Robust and efficient white matter analysis using tract shape modelling and principal components analysis

J. D. Clayden[1]
[1]Institute of Child Health, University College London, London, United Kingdom

Introduction

The problem of correction for multiple comparisons is a pervasive one in the statistical analysis of magnetic resonance image data. With regard to diffusion MRI (dMRI) in particular, methods such as tract-based spatial statistics have been developed partly with the aim of reducing the multiple comparisons problem faced by more general voxelwise analysis methods [1]. For many studies, analysis at the scale of whole white matter tracts is sufficient to investigate the hypotheses of interest, but even in this case there is often a substantial amount of shared variance between structures due to the influence of exogenous common factors [2,3]. In this abstract we describe how probabilistic neighbourhood tractography (PNT), an automated approach to tract segmentation based on a statistical tract shape model [4], along with principal components analysis (PCA; see [5]), can be used to generate and analyse tract-based information in a way that is robust and efficient with data.

Methods

Eight healthy young adults (4 male, mean age 32.0 ± 5.6 yr) underwent a diffusion MRI protocol on a GE Signa LX 1.5 T clinical system. Echo-planar diffusion-weighted images were acquired along 64 noncollinear directions at a *b*-value of 1000 s mm^{-2}, along with 7 *b*=0 images. Reconstructed image resolution was 2 x 2 x 2 mm. Scan time was approximately 20 min.

Standard dMRI preprocessing steps, including correction for eddy-current induced distortions and brain extraction, were carried out on each data set using FSL tools (http://www.fmrib.ox.ac.uk/fsl). Diffusion tensors were estimated in each voxel using standard least-squares fitting, and mean diffusivity (MD) and fractional anisotropy (FA) maps were generated.

The PNT method, as implemented in the TractoR software package (http://code.google.com/p/tractor), was used to segment the arcuate fasciculi, cingulum bundles, corticospinal tracts (CSTs), inferior longitudinal fasciculi (ILFs), uncinate fasciculi, and corpus callosum genu and splenium, using standard reference tracts provided with TractoR. False positive streamlines from the final tractography step were pruned probabilistically as described previously [6]. Mean FA and MD within each tract, weighted by the visitation count after pruning, were calculated in each case. PCA was performed on FA and MD measurements independently, and component scores related to age and gender for illustration.

Results

Fig. 1 shows the correlation matrix for mean MD across all of the tracts segmented in this study. It can be immediately observed that there is a high degree of correlation between tracts, as has been reported previously [2,3], going well beyond links between equivalent tracts in the left and right hemispheres. In light of this, we may expect PCA to allow us to represent most of the variability in the data set more compactly. Indeed, Fig. 2 shows that more than 40% of the variance in FA and MD may be explained using a single principal component (PC) in each case. Taking the first two components will in both cases capture well over half of the variance, while there are three components with more than the 12.5% that would be expected if the variance were spread evenly over the components (dashed line in Fig. 2).

The loadings for each PC, or factor, establish the links back to the original data (Fig. 3). For MD, we observe that every tract except the splenium is positively loaded on the most important first component, indicating a high degree of commonality (cf. Fig. 1). For FA, the pattern is less consistent—the arcuates, CSTs, right ILF, left uncinate and splenium behave differently from the remaining tracts.

Together with the loadings, the principal component scores for each subject allow the original data to be reconstructed. These scores—the value of each factor for each subject—may be used in subsequent analysis rather than the raw FA and MD values to reduce the multiple comparisons correction burden. For example, we may observe using ANCOVA that the first PC in MD differs between the genders in this data set ($F_{1,4} = 8.06$, $P < 0.05$), and although there is no main effect of age in days, it does interact with gender ($F_{1,4} = 16.41$, $P < 0.05$). FA PC1 does not show these effects. (We do not attach much significance to these findings in such a small data set—rather, this analysis is given as an example of how the PCA scores may be used.)

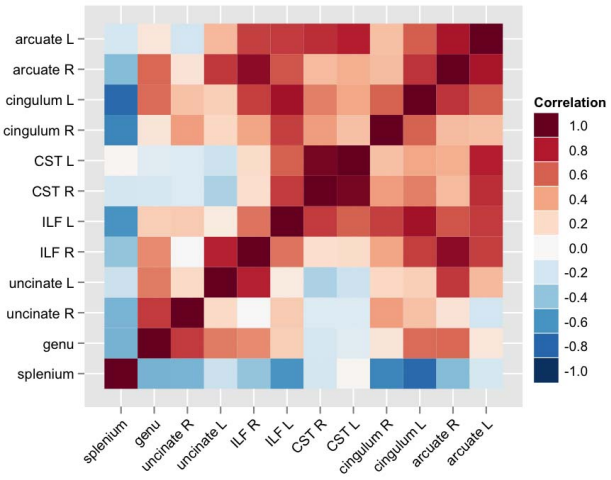

Fig. 1: *Correlation matrix for mean MD within the 12 tracts in this study. Substantial covariance between tracts is visible.*

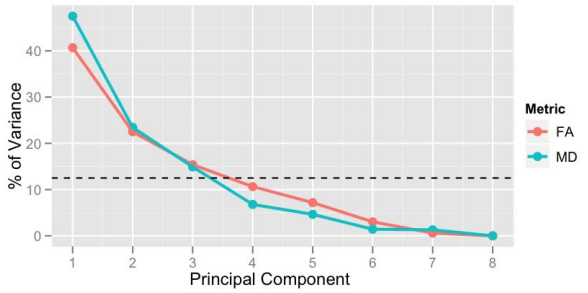

Fig. 2: *Scree plots showing the proportion of variance associated with each principal component in FA and MD for our data set.*

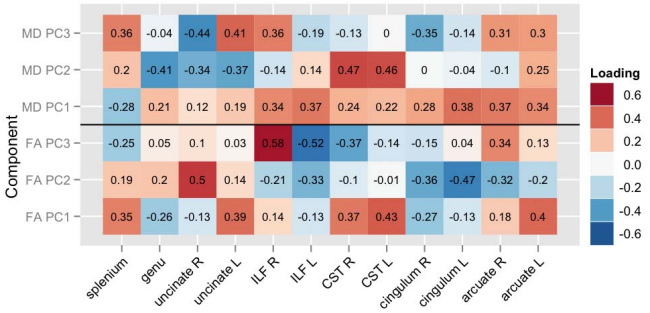

Fig. 3: *Loading matrix for the first 3 principal components in FA and MD.*

Discussion

Here we have described how probabilistic neighbourhood tractography and principal components analysis may be used as a powerful combination for white matter analysis. By applying the PCA transformation to FA and MD data, the major sources of variance may be identified and linked to exogenous factors of interest [7]. This approach is robust and efficient with data, requiring minimal correction for multiple comparisons.

References: **[1]** S.M. Smith et al., *NeuroImage* **31**:1487 (2006); **[2]** M. Wahl et al., *NeuroImage* **51**:531 (2010); **[3]** L.T. Westlye et al., *Cereb Cortex* **20**:2055 (2010); **[4]** J.D. Clayden et al., *NeuroImage* **45**:377 (2009); **[5]** N. Bratchell, *J Chemometr* **3**:579 (1989); **[6]** J.D. Clayden et al., *Lect Notes Comp Sci* **5762**:150 (2009); **[7]** L. Penke et al., *J Neurosci* **30**:7569 (2010).