions, either using a real MR scanner (*physical phantom*) or by the help of software in a simulation setup (*software phantom*).

3.1 Hardware Phantoms

Hardware phantoms to assess DTI can be created from physical materials such as silk threads or dialysis tubes [20, 21] and placed in a water basin for acquiring the diffusion weighted images. In our own experiments, we have used $8\mu m$ polyfil fibers for constructing kissing, crossing, and straight fiber bundles (see Figures 1 and 2). We have successfully tested them with two different scanners (Siemens Magnetom Verio 3T as well as Siemens Allegra 3T head scanner) and have examined the variability of DTI quantification results. Both scanners show the same behavior that the results strongly depend on the selected resolution, the number of gradients, the measurement time, the number of repetitions, the necessary postprocessing of DTI data including optional filtering and smoothing, as well as on the type of fiber tracking algorithm used.

Figure 3 shows some quantification results of the axial and radial diffusivity that can be derived from the tensor's eigenvalues (λ_i). For filtering, a Gaussian smoothing (using an infinite impulse response algorithm with $\sigma = 1$), followed by a resampling to an isotropic image resolution (we selected $1.5mm^3$) using a cubic B-spline filter was performed. On the one hand, filtering leads to better fiber tracking results as single outliers do not lead to a stopping of the tracking process. On the other hand, however, it changes the quantification results. Not only the absolute parameter values change (which can be explained by PVE and which is described in detail in the next subsection), but also the proportions change.



	# gradients	NEX (number of excitations)	resolution	TR / TE
1	30	2	1.8 x 1.8 x 2.2	12.500/108
2	30	2	2.3 x 2.3 x 2.5	12.500/108
3	6	9	2.3 x 2.3 x 2.5	12.500/108
4	6	9	2.3 x 2.3 x 5.5	6.500/108

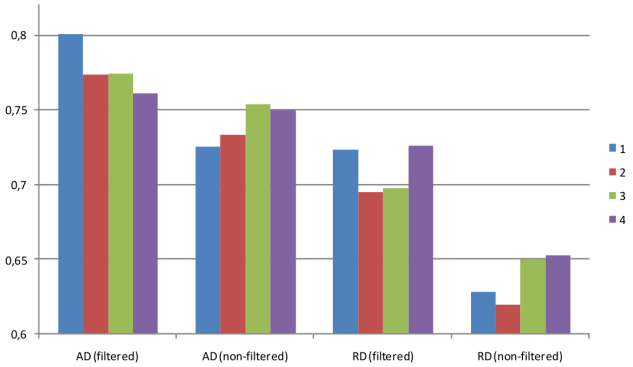


Figure 3: Quantification results using a straight hardware phantom. Axial diffusivity AD $= \lambda_1$ and radial diffusivity RD $= \frac{1}{2}(\lambda_2 + \lambda_3)$ are shown for different acquisition protocols. The data set has been acquired on a 3T Siemens Magnetom Verio scanner.

Instead of increasing axial diffusivity with decreasing resolution or fewer directions respectively, the filtering flips this behavior. This is also due to PVE, but the flipping results from the fact that the data is resampled to a finer resolution compared to their original ones. Furthermore, we also confirm the results of Goodlett et al. [28] where it has been shown analytically that low direction schemes introduce a statistical bias with a clinically relevant magnitude.

Figure 4 (left) shows fiber tracking results for different tractography algorithms (advection, deflection) of a hardware phantom, which models crossing fibers (see Figure 2, left). Using our quantification tool [39], we can see that deflection-based fiber tracking leads to remarkably higher FA values (about 50%) as well as to an increased number of completely tracked fibers (also about 50%). Figure 4 (right) gives an insight into the underlying tensor data and shows a volume rendering of the fractional anisotropy. It clearly depicts the problem in the area of crossing fibers. For one of the two bundles, the fractional anisotropy is very low and, thus, fiber tracking may abort, especially, if only the main eigenvectors are used for tracking.

Hardware phantom experiments for high angular resolution diffusion-weighted imaging (HARDI) data have been proposed recently [48, 59]. In [59] three different techniques are compared, namely constrained spherical deconvolution (CSD), super-resolved CSD and Q-ball imaging. It is shown that fiber tracking results, and as a consequence DTI quantification, depend on the employed algorithm's ability to resolve crossing fibers, and to provide accurate estimates of their orientations.

Overall, the advantage of hardware phantoms compared to software phantoms is that MR images can be acquired under realistic circumstances where images are disturbed by noise, distortions and other imaging artifacts. The disadvantage is the high manual effort which is needed for constructing and modifying such phantoms as well as the difficulty of constructing complex or anatomically realistic structures.

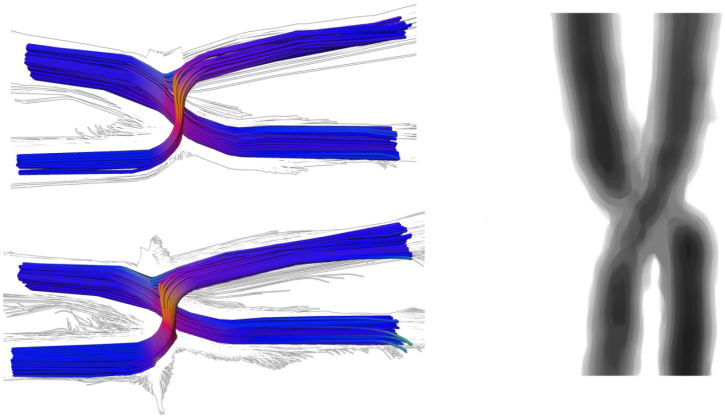


Figure 4: Left: fiber tracking results of our hardware phantom, where crossing fibers are modeled. The upper image shows fibers obtained by advection-based fiber tracking. The lower image displays the results of deflection-based fiber tracking, where about 50% more fibers have been tracked passing through the critical region. Right: a volume rendering of the fractional anisotropy (FA) shows that in the crossing area the FA is very low for one of the two bundles and, thus, fiber tracking may abort. The data set has been acquired on a 3T Siemens Allegra head scanner.

3.2 Software Phantoms

In contrast to hardware phantoms, software phantoms allow for an easy an exact geometrical description of arbitrarily shaped fibers and of an automatic computation of the corresponding diffusion weighted-images so that no MR scanner is needed. Basser et al. [4, 5] describe fibers by simple 2D rings in tensor fields, whereas other authors [30, 45, 60] define fibers by cylindric tubes in 3D tensor fields. Thereby, tracts are defined by circular helixes. A mathematical framework for simulating the partial volume between fiber and background tissue has been proposed in [42, 43]. The authors obtain a model of a fiber bundle by parameterizing the various features which characterize the bundle. Their results show that a higher correspondence between experimental and synthetic DTI data exists when the modeling a nonconstant fiber density across bundles.

As mentioned in Section 2.3, DTI quantification has become very popular for monitoring the disease progression of MS. However, axonal loss is coherent with progressive brain atrophy, and thus, PVE become more and more important. That means, that for a specific fiber bundle the fraction of partial volume voxels containing both fiber tissue and surrounding tissue or cerebrospinal fluid (background) increases if demyelination and axonal loss occur. To examine the impact of such PVE, we have developed a DTI software phantom, which is simple by construction and which can easily be adapted to own requirements with respect to assumed diffusion properties, image resolution, pixel noise, and the amount of axonal loss or atrophy. The phantom, a tensor field which is used for fiber tracking and quantification, consists of background, tissue as well as partial volume voxels and describes a straight fiber bundle. For constructing the phantom, a quasi-continuous tensor field is modeled where initially only background and tissue exist. The background is described by spherical tensors (FA=0) whereas the tissue is modeled by ellipsoids with FA=0.578, see

Figure 5. The diffusion strength of background voxels has been chosen four times higher than in tissue which is comparable to the proportion as it can be found in the human brain. We emulate axonal loss arising from progressive brain atrophy by decreasing the thickness of the tissue (see Figure 5), i.e., we replace tissue voxels by background voxels. Afterwards, for such a tensor field, diffusion weighted images are derived using a fixed gradient scheme with 30 directions [37]. To simulate an acquisition process where PVE as well as noise effects are present, the images are sampled to a fixed resolution of $2.0mm^3$ (using a triangle/linear filter) and complex Gaussian noise is added [38]. Finally, the tensor field is determined and fiber tracking as well as quantification can be performed [39]. It is important to choose sufficiently large seed ROIs for fiber tracking to cover the whole tissue as well as partial volume.

Obviously, the smaller the thickness, the higher is the fraction of partial volume voxels on the total amount of tissue voxels. These partial volume voxels consist of a mixture of background and pure tissue. As all eigenvalues of the background voxels are higher than the eigenvalues of tissue voxels (in the background the ADC is four times higher), partial volume voxels must also have higher eigenvalues than tissue voxels. As a consequence, fibers resulting from partial volume voxels will have higher AD (axial diffusivity = λ_1) , RD (radial diffusivity = $\frac{1}{2}(\lambda_2 + \lambda_3)$) and ADC values than pure fiber voxels. Furthermore, the FA values must decrease. Overall spoken, the smaller the thickness, the higher is the difference to the real DTI parameters. Specific results can be found in Figure 6. The RD values differ by up to 70%, the ADC values by 45% and the AD values by up to 23%. The fractional anisotropy differs (Figure 6, right) by about 20%.

These phenomenons and differences should be considered if performing DTI-based quantification. ROI-based quantification tools have already been developed where PVE are considered. Schlueter et al. [52] use an EM clustering technique to classify between tissue, background and partial volume and are able to quantify within those specific regions afterwards. This technique has also been extended to tract-based quantification [40]. However, a user-independent quantification process cannot be guaranteed under all circumstances, as for the inherent clustering process only geometric affinity measures are used without knowledge about the underlying true anatomy.

4 Conclusion

Technical challenges [47] like improved spatial resolution, whole brain coverage, signal to noise ratio, or magnetic susceptibility artifacts constitute the basis for reliable quantification techniques in neuroimaging. For example, high-resolution 3D imaging sequences facilitated by parallel imaging will strongly contribute towards quantitative reliability. Still, in most cases, partial volume modeling will be key to yield highly reliable quantitative measurements due to the complexity or small spatial extent of both anatomical features and pathological alterations. For example, there is increasing evidence that subtle or even significant gray matter alterations play an important role in MS pathology [66]. Furthermore, preprocessing algorithms for registration, regularization, or outlier rejection are substantial influencing factors.

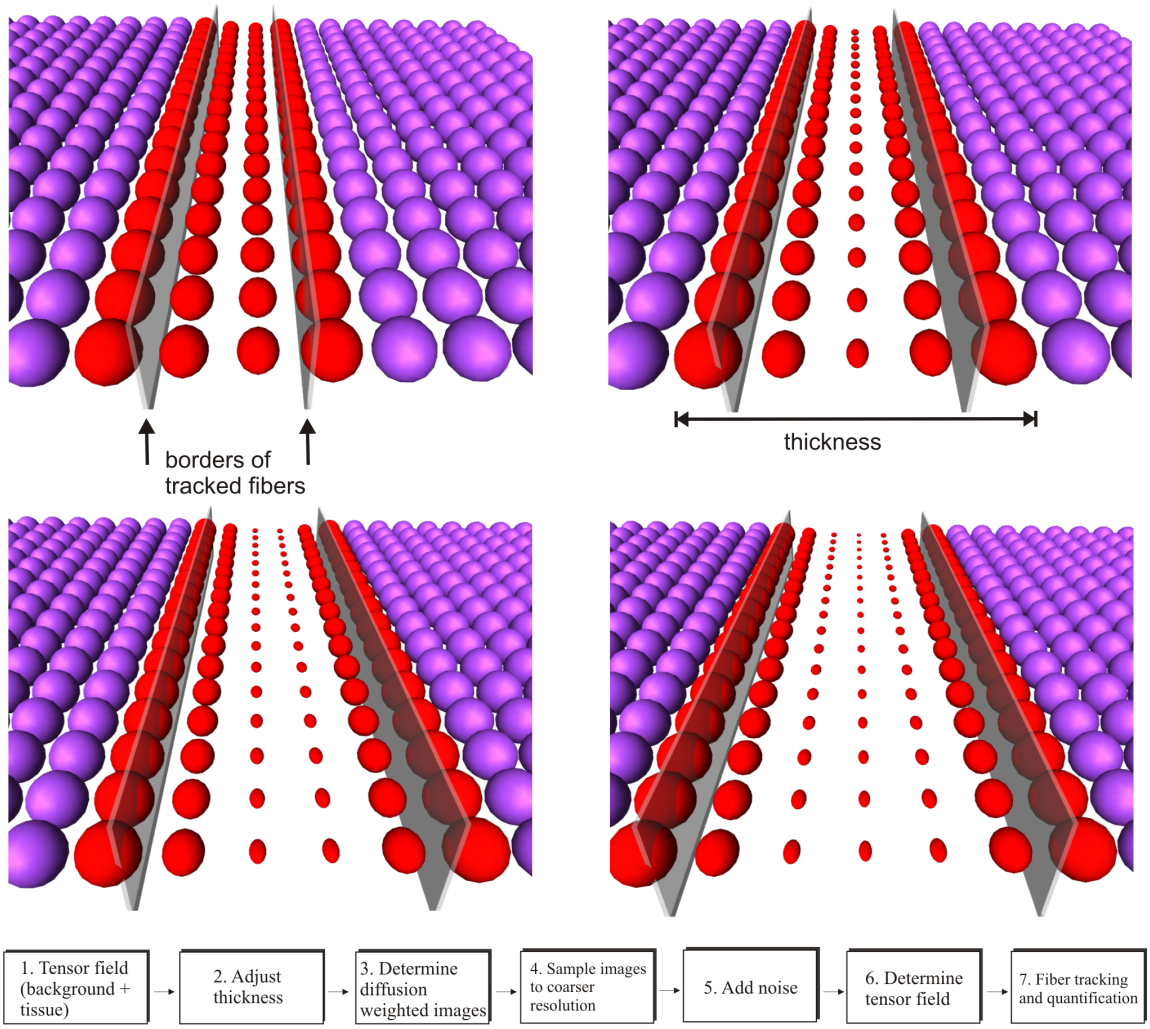


Figure 5: A DTI software phantom which allows to explore the impact of axonal loss which is coherent with progressive brain atrophy. The basic idea is to define a tensor field where the thickness of the anisotropic region can be modified and where both the extent of PVE and the amount of noise can be defined by the user. The images show the borders of tracked fibers for differently chosen thicknesses.

In the case of quantitative DTI, the assumption of a Gaussian diffusion process may not be adequate in areas of complex fiber structures like crossing or kissing fibers not only for fiber reconstruction but also for quantitative assessment. This problem has recently been addressed by multiple-compartment models, diffusion spectrum imaging, spherical deconvolution and persistent angular structure MRI (PAS-MRI), where higher order tensors or probability distributions describe the actual diffusion process. Assaf et al. [2] have already shown that with q-space imaging the difference of values in the normal appearing white matter of patients with multiple sclerosis is more pronounced than with DTI. However, virtually all techniques based on high-angular resolution diffusion imaging (HARDI) data

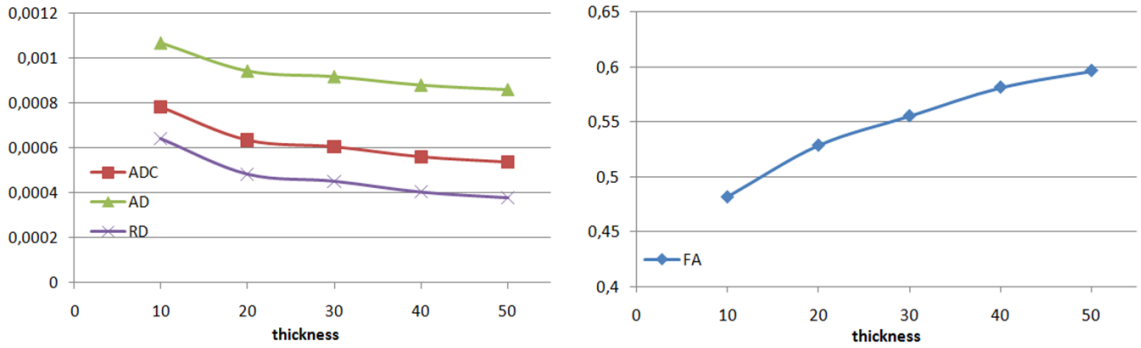


Figure 6: The plots show the dependency of DTI parameters on the chosen thickness of the anisotropic tissue of our software phantom.

are still in an early state and are subject to improvement with respect to acquisition and postprocessing time so that they become useful for clinical routine.

We have shown several examples in the context of neuroimaging where quantification techniques play an important role and have presented and discussed software and hardware phantoms for measuring their precision and reliability. Without such evaluation basis, several pitfalls and systematic errors might remain undetected. Although substantial methodological effort has been made by several authors in the area of quantitative neuroimaging, further examinations are still urgently required. This is due to the immense complexity and the amount of influencing parameters throughout the whole pipeline of acquisition, processing, and quantification.

References

[1] S. Aoki, N.K. Iwata, Y. Masutani, M. Yoshida, O. Abe, Y. Ugawa, T. Masumoto, H. Mori, N. Hayashi, H. Kabasawa, S. Kwak, S. Takahashi, S. Tsuji, and K. Ohtomo. Quantitative Evaluation of the Pyramidal Tract Segmented by Diffusion Tensor Tractography: Feasibility Study in Patients with Amyotrophic Lateral Sclerosis. *Radiation Medicine*, 23(3):195–199, 2005.

[2] Y. Assaf, D. Ben-Bashat, J. Chapman, S. Peled, I.E. Biton, M. Kafri, Y. Segev, T. Hendler, A.D. Korczyn, M. Graif, and Y. Cohen. High b-value q-space analyzed diffusion-weighted MRI: Application to multiple sclerosis. *Magnetic Resonance in Medicine*, 47(1):115–126, 2002.

[3] D.P. Barboriak. Imaging of brain tumors with diffusion-weighted and diffusion tensor MR imaging. *Magn Reson Imaging Clin N Am*, 11:379–401, 2003.

[4] P. Basser, S. Pajevic, C. Pierpaoli, and A. Aldroubi. Fiber tract following in the human brain using DT-MRI data. *IEICE Trans. Inf. & Syst.*, E85-D(1):15–21, 2002.

[5] P. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magn Reson Med*, 44(4):625–632, 2000.

[6] P.J. Basser and S. Pajevic. Dealing with Uncertainty in Diffusion Tensor MR Data. *Israel Journal of Chemistry*, 43:129–144, 2003.

[7] M. Battaglini, S.M. Smith, S. Brogi, and N. De Stefano. Enhanced brain extraction improves the accuracy of brain atrophy estimation. *Neuroimage*, 40(2):583–589, 2008.

[8] T.E.J. Behrens, M.W. Woolrich, M. Jenkinson, H. Johansen-Berg, R.G. Nunes, S. Clare, P.M. Matthews, J.M. Brady, and S.M. Smith. Characterization and Propagation of Uncertainty in Diffusion-Weighted MR Imaging. *Magnetic Resonance in Medicine*, 50:1077–1088, 2003.

[9] A. Brunetti, A. Postiglione, E. Tedeschi, A. Ciarmiello, M. Quarantelli, E.M. Covelli, G. Milan, M. Larobina, A. Soricelli, A. Sodano, and B. Alfano. Measurement of global brain atrophy in Alzheimer's disease with unsupervised segmentation of spin-echo MRI studies. *J Magn Reson Imaging*, 11(3):260–266, 2000.

[10] S. Cader, H. Johansen-Berg, M. Wylezinska, J. Palace, T.E. Behrens, S. Smith, and P.M. Matthews. Discordant white matter N-acetylasparate and diffusion MRI measures suggest that chronic metabolic dysfunction contributes to axonal pathology in multiple sclerosis. *Neuroimage*, 36(1):19–27, 2007.

[11] M. Cercignani, R. Bammer, M.P. Sormani, F. Fazekas, and M. Filippi. Inter-sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. *American Journal of Neuroradiology*, 24: 638–643, 2003.

[12] T.L. Chenevert, L.D. Stegman, J.M.G. Taylor, P.L. Robertson, H.S. Greenberg, A. Rehemtulla, and B.D. Ross. Diffusion Magnetic Resonance Imaging: an Early Surrogate Marker of Therapeutic Efficacy in Brain Tumors. *J Natl Cancer Inst*, 92:2029–2036, 2000.

[13] N. De Stefano, M. Battaglini, and S.M. Smith. Measuring brain atrophy in multiple sclerosis. *J Neuroimaging*, 17(Suppl 1):10S–15S, 2007.

[14] N. De Stefano, P.M. Matthews, M. Filippi, F. Agosta, M. De Luca, M.L. Bartolozzi, L. Guidi, A. Ghezzi, E. Montanari, A. Cifelli, A. Federico, and S.M. Smith. Evidence of early cortical atrophy in MS: Relevance to white matter changes and disability. *Neurology*, 60(7):1157–1162, 2003.

[15] M. Diaz and J.S. Rao. Non-parametric bootstrap ensembles for detection of tumor lesions. *Pattern Recognition Letters*, 28(16):2273–2283, 2007.

[16] Z. Dinga, J. Preiningerovab, C.J. Cannistracia, T.L. Vollmerb, J.C. Gorea, and A.W. Andersona. Quantification of multiple sclerosis lesion load and brain tissue volumetry using multiparameter MRI: methodology and reproducibility. *Magnetic Resonance Imaging*, 23(3):445–452, 2005.

[17] Y. Duan, P. Hildenbrand, M. Sampat, D. Tate, I. Csapo, B. Moraal, R. Bakshi, F. Barkhof, D. Meier, and C. Guttmann. Segmentation of subtraction images for the measurement of lesion change in multiple sclerosis. *AJNR Am J Neuroradiol*, 29(2): 340–346, 2008.

[18] D. Dutendas, L. Moreau, F. Ghorbel, and P.M. Allioux. Unsupervised Bayesian segmentation with bootstrap sampling application to eye fundus image coding. *Nuclear Science Symposium and Medical Imaging Conference, 1994., 1994 IEEE Conference Record*, 4:1794–1796, 1994.

[19] F. Fazekas, P. Soelberg-Sorensen, G. Comi, and M. Filippi. MRI to monitor treatment efficacy in multiple sclerosis. *J Neuroimaging*, 17(Suppl 1):50S–55S, 2007.

[20] E. Fieremans, Y. De Deene, S. Delputte, M. Ozdemir, E. Achten, and I. Lemahieu. The design of anisotropic diffusion phantoms for the validation of diffusion weighted magnetic resonance imaging. *Physics in Medicine and Biology*, 53:5405–5419, 2008.

[21] E. Fieremans, Y. De Deene, S. Delputte, M. Ozdemir, Y. D'Asseler, J. Vlassenbroeck, K. Deblaere, E. Achten, and I. Lemahieu. Simulation and experimental verification of the diffusion in an anisotropic fiber phantom. *Journal of Magnetic Resonance*, 190 (1):189–199, 2008.

[22] M. Filippi, F. Agosta, and M.A. Rocca. Regional assessment of brain atrophy: A novel approach to achieve a more complete picture of tissue damage associated with central nervous system disorders? *Am J Neuroradiol*, 28(2):260–261, 2007.

[23] P. Fillard, J. Gilmore, J. Piven, W. Lin, and G. Gerig. Quantitative Analysis of White Matter Fiber Properties along Geodesic Paths. In *MICCAI'03*, pages 16–23, 2003.

[24] F. Fink, J. Klein, T. Mitrovics, and H. Hildebrandt. Mirroring disease progression and executive functioning in MS patients by DTI-based fibertracking and various brain atrophy measures. *Brain and Cognition*, 2009. to appear.

[25] N.C. Fox, P.A. Freeborough, and M.N. Rossor. Visualisation and quantification of rates of atrophy in Alzheimer's disease. *Lancet*, 348(9020):94–97, 1996.

[26] O. Friman, G. Farnebäck, and C.-F. Westin. A Bayesian Approach for Stochastic White Matter Tractography. *IEEE Transactions on Medical Imaging*, 25(8):965–978, 2006.

[27] J.H. Gillard. *Clinical MR Neuroimaging. Diffusion, Perfusion and Spectroscopy.* Cambridge University Press, 2004.

[28] C. Goodlett, P.T. Fletcher, W. Lin, and G. Gerig. Quantification of Measurement Error in DTI: Theoretical Predictions and Validation. In *MICCAI (1)*, vol. 4791 of *Lecture Notes in Computer Science*, pages 10–17, 2007.

[29] C.M. Griffin, D.T. Chard, O. Ciccarelli, R. Kapoor, G.J. Barker, A.J. Thompson, and D.H. Miller. Diffusion tensor imaging in early relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, 7(5):290–297, 2001.

[30] C. Gössl, L. Fahrmeir, B. Pütz, L. Auer, and D. Auer. Fiber tracking from DTI using linear state space models: detectability of the pyramidal tract. *Neuroimage*, 16(2): 378–388, 2002.

[31] R.G. Guttmann, R. Benson, S.K. Warfield, X. Wei, M.C. Anderson, C.B. Hall, K. Abu-Hasaballah, J.P. Mugler, and L. Wolfson. White matter abnormalities in mobility-impaired older persons. *Neurology*, 54:1277–1283, 2000.

[32] H.K. Hahn, B. Jolly, M. Lee, D. Krastel, J. Rexilius, J. Drexl, M. Schlüter, B. Terwey, and H.-O. Peitgen. How Accurate is Brain Volumetry? A Methodological Evaluation. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 3216, pages 335–342, 2004.

[33] H.K. Hahn, J. Klein, C. Nimsky, J. Rexilius, and H.-O. Peitgen. Uncertainty in Diffusion Tensor Based Fibre Tracking. *Acta Neurochir Suppl*, 98:33–41, 2006.

[34] E. Heiervang, T.E. Behrens, C.E. Mackay, M.D. Robson, and H. Johansen-Berg. Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage*, 33(3):867–77, 2006. 1053-8119 (Print)Journal Article.

[35] B. Inglis, D. Neubauer, L. Yang, D. Plant, T. Mareci, and D. Muir. Diffusion tensor MR imaging and comparative histology of glioma engrafted in the rat spinal cord. *AJNR Am J Neuroradiol*, 20:713–716, 1999.

[36] D.K. Jones. Determining and Visualizing Uncertainty in Estimates of Fiber Orientation From Diffusion Tensor MRI. *Magnetic Resonance in Medicine*, 49:7–12, 2003.

[37] D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42(3):515–525, 1999.

[38] J. Klein, H.K. Hahn, J. Rexilius, P. Erhard, M. Althaus, D. Leibfritz, and H.-O. Peitgen. Efficient Visualization of Fiber Tracking Uncertainty based on Complex Gaussian Noise. In *Proc. 14th ISMRM Scientific Meeting & Exhibition (ISMRM 2006)*, page 2753, 2006.

[39] J. Klein, S. Hermann, O. Konrad, H.K. Hahn, and H.-O. Peitgen. Automatic Quantification of DTI Parameters along Fiber Bundles. In *Proceedings of Image Processing for Medicine*, pages 272–276, 2007.

[40] J. Klein, H. Stuke, J. Rexilius, B. Stieltjes, H.K. Hahn, and H.-O. Peitgen. Towards User-Independent DTI Quantification. *Proc. SPIE Medical Imaging*, 6914:69142E–1–69142E–8, 2008.

[41] M. Lanz, H.K. Hahn, and H. Hildebrandt. Brain atrophy and cognitive impairment in multiple sclerosis: A review. *J Neurol*, 254(Suppl 2):II43–48, 2007. Erratum in: J Neurol 255(2):309–10, Feb 2008.

[42] A. Leemans, J. Sijbers, M. Verhoye, A. van der Linden, and D. van Dyck. A simulated phantom for diffusion tensor MRI fiber tracking. In *Proceedings of ACIVS 2003 (Advanced Concepts for Intelligent Vision Systems)*, pages 281–285, 2003.

[43] A. Leemans, J. Sijbers, M. Verhoye, A. van der Linden, and D. van Dyck. Mathematical framework for simulating diffusion tensor MR neural fiber bundles. *Magnetic resonance in medicine*, 53(4):944–953, 2005.

[44] C.-P. Lin, W.-Y. I. Tseng, H.-C. Cheng, and J.-H. Chen. Validation of Diffusion Tensor Magnetic Resonance Axonal Fiber Imaging with Registered Manganese-Enhanced Optic Tracts. *Neuroimage*, 14:1035–1047, 2001.

[45] N. Lori, E. Akbudak, J. Shimony, T. Cull, A. Snyder, R. Guillory, and T. Conturo. Diffusion tensor fiber tracking of human brain connectivity: acquisition methods, reliability analysis and biological results. *NMR Biomed*, 15:493–515, 2002.

[46] C. Lukas, H.K. Hahn, B. Bellenberg, J. Rexilius, G. Schmid, S.K. Schimrigk, H. Przuntek, O. Köster, and H.-O. Peitgen. Sensitivity and reproducibility of a new fast 3D segmentation technique for clinical MR-based brain volumetry in multiple sclerosis. *Neuroradiology*, 46(11):906–915, 2004.

[47] D.C. Noll. Technical Challenges in Functional Neuroimaging. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1208–1214, 2004.

[48] C. Poupon, B. Rieul, I. Kezele, M. Perrin, F. Poupon, and J.-F. Mangin. New diffusion phantoms dedicated to the study and validation of HARDI models. *Magnetic Resonance in Medicine*, 2008. in press.

[49] C. Pul, J. Buijs, A. Vilanova, F.G. Roos, and P.F.F. Wijn. Fiber Tracking in Newborns with Perinatal Hypoxic-Ischemia at Birth and at 3 Months. *Radiology*, 240(1):203–214, 2006.

[50] J. Rexilius, H.K. Hahn, M. Schlüter, H. Bourquain, and H.-O. Peitgen. Evaluation of Accuracy in MS Lesion Volumetry Using Realistic Lesion Phantoms. *Acad. Radiol.*, 12(1):17–24, 2005.

[51] R.A. Rudick, E. Fisher, J.C. Lee, J. Simon, and L. Jacobs. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. Multiple Sclerosis Collaborative Research Group. *Neurology*, 53(8):1698–1704, 1999.

[52] M. Schlüter, B. Stieltjes, H.K. Hahn, J. Rexilius, Konrad-Verse O., and H.-O. Peitgen. Detection of Tumour Infiltration in Axonal Fibre Bundles Using Diffusion Tensor Imaging. *Int J Medical Robotics and Computer Assisted Surgery*, 1:80–86, 2005.

[53] J.H. Simon. Brain atrophy in multiple sclerosis: What we know and would like to know. *Mult Sclerosis*, 12(6):679–687, 2006.

[54] S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T.E. Nichols, C.E. Mackay, K.E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews, and T.E.J. Behrens. Tract-Based Spatial Statistics: Voxelwise Analysis of Multi-Subject Diffusion Data. *Neuroimage*, 2006.

[55] R. Stahl, O. Dietrich, S. Teipel, H. Hampel, M.F. Reiser, and S.O. Schoenberg. Assessment of axonal degeneration on Alzheimer's disease with diffusion tensor MRI. *Radiologe*, 43(7):566–575, 2003.

[56] B. Stieltjes, J. Klein, B. Hyman, L.G. Naul, V. Runge, and M. Essig. Reproducible quantification of fiber integrity profiles in the cingulum and the fornix using an experimental 32 channel head coil. In *ISMRM 2008*, page 1842, 2008.

[57] A. Tievsky, T. Ptak, and J. Farkas. Investigation of apparent diffusion coefficient and diffusion tensor anisotropy in acute and chronic multiple sclerosis lesions. *Am. J. Neuroradiol.*, 20:1491–1499, 1999.

[58] P.S. Tofts, N.C. Silver, G.J. Barker, and A. Gass. Object strength–An accurate measure for small objects that is insensitive to partial volume effects. *MAGMA*, 18(3): 162–169, 2005.

[59] J.-D. Tournier, C.-H. Yeh, F. Calamante, K.-H. Cho, A. Connelly, and C.-P. Lin. Resolving crossing fibres using constrained spherical deconvolution: Validation using diffusion-weighted imaging phantom data. *Neuroimage*, 42(2):617–625, 2008.

[60] J.D. Tournier, F. Calamente, M. King, D. Gadian, and A. Connelly. Limitations and requirements of diffusion tensor fiber tracking: an assessment using simulations. *Magn Reson Med*, 47(4):701–708, 2002.

[61] A. Tropine, G. Vucurevic, P. Delani, S. Boor, N. Hopf, J. Bohl, and P. Stoeter. Contribution of diffusion tensor imaging to delineation of gliomas and glioblastomas. *J Magn Reson Imaging*, 20(6):905–912, 2004.

[62] D.S. Tuch. Q-Ball Imaging. *Magnetic Resonance in Medicine*, 52:1358–1372, 2004.

[63] M. Wiegell, D. Tuch, H. Larsson, and V. Wedeen. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *Neuroimage*, 19:391–401, 2003.

[64] H. Zaidi. *Quantitative Analysis in Nuclear Medicine Imaging*. Springer US, 2006.

[65] J. Zhang, P. van Zijl, and S. Mori. Three-dimensional diffusion tensor magnetic resonance microimaging of adult mouse brain and hippocampus. *Neuroimage*, 15:892–901, 2002.

[66] R. Zivadinov and J.L. Cox. Neuroimaging in multiple sclerosis. *Int Rev Neurobiol*, 79: 449–474, 2007.