The Effects of Inter-slice Gap Thickness on Parametric Methods for Diffusion Tensor Imaging

Beth Hutchinson; Neuroscience Training program, Statistics 592

Abstract

Although high-resolution, whole-brain coverage is desirable for the acquisition of diffusion tensor (DT) images, it is not always acquired clinically due to time and scanner constraints. Although region of interest (ROI) analysis is less affected by this slice undersampling, there are likely more substantial implications for the use of parametric methods. This work will explore the effects of inter-slice gap size, or the absent data between acquired slices, on the sensitivity of statistical parametric mapping for detecting group differences in signal and spatial extent using DTI data. Simulated data sets were generated from a high-resolution, whole brain DTI acquisition with either no mask, a spherical or diffuse white matter mask applied to the unprocessed diffusion data. K-space manipulations were used to create data sets with larger voxels and these were additionally undersampled to simulate gap thickness. These simulated data sets were processed and the fractional anisotropy maps analyzed using statistical parametric mapping to calculate a voxel-wise T-test between masked and non-masked groups. The T-statistic map was then used as a classifier and the original mask as an indicator for true positives in receiver operating characteristic (ROC) analysis. From this analysis, it was shown that undersampled data sets diminish the effectiveness of parameteric mapping differentially based on the type of mask.

Introduction

Voxelwise methods of image analysis offer multiple advantages for comparing groups of magnetic resonance imaging (MRI) data based on the nature of these methods, which is to compare groups of data statistically at each voxel in the image. The most used structural MRI application of voxelwise analysis is voxel based morphometry, which is typically employed to compare white and grey matter density between groups. Voxelwise methods differ from region of interest approaches, which are more commonly used for T2 and diffusion image analysis. Voxel based approached do not require an *a priori* hypothesis about brain structures, which might be different between groups and consequently, experiments may yield information that would otherwise be overlooked using other methods. The automation of the method and necessity for normalization of data into a standard space reduce the user bias inherent in techniques such as region of interest (ROI) analysis. Additionally, the standard strereotaxic coordinates describing activation or group differences may be used to compare proposed biological mechanisms across studies. Although parametric methods have numerous strengths, there are also caveats for this type of analysis, which must not be ignored. Multiple comparisons errors are one of the largest issues associated with parametric techniques and a number of post processing corrections have been developed to describe and account for the error introduced by making thousands of statistical tests of one data set. Additionally, the statistical power is reduced and it is more likely that some regions will not be detected.

Recently, there has been much interest in using diffusion tensor imaging to characterize normal human brains as well as patient groups since the technique is very sensitive to changes in diffusion and anisotropy (LeBihan et al., 2002). The most commonly used method for analyzing this data is ROI based, however there is a growing use of parametric approaches based on the previously mentioned advantages. A recent study (Snook et al., 2007) compared the results of ROI and voxelwise techniques for reporting changes in diffusivity measures during childhood brain development. The authors found that there was reasonable agreement between the two approaches for detecting differences in apparent diffusivity (AD) and fractional anisotropy (FA), however neither technique was able to detect all changes. The ROI approach was more powerful statistically, however the voxelwise approach showed changes in diffuse white matter that the ROI based approach was not able to detect. Based on their results, the authors recommend application of both techniques for a complete account of diffusivity measures.

In order to detect differences using parametric techniques, it is necessary to compare the

information from multiple brains. Clearly the ideal data for use in such applications incorporates the entire brain, however in practice the acquisition does not always include the entire brain due to time constraints or scanner capability. It is not uncommon to collect slices that are thick or that have gaps in between. In such cases, it is instructive to know how much this undersampling affects the detection of group differences when using parametric. Although the effects of gap thickness are less commonly investigated, there are multiple studies addressing the effects of resolution on image analysis and in particular attempting to quantify and correct for the partial volume effect, which arises due to the presence of more than one tissue type within the same voxel. However, it is not always ideal to use the highest resolution possible for MRI because with higher resolution there is decreased signal to noise ratio, based on the physics of MRI acquisition.

The implications of resolution on grey or white matter density analyses, such as voxel based morphometry have been discussed in the literature and the recommendation for these types of analyses is to acquire the T1-weighted images used for this with high resolution (1-1.5 mm) and isotropic voxel dimensions (Ashburner and Friston, 2000). At a more basic level, the effects of imaging parameters such as slice thickness and slice orientation have been shown the affect the exactness of volumetry measures (Luft et al. 1996). A series of phantoms were scanned with varied imaging parameters and processed to calculate the geometric volume of the phantoms. It was found that the error in volume measurement increased with slice thickness in a slice orientation dependent fashion.

Studies of the effects of resolution on DTI are more complicated due to the tensor calculation and typically lower in-plane resolution of DTI compared with other structural MRI techniques. In a study by Alexander et al. (2001), the advantages of using a two compartments tensor model as opposed to the standard one tensor model were considered to investigate the presence and extent of partial volume effects in DT tensor model may lead to false reporting of diffusion values in some cases and the two compartment model may be advantageous for certain model parameters. Furthermore, the authors found that experimental DTI scans at high b values and large voxel size were susceptible to partial voluming even though at typically used values, the effects were more difficult to detect. The implications of resolution for DTI extend beyond the calculation of diffusivity measures to influence tractography, which is becoming popular for mapping white matter tracts in the brain. Kim et al. (2006), have recently shown that voxel size can change the detection of fiber tracts and consequently the performance of tractography algrorithms. They suggest a voxel resolution on the order of 2 mm.

Simulation of data can offer methodological insight for image acquisition and processing development and consequently has been used for a variety of MRI approaches. In particular the generation of image data sets is useful for validating techniques and there are several projects intended

to supply the ability to create such data sets. The Brainweb simulator (Cocosco et al., 1997) is a freely available resource intended to provide "ground truth" images for validation studies. The simulator is able to generate T1, T1 and proton density weighted images based on the manipulation of a digital phantom. Using the fundamental Bolch equations for MRI signal, the simulator applies mathematical manipulations to the digital phantom to allow a variety of simulated data.

The primary motivation for the work presented here is to develop a simulation and analysis paradigm by which the effects of undersampled DTI data can be assessed. Also presented are results obtained using this method to consider detection of anisotropy differences using data with substantial gap thickness. Specifically, the aim is to employ masking techniques and k-space manipulations to simulate undersampled DTI datasets and to determine the performance characteristics of voxelwise analyses for these data sets when compared to normally sampled, full coverage, datasets.

Background

K-space

Simulation of datasets in this study was accomplished performing a discrete Fourier transform on the image data and manipulating the resulting k-space information followed by and inverse Fourier transform back into the image domain. MRI applications require discrete sampling, so discrete Fourier and inverse Fourier transform functions are typically used. The mathematics behind these transforms are the fundamental basis for MRI acquisition (Sorenson, 1998) and also allow manipulations of the data based on the different domain properties. Equations one and two are the discrete forms of the Fourier transform and its inverse.

$$F(k) = \sum_{n=1}^{N} f(x) e^{-i2\pi(k-1)(n-1)/N} \qquad (1)$$

$$f(x) = \sum_{n=1}^{N} F(k) e^{i2\pi(k-1)(n-1)/N} \qquad (2)$$

Linear shifts in image space can be accomplished by applying a linear phase component to k-space in the direction of desired shift. This is shown below:

$$F'(k_z) = F(z) e^{-i2\pi k_z R/N} \qquad (3)$$

Because N, the number of discrete samples taken from k-space in the z-direction, determines the

resolution, the R/N term will control the shift of the image in the image domain and this will be equal to R voxels. Because R may be a fraction, manipulation of the data in k-space will effectively allow reslicing of the data based upon the shift applied.

A second way that alterations of k-space can change image parameters is by truncation. The truncation of the field of view in k-space results in a resolution decrease in the image domain formalized by the following equation:

$$\Delta x = \frac{1}{\Delta k_x N_x} \qquad (4)$$

Here Δx is the voxel size in one dimension, Δk_x is the size of the sampling interval in k-space in that same dimension and N_x is the number of k-space samples taken. Reducing the number of k-space samples taken and maintaining the sampling interval increases the voxel dimension. This extends also to higher dimensions.

Receiver operating characteristic analysis.

The statistical approach used to compare the performance of undersampled data to detect different simulated differences, was ROC analysis. This is commonly used to visually describe the performance of classifiers, such as medical diagnostic tests. The basis for this approach is fairly straightforward (For a review see Fawcett, 2000) and stems from the confusion matrix (shown in figure 1a). The True positive (TP) rate is the number of events that were classified as true, which were also known to be true divided by the number of total detected true events and the false positive (FP) rate is the number of events that were classified as true, which were known to be false divide by the total number of events detected as true. Choosing from an evenly mixed population of “trues” and “falses” would yield a value of .5 for both parameters. ROC analysis visualizes the performance of the classifier by plotting the TP rate against the FP rate. The line with a slope equal to one is the visual representation of choosing at chance. This technique can be expanded to include non-discrete classification methods such as T-statistics, which report the probability that a given values belongs to a certain group. Curves can be made from the pairs of classifier value and known value and these pairs can be ordered according to descending classifier value. The ROC curve can then be constructed by incrementing the TP direction for every instance of a known true and by incrementing the FP direction for every instance of a false in the ordered known set. The curve representing a line with a slope of one will be equivalent to selecting randomly and a high performing curve will begin with a steep positive slope and remain close to the y-axis until eventually veering to the right, in other words it will hug the

upper left corner as much as possible. The area under the curve is often used to quantify the performance as well. This technique was well suited to the needs of this experiment to visualize the performance curve for voxelwise analysis using different simulated sample sets.

Methods

Data Acquisition

A relatively high resolution DTI data set was acquired on a GE 3T scanner using the following parameters: TE, TR, 38 contiguous slices with 3mm thickness and 0.98x0.98 in-plane resolution, 26 scans were acquired, one with a zero b-value and 25 with b=1000 s/m^2 at different and equally distributed gradient directions.

Simulation of Datasets

From the single high intensity acquisition, simulated datasets were created with larger voxel dimensions (4mm isotrpoic) and interslice gaps (4mm between slices). This was accomplished using matlab and spm tools (please see appendix A for code). There were six undersampled datasets, each containing twenty simulated subject sets. Each simulated subject set ultimately became one FA map, however initially each subject set contained 26 images including one T2 reference image and 25 diffusion weighted images. These were simulated from the high-resolution data set containing the same 26 images, but with the image parameters specified previously.

Two of the simulated data sets contained images with no mask manipulations and these served as control simulated subjects. Two of the data sets contained images that were simulated from the high resolution image set after application of a spherical mask. This mask was a uniform sphere, 16 mm in diameter and positioned in the posterior portion of the brain centered on the midline. The final two data sets contained images with a diffuse white matter mask applied first to the high resolution set. In both cases, one of the two data sets contained a mask with high intensity and the other with low intensity.

The basis for simulations was to apply a 3D fast Fourier transform (FFT) to the 3D image in order to convert the data into the k-space domain for alteration. Manipulations were made on k-space in two ways (based upon the mathematics discussed previously). First, the locations of slices in the z-direction were altered by applying a linear phase to k-space in the z- direction as specified in equation (5). Here R was selected randomly for each simulated subject set over a range between 0 and 2 in order to shift the data set that number of voxels in the Z-direction. This is an important parameter, because it ensures that all information in the data set will be sampled equally in the simulation.

After the phase component was applied, the k-space data was truncated and an inverse 3D FFT

was performed to transform the data back into the image domain, now with new parameters. This simulation generated 20 unique subject sets for each data set.

Tensor Calculation

The diffusion tensor for each simulated subject set was calculated using the Diffusion II toolbox (Volkmar Glauche) in spm5 (Wellcome Department of Cognitive Neurology, Institute of Neurology, London, UK). Briefly, diffusion weighted images were motion corrected and the diffusion tensor, **D**, was calculated at each voxel. From **D** and the known scanner reference frame, the three eigenvalues λ_1, λ_2 and λ_3 and eigenvectors at each voxel were determined. Fractional anisotropy (FA) maps were calculated using the standard definition as below:

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{\frac{1}{3}\left[(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2\right]}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \qquad (5)$$

Voxelwise analysis

Voxelwise statistics were performed on the simulated datasets using the matlab commands and the spm5 graphical user interface. First, the reference T2 images (b = 0) were coregistered to the anatomical T1 weighted scan of the same brain. The anatomical scan was normalized to standard stereotaxic space (MNI atlas brain avg152T1) and the same 12 parameter affine transformation was applied to the DTI maps. Registration was investigated manually to ensure appropriate alignment between data sets.

Next, the data were smoothed using a 12 mm isotropic Gaussian kernel. This step may play a crucial role in the statistical outcomes for this study and is discussed later in this paper, however this size was chosen based on the largest dimension of the image voxels, which was effectively 8mm due to missing information.

A second level design matrix was defined and estimated using the spm5 graphical user interface. The contrasts for each test were defined to subtract the non-masked images from the masked images and one T-statistic map for each test was created automatically by the software and used to create ROC curves as described below.

Validation comparisons

In order to create ROC curves, the T-statistic maps were used as the classifier and the masks used to create the original masked data sets served as the indicator of true values. These were also both

masked to include only voxel values corresponding to brain. The values for both input images (the map and the mask) were then read into a one-dimensional ordered vector. The vector containing the T-statistic information was the sorted from largest value to smallest and the index of sorting was applied to the mask vector so that each of the elements correcponded to its spatial pair from the other vector, but both were sorted according to T-statistic. Next the algorithm read in the sorted mask values and for each "true" instance, of encountering a new true positivce vector was incremented and each time the algorithm encountered a "false", or zero, event a false positive (FP) vector was incremented. Ultimately, when TP was plotted against FP, an ROC curve resulted. ROC analysis was performed for T-statistic maps created for each of the four low resolution mask conditions and the two high resolution conditions against the respective control sets. In this way performance could be determined by comparing low resolution to high resolution curves.

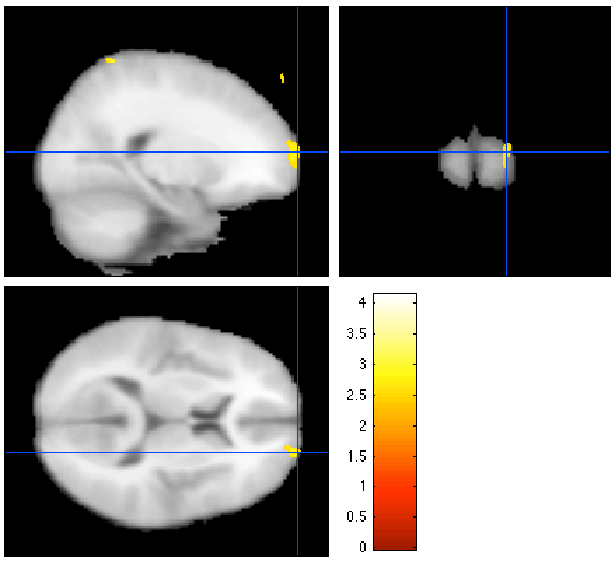

Figure 1. T-statistic map thresholded at p = 0.01 and overlayed on standard atlas. Color bar indicates T-statistic, degrees of freedom = 38

Results

The simulation produced the intended data sets with resolutions and slice locations as predicted. From these sets, it was possible to test for differences between two identical simulations, varied only by the random slice shift parameter. The results of these tests are shown in figure 1 and summarized briefly. There were some false positives reported by tests between two different iterations of the simulation algorithm (see figure 1), however, the extent of these false positives was not large and can be explained by multiple comparisons in that the number of voxels is so large that at a P-value of .05 there is expected to be 5% false positive rate.

ROC analysis

The ability of voxelwise methods to correctly report differences between groups was found to be different depending upon which mask was used. In the case of the spherical mask, there was a good performance of voxel-wise analysis for the undersampled data. This can be seen by the ROC curves shown in figure 5. The red trace is the performance curve for a full (not undersampled) data set. This

follows a steep initial incline, indicative of high levels of reporting the true value. The blue and green curves show the high and low intensity masks respectively and both curves also perform well, almost at the level of the full data set. However upon closer inspection of the T-statistic maps, it becomes evident that the bulk of the true positive signal is coming from gray matter areas. There is an anatomical correlation of report, with the white matter areas showing diminished significance for the same applied signal as the adjacent gray matter.

This behavior is further exemplified by the behavior of the analysis on data sets created using diffuse white matter masks. These show the opposite trend form the sphere masks in that the roc curve for the high resolution set is closer to the slope of one line and the ROC curve for a high intensity mask shows worse performance that that for a low intensity mask. When the statistic maps are considered (figure 4) we can see that this trend is coming from an overestimation of the difference in groups at lower mask value and in the undersampled group. There is a striking reduction in the extent of reported differences using the same smoothing, thresholding and correction parameters for the high resolution data set and also a noticeable difference in extent between the low and high intensity mask conditions.

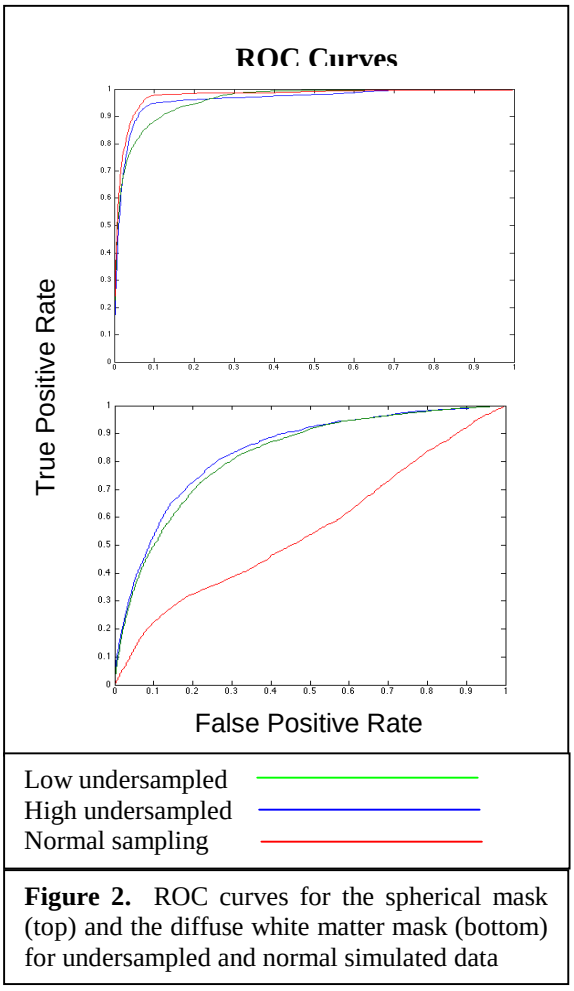

Figure 2. ROC curves for the spherical mask (top) and the diffuse white matter mask (bottom) for undersampled and normal simulated data

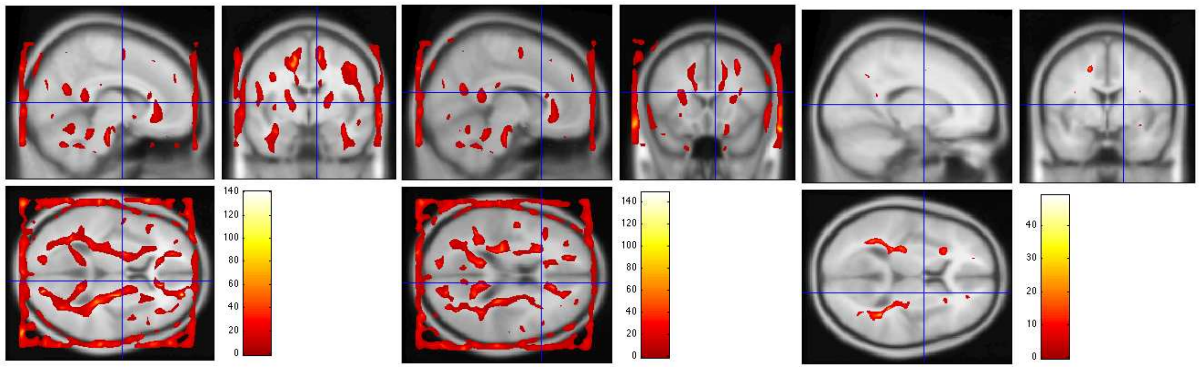

Figure 3. T-Statistic overlays on an atlas brain show the differences detected between masked and non-masked data, using the diffuse white matter mask. Two intensities of masking are shown for the undersampled data set and one intensity (high) is shown for the high resolution data. low high HR

Discussion

The primary outcome of this work was the development of a simulation and analysis technique appropriate for the characterization of DTI datasets with variable resolution and sampling parameters. The simulated data sets were able to sample slices from the entire brain equally whit user defined voxel size.

The performance of voxelwise group tests using the undersampled data was consistent with what was expected for the spherical mask in that the performance was less for undersampled data than for the full data set. However, the ROC analysis indicates that the performance of the tests on undersampled data were still reasonable and even well performing. This indicates that undersampled data has a small effect on the detection of large focal differences, but the effect is not substantial enough to exclude undersampled data from being used for this purpose.

Diffuse white matter analyses were different from what was expected in that the normally sampled data set showed far worse performance that the undersampled data sets. This may be due to the nature of the statistics used, as the number of voxels contained in the normally sampled set is double that in the undersampled set. Although more mask types need to be considered before making any further inferences,it is tempting to speculate that the undersampling of white matter results in an exaggeration of effect.

The technique developed in this work may be applied to multiple variations of resolution and undersampling designs as well as to other diffusion parameters. The flexible framework allows for numerous manipulations to determine the effects of parameters such as registration, smoothing and tensor calculation on performance of voxelwise methods. In all, this algorithm is simple, but effected for its intended purpose.

References

1. Snook L, Plewes C and Beaulieu C. *Voxel based versus region of interest analysis in diffusion tensor imaging of neurodevelopment.* Neuroimage. 2007 Jan 1;34(1):243-52. Epub 2006 Oct 27
2. C. A. Cocosco, V. Kollokian, R. K.-S. Kwan and A. C. Evans, Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage* **5** (1997), p. S425
3. K. J. Worsley, M. Andermann, T. Koulis, D. MacDonald and A. C. Evans, Detecting changes in non-isotropic images. *Hum. Brain Mapp.* **8** (1999), pp. 98–101.
4. Kim M, Ronen I, Ugurbil K and Kim DS. *Spatial resolution dependence of DTI tractography in human occipito-callosal regio*n. Neuroimage. 2006 Sep;32(3):1243-9. Epub 2006 Jul 24
5. A.L. Alexander, K.M. Hasan, M. Lazar, J.S. Tsuruda and D.L. Parker, Analysis of partial volume effects in diffusion-tensor MRI, *Magn. Reson. Med.* **45** (2001) (5), pp. 770–780
6. J. Ashburner and K. Friston, *Voxel-Based Morphometry – The Methods*, NeuroImage 11, 805–821 (2000)
7. T. Fawcett, *ROC Graphs: Notes and Practical Considerations for Researchers* 2004, Kluwer Academic Publishers, pp. 1-38
8. AR Luft, M Skalej, D Welte, R Kolb and U Klose, *Reliability and exactness of MRI-based volumetry: a phantom study* J Magn Reson Imaging. 1996 Jul-Aug; 6(4):700-4
9. D. Le Bihan, J.F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko and H. Chabriat. *Diffusion tensor imaging: concepts and applications,* J Magn Reson Imaging. 2001 Apr; 13(4):534-46
10. J. Sorenson *Physics of Magnetic Resonance Imaging,* formal course notes, University of Wisconsin, Madison, copyright 1998, Chapters 2 and 11.