10

Conclusions

The aims of this thesis were to develop methods to facilitate the robust segmentation of specific white matter structures from multiple dMRI brain volumes, and the subsequent comparative analysis of the segmented regions. In this final chapter we review the extent to which the work described in the previous chapters has met these aims, and discuss the work that still needs to be done.

10.1 Tract segmentation

As we described in the introduction, the study of structural human brain connectivity *in vivo* really only began with the invention of DTI in the mid-1990s. Over the course of less than a decade, since the possibility of using tensor-derived metrics to probe white matter integrity took hold, a sizeable clinical literature based on the method has amassed; but the techniques are still quite immature. Ideally, one would begin studying a disease in which a loss of connective efficacy is a suspected factor by applying a whole-brain analysis technique such as VBM to suggest regions of localised contrast between patient and control populations. A replication study might then hope to characterise the effect on any implicated white matter structures more clearly, and look for evidence that particular white matter degradation is specifically linked to the pathology in question. Unfortunately the reality is less straightforward.

When applied to maps of diffusion anisotropy, the VBM method is not robust. As we discussed in §6.1, the choice of smoothing kernel can have a very substantial effect on the results—not just quantitatively, but qualitatively too, with regions of contrast appearing in quite different brain areas as the kernel width is altered. Since this parameter of the method is usually chosen for each individual study without recourse to any firm principles, the scope for spurious and misleading results is unsettling. Moreover, it is easy—although unwise—to forget that FA itself has limitations as a proxy for integrity. We discussed in §9.4 that FA would be expected to *increase* if one of a pair of crossing fibre populations were to be preferentially degraded, and therefore it cannot necessarily be trusted as a reliable indicator of disease in crossing fibre regions. Even if this shortcoming did not exist, the expressiveness of a single scalar parameter will always be limited.

Tract-based spatial statistics may in practice take over as the method of choice when looking for localised differences between populations with limited or no prior knowledge. Its only overt parameter is an FA threshold that is applied to the skeletonised anisotropy maps, which will usually affect the results only quantitatively due to its impact on the number of voxels surviving to the multiple comparisons correction stage. The technique does perform time-consuming nonlinear registration of each subject's brain volume to every other, which makes it scale badly to large data sets, but the introduction of a standardised template skeleton might be possible to remove this issue. TBSS is certainly an attractive approach, although it is not truly "tract-based" since the skeletonisation process will find any ridge in the anisotropy map and has no concept of white matter structure or connectivity. For most purposes, the approximation is however an adequate one.

On the other hand, automated methods for tract-specific segmentation and comparative analysis are more or less nonexistent. Regions of interest can be defined in standard space and then transferred to native space using registration, and used to constrain tractography; but this multiple ROI approach has a number of drawbacks, as we discussed in §6.5. Like any registration-based transformation, this one will engender some inaccuracy in the placement of the ROIs in native space; but in any case these ROIs encode prior knowledge about the topology of tracts in an unintuitive manner, which is informed primarily by experience with tractography rather than direct knowledge of anatomy. The use of "termination" and "removal" masks by Heiervang *et al.* (2006), for example, is presumably founded on past experience, during which the authors observed some pathways straying into these regions and deemed them aberrant or undesirable. The problem is that the ROIs might need to be redrawn for use with a different tractography algorithm.

We would argue that our representation of prior knowledge about tract topology in terms of reference tracts is more intuitive, more transferrable and ultimately more reliable. Information about the expected route of the tract is given along its entire length, but this richer prior information is not used to directly constrain the fibre tracking algorithm—rather, it guides the choice of tractography results from among a number of candidate seed points. The combined process of matching tracts to a reference and choosing a segmentation *a posteriori* based on these matches is neighbourhood tractography, a largely automated approach that we have invented and refined over the course of the thesis.

In chapter 6, we described a heuristic similarity measure for matching tracts and outlined the principle of NT. We demonstrated that the method improved segmentation consistency over a naïve alternative method in a group of healthy volunteers; and then, in chapter 7, we found similar benefits in a healthy agèd cohort. We were able to use a reference tract taken from an agèd brain to successfully select tracts from the younger group, and thereby to show anisotropy differences between the groups in a specific tract where previous whole-brain studies have suggested that one might be present. We have also discussed how standardised reference tracts can be generated from a white matter atlas, and used these references in a practical study.

To ameliorate some of the shortcomings of our simple first approach to tract matching for NT, we reformulated the problem in formal probabilistic terms in chapter 8, and took a machine learning perspective toward its solution. The models that we used to represent the relationships between matching tracts were parameterised such that the extent of deviation from the route of the reference tract can vary along its length, meaning that large variability within the data used to fit the model will result in only small penalties for straying from the reference. To learn suitable parameters, we initially took a supervised maximum likelihood approach, in which a group of training tracts is selected by hand in addition to the reference tract; but later showed that an EM algorithm could be used to successfully find matches in a data set without a separate training phase.

The main parameter of NT methods is the neighbourhood width. If this is set too small then no appropriate match to the reference will be found, and if it is set too large then the process will take a very long time to run. The limiting case of seeding throughout the brain is theoretically optimal in the sense that if a matching tract can be produced then it should be found this way—unless there happens to be another fasciculus with very similar shape and length in another part of the brain—but the practical consideration of run time makes this an unwise strategy. In any case, the use of tract similarity measures or matching models gives us an indication of the acceptability of the best match that we can use to our advantage. As we mentioned in §8.6, the null-match posterior probability that is available in the unsupervised probabilistic case could be used as the basis of a rejection criterion. To minimise run time, the neighbourhood width could be chosen separately for each brain volume, being increased in steps until the null-match posterior drops below a certain level. Nonetheless, a proper analysis of the effect of neighbourhood width would be a useful avenue for future work.

Other parameters arising in the model-based NT methods, such as the residual error threshold, η, and the streamline length quantile, ξ, may also have some effect on the outcome. But the former is relevant only to the reference tract, and we have found no reason to vary the latter from one brain volume to another, so in practice there should be little reason for them to

vary between studies and therefore become a point of weakness in any results.

We have not yet had time to apply the probabilistic model-based variants of NT to clinical data sets, or to develop atlas-based reference tracts for use with them; and these are important areas for future work. With them in place, however, we feel that the approach could represent a useful, robust and automated technique for the segmentation of specific tracts.

10.2 Comparative analysis

Once similar regions are segmented from a number of brain volumes, the simplest approach to comparative analysis between groups is to average a scalar measure of interest within each region and statistically compare the range of values thus obtained. This average can be weighted using the voxelwise likelihoods of connection to the seed point produced by an algorithm such as FSL ProbTrack—as we did in chapter 7. Using probabilistic neighbourhood tractography, we can also include data derived from multiple seed points, weighting according to the corresponding matching posteriors as in §8.5. It would be constructive to examine the benefits (or otherwise) of these weighting schemes more closely than we have done above.

In chapter 9, we explored the possibility of profiling anisotropy along tracts rather than simply averaging its value within the regions representing the relevant fasciculus. This raises some difficult questions about point homology in different brain volumes, but our initial results nevertheless suggest that this kind of approach may be able to yield some additional meaningful information.

Ultimately, comparing anisotropy between populations—however it is done—is only going to take *in vivo* white matter studies so far. Mean diffusivity is a mathematically independent measure for characterising diffusion, but in practice it is generally negatively correlated with FA. Combination of diffusion data with information from other magnetic resonance methods may prove more fruitful: spatially localised brain "activation" data from functional MRI, or metrics derived from magnetisation transfer imaging (which was briefly mentioned in §4.5) may help, if the concomitant coregistration issues can be worked out. Even more broadly, there is scope for incorporating data from fields such as genetics into advanced studies.

10.3 Final remarks

It is our hope that the methodological developments set out in this thesis will be helpful for ongoing work investigating whether connective changes are systematically linked to outwardly visible pathology. We believe that we have made useful progress towards robust segmentation of tracts of interest, which, so long as it remained problematic, has been a distracting prerequisite for meaningful investigation of the differences and similarities between comparable white matter structures.

As its resolution and noise properties improve, the potential of DMRI should continue to increase, although these developments will probably bring new challenges as well. Methods for examining connectivity may need to become more sophisticated, but a definitive test of the disconnection hypothesis could be at hand in the foreseeable future.