



## **TractLearn: A geodesic learning framework for quantitative analysis of brain bundles**

Arnaud Attie, Felix Renard, Monica Baci, Elise Roger, Laurent Lamalle, Patrick Dehail, Helene Cassoude, Fernando Calamante

### **► To cite this version:**

Arnaud Attie, Felix Renard, Monica Baci, Elise Roger, Laurent Lamalle, et al.. TractLearn: A geodesic learning framework for quantitative analysis of brain bundles. NeuroImage, 2021, 233, pp.117927. <10.1016/j.neuroimage.2021.117927>. <hal-03222320>

**HAL Id: hal-03222320**

**<https://hal.science/hal-03222320v1>**

Submitted on 10 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License



# TractLearn: A geodesic learning framework for quantitative analysis of brain bundles

Arnaud Attyé, Conceptualization Methodology Software Investigation Resources Coding & Writing<sup>a,b,\*</sup>, Félix Renard, Conceptualization Methodology Software Investigation Resources & Coding<sup>a</sup>, Monica Baciú, Software Review & Editing<sup>a</sup>, Elise Roger, Software Review & Editing<sup>a</sup>, Laurent Lamalle, Review & Editing Visualization Software Server administration<sup>c,d</sup>, Patrick Dehail, Data provider review & editing<sup>e,f</sup>, Hélène Cassoudeulle, Data provider Review & Editing<sup>e,f</sup>, Fernando Calamante, Conceptualization Methodology Writing Review Editing & Supervision<sup>b,g</sup>

<sup>a</sup> Neuroradiology and MRI, Grenoble Alpes University Hospital, Grenoble, France

<sup>b</sup> School of Biomedical Engineering, The University of Sydney, Sydney, NSW 2006, Australia

<sup>c</sup> University Grenoble Alpes, Grenoble, France

<sup>d</sup> IRMaGe, Inserm US 17, CNRS UMS 3552, Grenoble, France

<sup>e</sup> Physical and Rehabilitation Medicine department, University Hospital of Bordeaux, Bordeaux, France

<sup>f</sup> 2EA4136-HACS, The University of Bordeaux, Bordeaux, France

<sup>g</sup> Sydney Imaging – The University of Sydney, Sydney, Australia

## ARTICLE INFO

### Keywords:

Diffusion MRI  
Fiber tractography  
Precision medicine  
Manifold learning

## ABSTRACT

Deep learning-based convolutional neural networks have recently proved their efficiency in providing fast segmentation of major brain fascicles structures, based on diffusion-weighted imaging. The quantitative analysis of brain fascicles then relies on metrics either coming from the tractography process itself or from each voxel along the bundle.

Statistical detection of abnormal voxels in the context of disease usually relies on univariate and multivariate statistics models, such as the General Linear Model (GLM). Yet in the case of high-dimensional low sample size data, the GLM often implies high standard deviation range in controls due to anatomical variability, despite the commonly used smoothing process. This can lead to difficulties to detect subtle quantitative alterations from a brain bundle at the voxel scale.

Here we introduce *TractLearn*, a unified framework for brain fascicles quantitative analyses by using geodesic learning as a data-driven learning task. *TractLearn* allows a mapping between the image high-dimensional domain and the reduced latent space of brain fascicles using a Riemannian approach.

We illustrate the robustness of this method on a healthy population with test-retest acquisition of multi-shell diffusion MRI data, demonstrating that it is possible to separately study the global effect due to different MRI sessions from the effect of local bundle alterations. We have then tested the efficiency of our algorithm on a sample of 5 age-matched subjects referred with mild traumatic brain injury.

Our contributions are to propose:

1/ A manifold approach to capture controls variability as standard reference instead of an atlas approach based on a Euclidean mean.

2/ A tool to detect global variation of voxels' quantitative values, which accounts for voxels' interactions in a structure rather than analyzing voxels independently.

3/ A ready-to-plug algorithm to highlight nonlinear variation of diffusion MRI metrics.

With this regard, *TractLearn* is a ready-to-use algorithm for precision medicine.

**Abbreviations:** TWI, Track-weighted imaging; GLM, General linear model; HDLSS, High dimensional low sample size data; UMAP, Uniform Manifold Approximation and Projection.

\* Corresponding author at: Neuroradiology and MRI, Grenoble Alpes University Hospital, Grenoble, France.

E-mail address: [arnaudatty@gmail.com](mailto:arnaudatty@gmail.com) (A. Attyé).

<https://doi.org/10.1016/j.neuroimage.2021.117927>

Received 20 May 2020; Received in revised form 25 February 2021; Accepted 1 March 2021

Available online 6 March 2021

1053-8119/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Introduction

A streamline data-set generated from diffusion MRI provides a wealth of information regarding structural connectivity between brain regions in medicine and neurosciences. An elegant approach to segment brain bundles has recently been proposed, TractSeg, based on a combination of three convolutional neural networks (Wasserthal et al., 2019). The first network proposes to segment major bundles using a voxel-wise binary classification to discern tract and non-tract voxels. The second network allows to learn the start and end regions of each brain fascicle, while the last one computes tract orientations mapping to obtain a single 3D peak vector per voxel. This architecture is then employed to start a probabilistic tracking algorithm for 72 brain bundles, tested in various acquisition conditions and on a few disease models. We could consider this segmentation process as a first step to reduce the dimensionality of data coming from a whole-brain diffusion MRI acquisition.

Usually, group study statistical analyses rely on the General Linear Model (GLM), a standard tool based on Gaussian distribution assumptions that returns for each voxel the mean of the control group. The subsequent step is either t-tests (in case of group vs. group studies) or Z-score (in case of the comparison between one individual versus a group); this is usually done analyzing each voxel independently. In addition, precision medicine requires to have algorithms suitable to manage high dimensional low sample size data (HDLSS). A potential limitation of HDLSS is the difficulties to capture all the physiological variation of MRI contrasts using a Euclidean mean, leading to high standard-deviation values (see supplementary figures 1 and 2, for a simulated illustrative example). It can theoretically limit the possibility to distinguish bundle subtle quantitative alterations.

One way to address these two limitations is by manifold learning. Manifold learning is a class of machine learning methods that is gaining success and attracting interest, allowing reconstruction of a manifold sub-space to represent, understand and visualize degrees of freedom of complex quantitative data (Tenenbaum et al., 2000). It is classically used before deep learning algorithms to identify outliers, for example to increase the robustness of images reconstructions (Zhu et al., 2018). While classical deep learning algorithms allow to segment, identify, and classify imaging features based on big data, a manifold framework usually requires multiple quantitative biomarkers per region of interest (in our case, white matter bundles) to estimate distances between subjects, while still remaining applicable to studies with a low number of imaging data.

UMAP (Uniform Manifold Approximation and Projection) is a recent nonlinear manifold learning technique for dimensionality reduction, which is constructed from a theoretical framework based on Riemannian geometry approximation and algebraic topology (McInnes et al., 2018). It constructs a high dimensional representation of the data, and then optimizes a low-dimensional graph while preserving the structural similarity of the data as much as possible, thus providing a Riemannian approximation of the original data. It aims at finding the shortest path between samples, whose length is called the geodesic distance. Compared to linear dimensionality reduction, geodesic approaches treat the geometry of all the voxel diffusion MRI metrics contained in brain white matter bundles as Riemannian manifolds<sup>1</sup>. Therefore, UMAP takes advantage of the Riemannian nature of the manifold when comparing two bundles in a reduced subspace, and the shortest path (here path refers to the difference in the Euclidean distance between 3-dimensional points in the reduced manifold) between the two bundles reflects the nonlinear similarity between these bundles.

A Riemannian manifold is a smoothed curved manifold, based on the assumption that the surface's curvature is solely determined by computing distances over the surface.

Previously published work on Riemannian framework for diffusion MRI analysis either exploited the geometry of diffusion tensor images to propose alternative diffusion MRI metrics (e.g. Batchelor et al., 2003; Fuster et al., 2014) or used it to improve the tractography segmenta-

tion step (e.g. Hao et al., 2014; Hauberg et al., 2015). Here we use the Riemannian framework for a different approach and will demonstrate that usual diffusion MRI metrics can also take advantage of a nonlinear global analysis.

The information contained in the streamline tractograms can also be exploited to generate further diffusion MRI metrics using the track-weighted imaging (TWI) framework (which can have super-resolution properties) (Calamante, 2017; Calamante et al., 2012). For example, TWI maps can be generated from various properties of the streamlines themselves, such as maps related to the number of streamlines in each voxel or their average length (Calamante, 2016; Pannek et al., 2011). Alternatively, TWI maps can also be computed from track-weighted version (e.g. an average along the track) of the contrast of an associated parameter (e.g. a fractional anisotropy map). The TWI contrast has proved to be useful for clinical care in providing fast identification of major alterations in brain fascicles (e.g. tumor proliferation (Barajas et al., 2013) or demyelination (Lyksborg et al., 2014)). Yet, it might be also relevant to highlight subtler quantitative alterations at the global level.

As the quantitative and reliability aspects of the TWI contrast has only been investigated in the context of whole brain tractography (Calamante et al., 2015; Willats et al., 2014), new test-retest data manifold learning analysis would be interesting to compare diseased individuals to a healthy cohort within a Riemannian framework. In other words, we raise the hypothesis that manifold learning can be used to detect local abnormalities in diffusion MRI metrics of white matter bundles, and that the detected abnormalities will be robust to the healthy control data set used (i.e. test/retest data) as reference.

Here, we propose a 3-step method to obtain a fast-quantitative analysis of brain bundles, through various quantitative diffusion MRI metrics (either standard diffusion MRI metrics or TWI metrics). Firstly, we will reduce the dimensionality of all voxels contained in each bundle, to have one point per bundle and subject in a manifold subspace. Secondly, we will learn the manifold from the test and the retest session of our healthy controls independently (Tilquin et al., 2019) before projection of each individual patient (mild trauma patients in our study) onto the learned manifolds. Finally, we propose to apply Riemannian Z-scores to the projection of a new subject onto the learned manifolds (from the test and retest sessions) before calculating inter-session agreement to detect local abnormalities.

## Methods

### Data and code availability statements

The code for *TractLearn* is available on <https://github.com/geodaisics/TractLearn>

The code for *UMAP* (McInnes et al., 2018) is also freely available on <https://github.com/lmcinnes/umap> and the TractSeg pre-trained DWI algorithm, openly available at <https://github.com/MIC-DKFZ/TractSeg> (Wasserthal et al., 2018a).

### Data acquisition

20 healthy males (mean age 20.9 [SD 3.3] years) were recruited for the study. Informed written consent was obtained from all subjects in accordance with ethical approval from the local human research ethics committee of Bordeaux university (France, IRB number 2016-A00765-46). MRI data were acquired on a 3 T Siemens Prisma scanner using a 48-channel head coil. High angular resolution diffusion imaging (HARDI) datasets were acquired for each subject over two sessions with a delay of 12 months. Five age-matched subjects referred with mild traumatic brain injury (mTBI) were also included with the same MR protocol and ethical statements, for one unique MR session acquired between the time of the test and retest session for the controls.

Diffusion-weighted imaging (DWI): single-shot spin-echo sequence; TE/TR: 80/4700 ms; voxel size: 1.5 mm isotropic; 99 slices; multiband

factor: 3; scan time: 7.5 min. In addition to 10 non-DWI volumes (which were averaged), 3 DWI shells were acquired, each with a different diffusion weighting and a unique set of diffusion-weighted directions. The set of directions was independently generated for each shell by electrostatic repulsion, as follow: directions/b-value (in  $\text{s/mm}^2$ ) = 60/2000, 15/800, and 10/300.

In order to allow pre-processing minimization of distortion artefacts, two additional  $b = 0 \text{ s/mm}^2$  images were acquired before each HARDI acquisition, with identical imaging parameters as above, but one had its phase encoding reversed to allow for susceptibility distortion correction (Holland et al., 2010).

T1-weighted anatomical images were acquired using the three-dimensional magnetization-prepared rapid gradient echo (3D MPRAGE) sequence (Mugler and Brookeman, 1990) with the following parameters:  $256 \times 256 \times 192$  matrix; 0.9 mm isotropic resolution; TE 2.6 ms; inversion time TI 900 ms; TR 1900 ms; flip angle  $9^\circ$ . Susceptibility-weighted imaging and 3D-FLAIR sequences were also acquired.

### Data pre-processing

DWI preprocessing included denoising of data (Veraart et al., 2016), eddy current correction and motion correction (Andersson et al., 2003), bias field and Gibbs artefacts corrections (Tustison et al., 2010), and up-sampling DWI spatial resolution in all three dimensions using cubic b-spline interpolation, to a voxel size of 1 mm isotropic (Raffelt et al., 2012a). We have estimated fiber orientation distributions (FODs) using the Constrained Spherical Deconvolution (CSD) model (Tournier et al., 2007) using group response function (RF); in particular, we used the multi-shell 3-tissue CSD variant (Jeurissen et al., 2014). We derived the SH peaks from the FOD maps. Spatial correspondence was achieved by first generating a group-specific population template with an iterative registration and averaging approach (Raffelt et al., 2011) using FOD images from 45 MR scans (5 mTBI, 20 healthy control subjects acquired in a test session, and 20 healthy control subjects acquired in a retest session). Each subject's FOD image was then registered to the template via a FOD-guided non-linear registration (Raffelt et al., 2011, 2012a).

All preprocessing steps were conducted using commands either implemented within MRtrix3 (Tournier et al., 2019) ([www.mrtrix.org](http://www.mrtrix.org)), or using MRtrix3 scripts that interfaced with external software packages.

### TractSeg deep learning bundle specific tractography

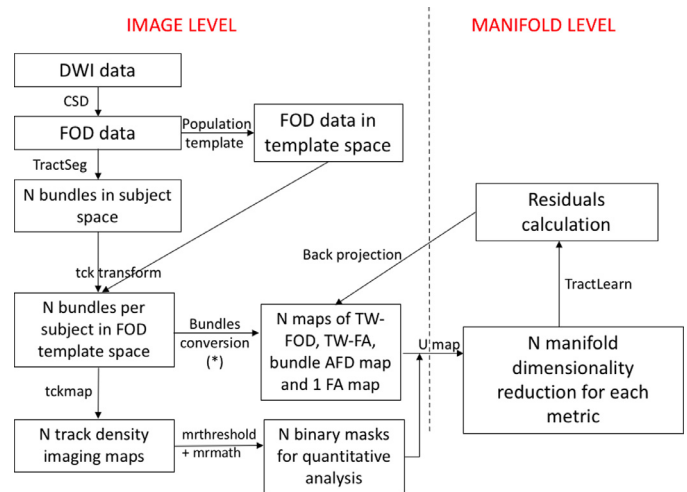
We have used the TractSeg pre-trained algorithm, to automatically identify 72 white matter bundles in each subject. Briefly, authors have proposed a custom probabilistic tracking algorithm that samples from a Gaussian distribution with fixed standard deviation centered on each spherical harmonic peak. They have used three convolutional neural networks (tract segmentation, start/end region segmentation and tract orientation mapping), all based on U-Net (Ronneberger et al., 2015) that receives as input the fiber orientation distribution function (FOD) peaks.

Importantly, for our application, we have warped all the resulting TractSeg tracks from each individual space into the common FOD template space, before the subsequent steps of our analysis.

### Estimating track-weighted imaging

As the manifold approach requires an accurate voxel to voxel matching between subjects, we first converted each track file into TWI maps based on the number of streamlines (also known as track-density imaging (TDI) maps (Calamante et al., 2010)). The objective was to crop the bundle masks, only keeping the voxels with top 80% TDI values (a value that was empirically chosen). This allows to decrease potential misregistration of brain bundles between subjects by removing the smallest cortical terminations, which can be highly variable at the group level.

Secondly, we computed TWI maps using a Gaussian-smoothed kernel for computing the parameter along each streamline; a full-width



**Fig. 1.** Overview of the workflow put forward in this work. The bundle-to-image conversion (\*) was obtained in MRtrix either using *tckmap* (for TW-FOD and TWI-FA contrasts) or *afdconnectivity* (for the AFD-related contrast). The FA contrast was obtained directly from the FA map. N represents the number of bundles successfully reconstructed using TractSeg. In the example data set included in this study,  $N = 68$  (from the original total 72 TractSeg bundles, see Section IIII for further details).

half-maximum kernel was set to 8 mm, as an empirical compromise between spatial smoothing (for increase signal to noise) and along-tract spatial blurring (to be able to detect localized effects) (Calamante, 2017; Willats et al., 2014). In particular, for the results presented here, we considered two types of TWI maps: TWI based on the Fractional anisotropy (TW-FA), and TWI based on the amplitude of the FOD (TW-FOD) along the direction of the track (Willats et al., 2014). Finally, for comparison purpose of not using the TWI approach, we have extracted from each bundle the voxel values from the FA and an FOD related map. To construct the FOD-related map (Raffelt et al., 2012b), we used the *afdconnectivity* command in MRtrix, which provides an estimate of the fibre volume of the pathway of interest at each voxel.<sup>1</sup> Fig. 1 shows an overview of the workflow of the analysis.

### Modeling brain fascicles quantitative values with manifold learning

We used a strategy similar to that developed in (Tilquin et al., 2019) to localize an abnormality among TractSeg bundles. Note that while TractLearn uses a Riemannian framework to estimate the distance between all diffusion MRI metrics contained in brain bundles, the statistics in the low-dimensional reduced manifold space (for example the distance between a point corresponding to a patient and a point corresponding to a healthy control in the manifold subspace) are then performed following classical Euclidean metrics. Here we assume that the residuals in the reduced subspaces follow a classic Gaussian law (Titsias and Lawrence, 2004).

TractLearn can be summarized as follow

- (i) Low-dimensional reduction to convert collection of voxels quantitative values from each bundle into a unique point in a manifold subspace. To test the effect of dimension reduction on the MRI data, we

<sup>1</sup> In MRtrix, *afdconnectivity* can be used to compute a map of the AFD estimated for each voxel, based on the sum of the FOD lobes aligned with the direction of the streamlines, divided by the streamline length (see [https://mrtrix.readthedocs.io/en/latest/concepts/afd\\_connectivity.html](https://mrtrix.readthedocs.io/en/latest/concepts/afd_connectivity.html) for more details).

performed optimization-based methods. We have chosen the number of dimension by repeating the reduction process using an increasing number of dimensions and evaluating whether incorporating more components achieves a significantly lower value of the loss function that the method minimizes (Nguyen and Holmes, 2019).

- (ii) Here we have used *UMAP* (McInnes et al., 2018). *UMAP* is a manifold learning approach, where the degrees of freedom of the data are captured by the latent variables. Importantly, the structure of points in the latent space (the reduced space) mimics the structure of data in the original space. Interpoint distances in the reduced space reproduce as much as possible interpoint distances in the original space; Euclidean and geodesic distances are respectively used. We investigate two issues: (a) interpreting the latent variables, and (b) determining the effect a change in the latent variables incurs in the MRI space, i.e. the corresponding changes in identified brain bundles voxels.
- (iii) We learn the manifold based on the data from the healthy controls:

$$Y = f(x) + \epsilon.$$

$Y$  being healthy control data in real space (i.e. quantitative values extracted from each *TractSeg* bundle),  $x$  the corresponding point in the reduced space, and  $\epsilon$  the residuals;  $f$  will be the regression equation between the reduced space and the real space.

The projection  $f$  of a new subject will correspond to the image closest to that of the tested subject, while belonging to the manifold of the healthy controls.

It is worthwhile to note that, in the Riemannian framework,  $f(x)$  represents the local mean value (i.e. the mean value of the closest healthy subject) of the tested diffusion MRI metrics in the reduced space (Titsias and Lawrence, 2004). For a comparison purpose, the GLM and a Principal Component Analysis (PCA) approach were also used, based on following equations:

$$Y = \text{Euclidean Mean} + \epsilon_{GLM}$$

$$Y = A(x) + \epsilon_{PCA}$$

$A(x)$  being a linear matrix in the case of Principal Component Analysis.

We have calculated the sum of squared residuals (SSR Archdeacon, 1994) as metric for comparison between the GLM, PCA and UMAP approaches. SSR is a statistical measure of the discrepancy between the data ( $Y_i$ ) and an estimation model, here defined:

$$SSR = \sum_i (Y_i - f(x_i))^2$$

where  $Y_i$  is the  $i$ th value of the variable to be predicted,  $x_i$  is the  $i$ th value of the explanatory variable, and  $f(x_i)$  is the predicted value of  $Y_i$ .

A small SSR indicates a tight fit of the model to the data. It is often used as an optimality criterion in parameter selection and model selection. A paired  $t$ -test was used to compare UMAP and PCA from one side, UMAP and the GLM from the other side

- (i) The back-projection step was done from the manifold atlas to the individual MRI space. As the Riemannian atlas has been built to capture the maximum variability, a new subject projected onto the learned manifold (for example the projection of one mTBI patient) will be synthesized using a local average to the closest individuals of the 20 healthy controls. For the estimation of the local average, we have used the Nadaraya-Watson kernel for high dimensional non-parametric regression (Conn and Li, 2017), here defined:

$$\hat{Y} = \frac{\sum_i K_h(x - x_i) y_i}{\sum_j K_h(x - x_j)}$$

where  $K_h$  represents a Gaussian kernel of bandwidth  $h$ . The training couples ( $x_i, y_i$ ) includes  $x$  as the quantity to regress and  $y$  as the prediction.

- (ii) We would like to emphasize that in our framework the residual  $\epsilon$  is calculated based on the difference between a synthetic bundle (we estimate the quantitative values coming from one brain bundle based on local mean values at a given location in the reduced subspace) and the true bundle quantitative values.

The selected healthy controls will be those who have closest quantitative values from the tested subject. Consequently, it is expected that  $f(x)$  will be generally less than the Euclidean mean, allowing  $\epsilon$  to be more representative of the pathological changes in the Riemannian framework. Indeed, a classical Euclidean  $\epsilon_{GLM}$  does represent both potential pathological changes and anatomical variability (i.e. the standard deviation of the mean) (Tilquin et al., 2019; Vik et al., 2007). The choice to work on a manifold (either Riemannian or using a more classical linear principal component analysis) is to shorten the standard deviation value so that  $\epsilon$  mainly represents the potential disease effect. A visual comparison of standard deviation between the GLM and *TractLearn* is provided in supplementary figure 1.

- (i) To ensure robustness in our dataset,  $\epsilon$  was estimated using a leave-one-out (LOO) strategy, by randomly selecting inter and intra-individual distances in the manifold from the test session. We consider that the residual  $\epsilon$  is representative of the abnormalities present in a new subject when it is greater than the model variability learned during the LOO on the healthy control group. A total of 1360 permutation tests were computed in the control group using the LOO. At this step, we have obtained  $\epsilon$  estimation in all controls allowing a comparison with the  $\epsilon$  estimated from the 5 mTBI patients. We make the assumption that  $\epsilon$  follows a multivariate Gaussian distribution with a standard deviation that varies across the voxels (Tilquin et al., 2019; Vik et al., 2007).
- (ii) We finally identified altered voxels in each fascicle in real space by applying a generalized Z-score, here defined in our Riemannian framework<sup>1</sup>:

$$z = \frac{Y - f(x)}{\sigma}$$

Here we assume that  $Y$  follows a Gaussian law as follow:  $N[f(x), \sigma^2]$ . This approach aims at generalizing the classic Z-score formula:

$$z = \frac{Y - \mu}{\sigma}$$

where  $\mu$  is the Euclidean mean of the population.

where  $f(x)$  is our learned healthy manifold,  $Y$  is the raw score (here the bundle quantitative values) to be converted into a standard score, and  $\sigma$  the standard deviation of the population.

- (i) A Bonferroni correction was performed for the testing of a mean of  $n$  voxels tests (dividing the p-value by a factor of  $n$ ). The number of  $n$  voxels corresponds to all the voxels which are contained in each *TractSeg* generated bundles (for example, a bundle of 20,000 voxels was analyzed by dividing the p-value by a factor of 20,000). A Bonferroni-corrected significance threshold of 0.05 was used. Furthermore, we have also generated a synthetic lesion on corpus callosum 3 to illustrate the difference between Z scores in our Riemannian frameworks and using the GLM (supplementary figure 2).

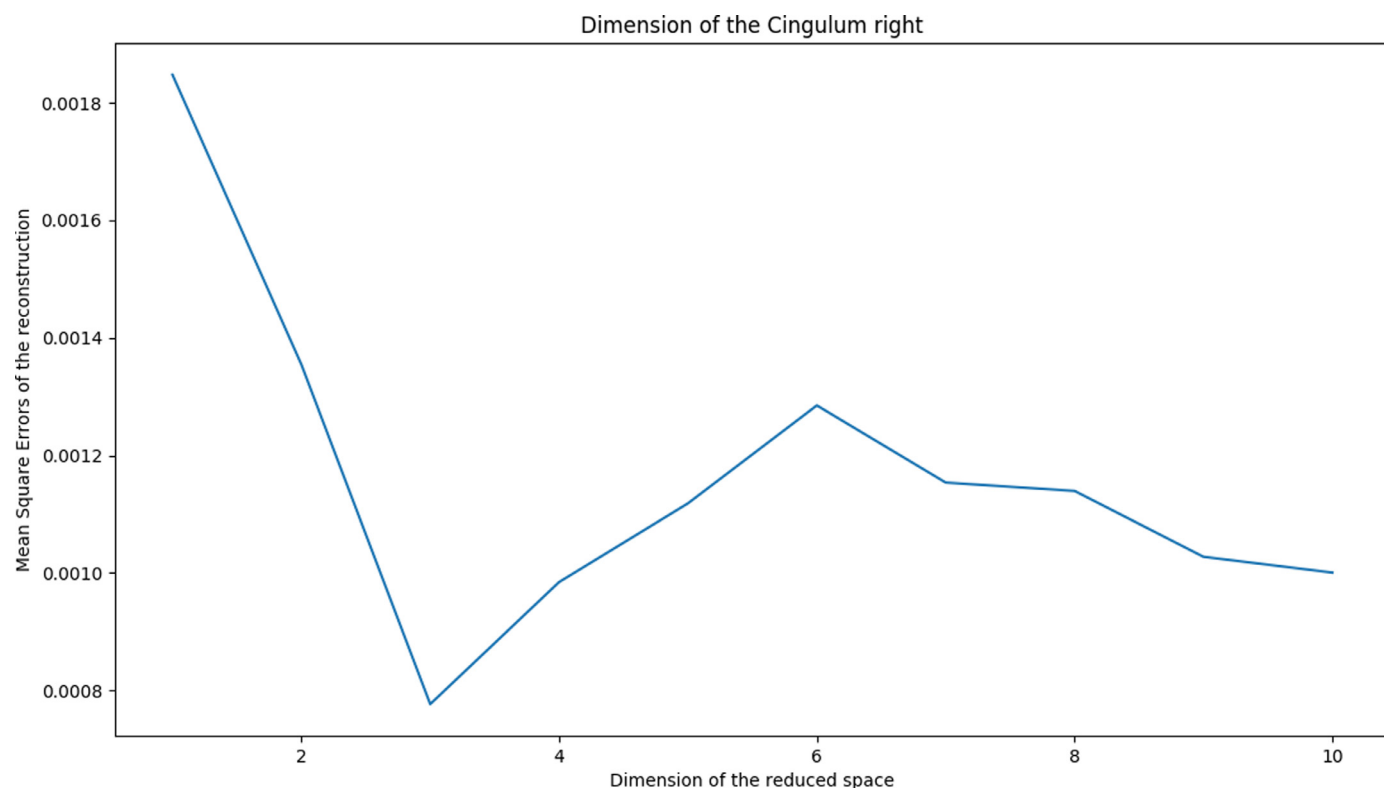
Fig. 2

Fig. 4

#### Test-Retest statistics

We have used three strategies to estimate the test-retest reliability of each voxel diffusion MRI metric in the *TractSeg* bundles.

Firstly, we have calculated intra-class correlation coefficients (ICC) for the different diffusion MRI metrics by comparing the mean values



**Fig. 2.** Example of the right cingulum, showing a minimum mean square error of the reconstructions for a dimension equal to 3. Each point of the curve represents a single execution of the *TractLearn* algorithm with a particular regularization: the x- and y-axes positions are determined by the dimension of the reduced space and mean square errors of the reconstruction, respectively.

of all voxels contained in one bundle between the test and the retest session. It is a classic approach in which a single-score ICC is based on a two-way model. A paired *t*-test was used to compare TW-FA and FA from one side, TW-FOD and AFD from the other side.

Secondly, we have compared the distance between the test and the retest session in the reduced space for each bundle with the distance between the controls and each patient, as illustrated Fig. 3.

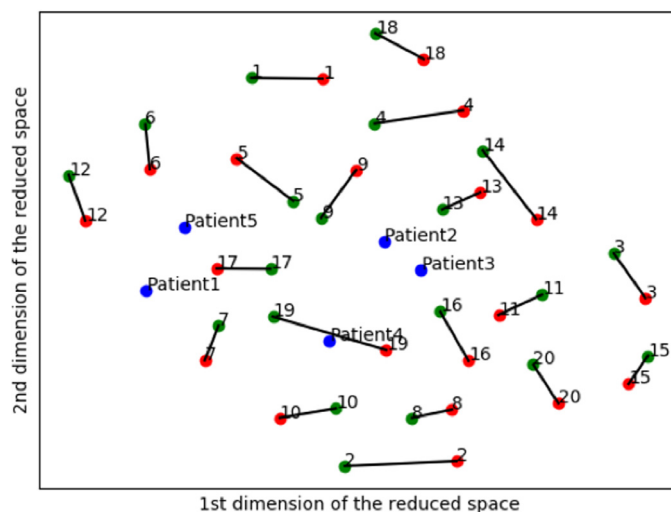
Finally, we have evaluated the number of injured bundles in the 5 mTBI patients depending on the choice of session for training the manifold space, i.e. when the test (or retest) session was used to compute the learned manifold of the healthy group. To allow loss functions comparisons (i.e. to statistically compare the inaccuracy of predictions in detecting the number of lesions between the test and retest session), we have considered lesions with at least 5 altered voxels per bundle.

## Results

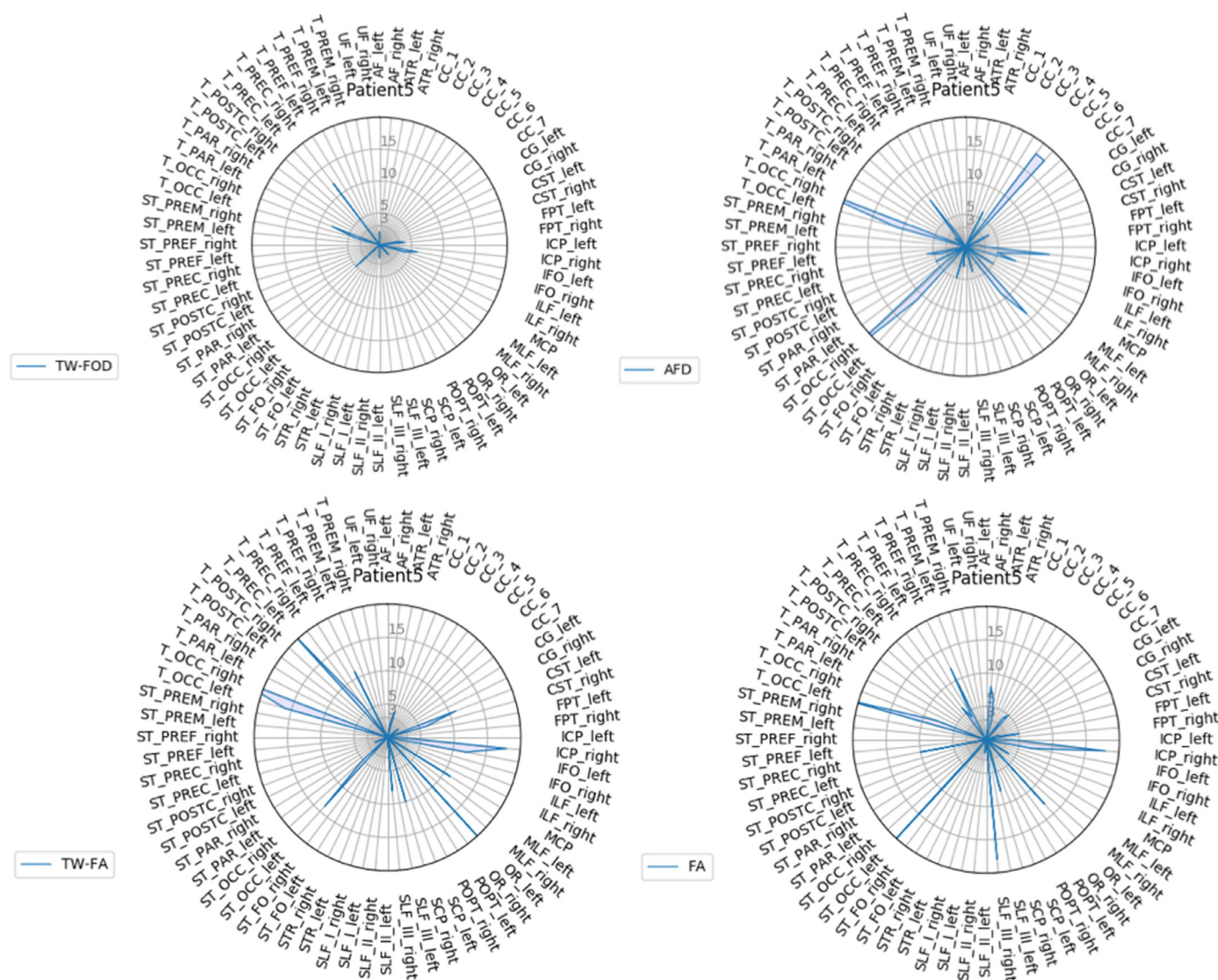
### *TractSeg and diffusion MRI metrics reliability*

All but two bundles (fornix and anterior commissure) of the 72 TractSeg brain bundles were successfully reconstructed in all subjects using TractSeg (Note however that this refers to the success of output from TractSeg and does not refer to the specific accuracy of each reconstruction, as no ground truth for those bundles is available). As the anterior commissure and the fornix were not robustly reconstructed in all subjects, they have been excluded from further quantitative analyses.

We have also chosen to study the corpus callosum through the seven subparts identified by *TractSeg* (rather than as a single anatomical structure), to help with the localization of potential lesions. Our analysis is therefore restricted to the remaining 68 bundles (all *TractSeg* bundles but left fornix, right fornix, anterior commissure, and non-segmented corpus callosum).



**Fig. 3.** 2D reduced space of left thalamo-prefrontal bundles. The units of each axes are usually arbitrary in dimensionality reduction. Here the TW-FOD metrics are provided as input for *TractLearn*. In this framework, a point represents a collection of TW-FOD values (one per voxel in a given bundle) in the reduced subspace. The farther two points are from each other, the more different the quantitative diffusion MRI metric values are. Healthy controls are numerated from 1 to 20, with green dots representing the test session, and red dots representing the retest session 12 months later; the mTBI patients are represented by the five blue dots. For most subjects, red and green dots are in close vicinity in the manifold, with the black lines representing the “MRI inter-session effect”. The patients bundle collection of diffusion MRI metrics here are located among those of the controls, suggesting that there is no or low global difference between controls and patients in this randomly selected bundle.



**Fig. 4.** Radar plots of *TractLearn* results for patient 5 showing the number of voxels presenting with significant altered Z scores in each bundle, as identified using *TractLearn* based on session 1.

The names and abbreviations of each brain bundle are as follows (Wasserthal et al., 2018b): Arcuate fascicle (AF), Anterior thalamic radiation (ATR), Corpus callosum (Rostrum (CC 1), Genu (CC 2), Rostral body (CC 3), Anterior midbody (CC 4), Posterior midbody (CC 5), Isthmus (CC 6), Splenium (CC 7)), Cingulum (CG), Corticospinal tract (CST), Middle longitudinal fascicle (MLF), Fronto-pontine tract (FPT), Inferior cerebellar peduncle (ICP), Inferior occipito-frontal fascicle (IFO), Inferior longitudinal fascicle (ILF), Middle cerebellar peduncle (MCP), Optic radiation (OR), Parieto-occipital pontine (POPT), Superior cerebellar peduncle (SCP), Superior longitudinal fascicle I (SLF I), Superior longitudinal fascicle II (SLF II), Superior longitudinal fascicle III (SLF III), Superior thalamic radiation (STR), Uncinate fascicle (UF), Thalamoprefrontal (T PREF), Thalamo-premotor (T PREM), Thalamo-precentral (T PREC), Thalamo-postcentral (T POSTC), Thalamo-parietal (T PAR), Thalamo-occipital (T OCC), Striato-fronto-orbital (ST FO), Strioprefrontal (ST PREF), Striato-premotor (ST PREM), Striato-precentral (ST PREC), Striato-postcentral (ST POSTC), Striato-parietal (ST PAR), Striato-occipital (ST OCC). Note: besides MCP and the 7 CC subdivisions, all other bundles were present in each hemisphere.

The MRI inter-session effect of the diffusion MRI metrics measured in the *TractSeg* bundles was firstly studied by the ICC coefficients. ICC

was on average 0.89, 0.88, 0.85 and 0.84 for TW-FA, FA, TW-FOD and bundle AFD, respectively; individual bundle ICC values are shown in [Table 1](#).

The t-tests, as paired per bundle, showed significantly higher ICC values for TW-FA in comparison to FA ( $p = 0.00014$ ) but not for TW-FOD as compared with Bundle AFD ( $p = 0.86$ ).

## TractLearn

The comparison of the GLM with PCA-based dimension reduction and UMAP nonlinear dimension reduction was done based on the SSR.

The mean SSR for all brain bundles was estimated as being 957.6 (SD 681.4) for the GLM; 101.4 (SD 69.8) for the PCA and for 74.5 (SD 57.3) UMAP. UMAP outperforms the PCA ( $p < 0.001$ ) and the GLM ( $p < 0.001$ ) for all brain bundles. As example of SSR differences, we have randomly selected a few bundles (Fig. 6).

We have firstly determined that the dimension with the minimum mean square error was equal to 3 for all 68 bundles. For visualization purpose, we have chosen to graphically present the reduced space as

**Table 1**  
ICC coefficients for all 4 diffusion MRI metrics mean value in each TractSeg bundle.

	TW-FA	FA	TW-FOD	Bundle AFD		TW-FA	FA	TW-FOD	Bundle AFD
AF_left	0,93	0,92	0,97	0,94	SLF_II_left	0,81	0,78	0,84	0,81
AF_right	0,95	0,94	0,96	0,94	SLF_II_right	0,9	0,87	0,92	0,89
ATR_left	0,9	0,9	0,82	0,84	SLF_I_left	0,8	0,78	0,7	0,77
ATR_right	0,89	0,9	0,78	0,78	SLF_I_right	0,87	0,85	0,8	0,57
CC_1	0,82	0,81	0,81	0,78	STR_left	0,91	0,91	0,73	0,88
CC_2	0,91	0,89	0,87	0,87	STR_right	0,91	0,9	0,82	0,87
CC_3	0,81	0,77	0,76	0,65	ST_FO_left	0,93	0,93	0,84	0,87
CC_4	0,82	0,81	0,83	0,79	ST_FO_right	0,8	0,78	0,71	0,52
CC_5	0,92	0,93	0,91	0,97	ST_OCC_left	0,95	0,94	0,96	0,92
CC_6	0,9	0,91	0,9	0,89	ST_OCC_right	0,91	0,91	0,97	0,94
CC_7	0,9	0,91	0,81	0,7	ST_PAR_left	0,91	0,9	0,92	0,92
CG_left	0,88	0,86	0,92	0,85	ST_PAR_right	0,89	0,88	0,95	0,94
CG_right	0,86	0,85	0,93	0,85	ST_POSTC_left	0,86	0,83	0,73	0,75
CST_left	0,96	0,96	0,92	0,92	ST_POSTC_right	0,9	0,89	0,85	0,81
CST_right	0,85	0,85	0,87	0,79	ST_PREC_left	0,94	0,94	0,89	0,93
FPT_left	0,93	0,92	0,9	0,88	ST_PREC_right	0,94	0,94	0,91	0,92
FPT_right	0,95	0,95	0,94	0,9	ST_PREF_left	0,79	0,74	0,66	0,55
ICP_left	0,79	0,79	0,67	0,66	ST_PREF_right	0,91	0,9	0,89	0,86
ICP_right	0,83	0,83	0,82	0,86	ST_PREM_left	0,84	0,71	0,77	0,64
IFO_left	0,94	0,94	0,94	0,94	ST_PREM_right	0,93	0,94	0,78	0,88
IFO_right	0,87	0,87	0,88	0,84	T_OCC_left	0,96	0,96	0,96	0,93
ILF_left	0,92	0,92	0,91	0,9	T_OCC_right	0,84	0,86	0,88	0,87
ILF_right	0,75	0,75	0,62	0,8	T_PAR_left	0,89	0,88	0,78	0,89
MCP	0,9	0,9	0,87	0,87	T_PAR_right	0,85	0,85	0,81	0,85
MLF_left	0,91	0,88	0,9	0,85	T_POSTC_left	0,91	0,89	0,75	0,88
MLF_right	0,95	0,95	0,86	0,93	T_POSTC_right	0,96	0,96	0,73	0,86
OR_left	0,97	0,97	0,96	0,93	T_PREC_left	0,91	0,91	0,67	0,85
OR_right	0,8	0,79	0,87	0,89	T_PREC_right	0,93	0,92	0,87	0,9
POPT_left	0,92	0,93	0,91	0,92	T_PREF_left	0,9	0,9	0,85	0,87
POPT_right	0,95	0,94	0,9	0,9	T_PREF_right	0,95	0,95	0,9	0,88
SCP_left	0,84	0,84	0,84	0,89	T_PREM_left	0,88	0,89	0,93	0,93
SCP_right	0,84	0,82	0,82	0,84	T_PREM_right	0,94	0,94	0,77	0,87
SLF_III_left	0,91	0,89	0,86	0,93	UF_left	0,84	0,8	0,9	0,88
SLF_III_right	0,89	0,86	0,89	0,89	UF_right	0,84	0,81	0,88	0,74

2-dimensional plots, though all the residual estimation was based on a 3-dimensional reduction of each brain bundle.

The inter-session effect was then directly studied in the manifold. Using *TractLearn*, the distance in the geodesic space for each bundle was significantly shorter between the test and the retest procedure (intra-subject) than between two different subjects ( $p < 0.001$ ) (e.g. see example in Fig. 3).

In all patients, the bundle AFD metric found the highest number of bundles with altered voxels. When taking into account the results from the manifold based on the test and the manifold based on the retest session, patient number 5 presented with the higher number of injured bundles (including corpus callosum parts 4 and 7, right thalamo-occipital bundle, right striato-occipital bundle, right optic radiation and middle cerebellar peduncle), with some of them including at least 20 altered voxels. Patient 2 showed no bundle with more than 10 altered voxels for any diffusion MRI metrics (Fig. 5) – see Supplementary Material for the corresponding results of the other three patients.

No morphological abnormalities were seen on morphological MRI sequences (see supplementary figure S6).

The back projections allow to identify the location of altered voxels within each bundle separately using the TW-FOD, TW-FA, bundle AFD and FA metrics, as illustrated in Fig. 6 and 7 for two of the patients.

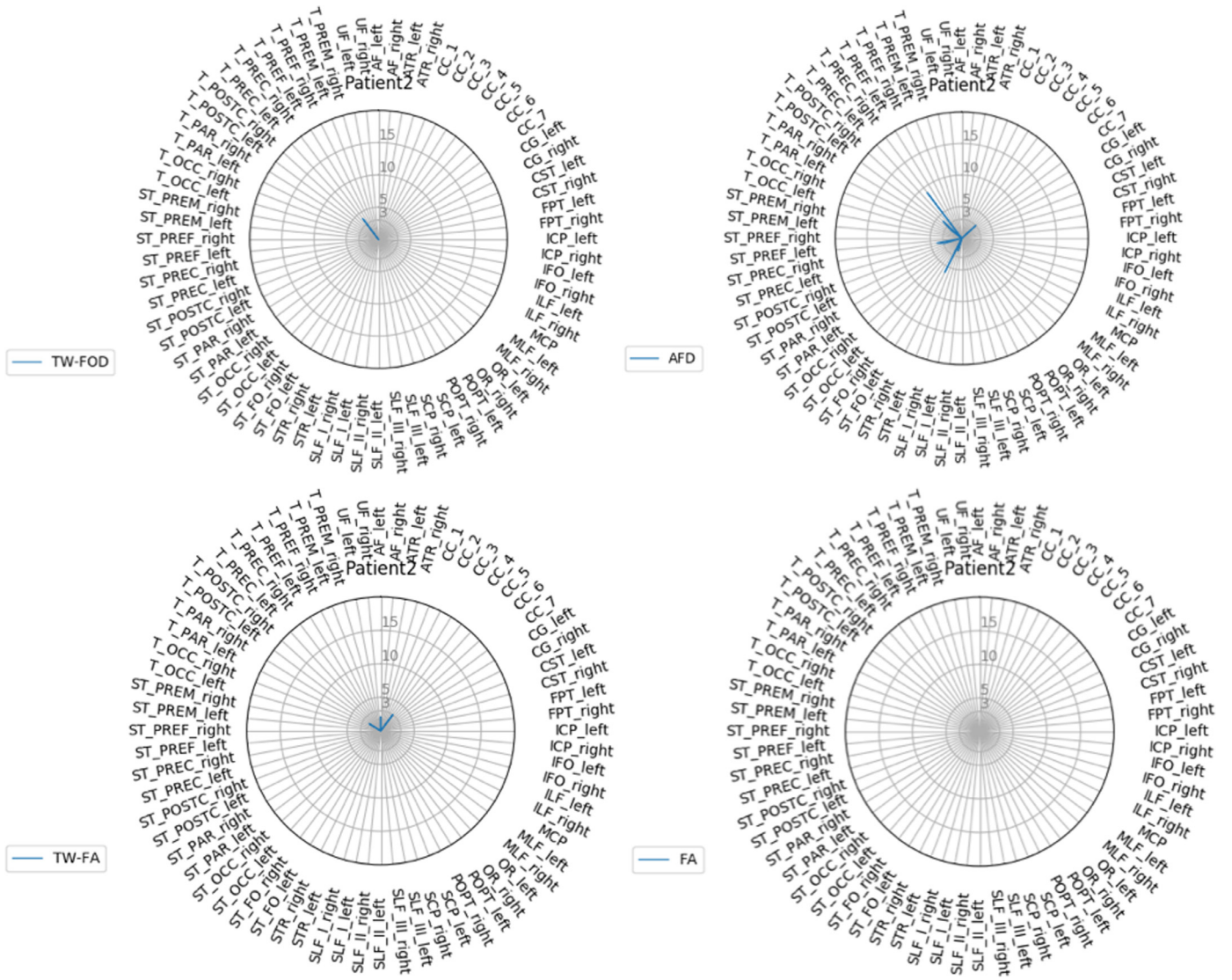
Using the Zero-one loss function for classification learning (i.e. to assign 0 to loss for a correct concordance between bundles with more than 5 altered voxels on the learned manifold built using the first MR session and bundles with more than 5 altered voxels on the learned manifold built using the second MR session: 0 is considered as a perfect score while 1 implies a total absence of concordance for all bundles), we have found scores as being 0.0029, 0.0088, 0.0147, 0.0029 for FA, TW-FA, AFD, TW-FOD, respectively (see Fig. 8 for an example).

## Discussion

We have proposed a novel unified framework for the quantitative analysis of properties of brain bundles by geodesic learning. *TractLearn*, a data-driven learning task, allows mapping between the high dimensional image domain and a reduced latent space of brain fascicles. In the patient group, *TractLearn* detected white matter abnormalities in a number of bundles; importantly, our results show that the identification of the abnormal bundles was reproducible, by comparing the number and location of abnormalities between a learned manifold built using the test session of controls and another learned manifold based on the retest session ~1 year later. The proposed framework provides a new means to quantitatively analyse the results from diffusion MRI fibre-tracking on a patient-specific basis.

The injuries location identified by *TractLearn* (mainly located in the frontal lobes, occipital lobes and in the corpus callosum) are compatible with the known pathophysiology of mTBI (Delouche et al., 2016; Stokum et al., 2015) and were not visible on conventional MR sequences. In addition, we proposed a statistical framework having the capabilities to detect a potential interaction between voxel alteration through their representation in the reduced manifold space. As the voxels are not analyzed independently, we can theoretically identify subtle diffusion MRI metrics variation in a given bundle. Finally, we proposed to use a method of back-projection from the highly sensitive low-dimensional manifold space to the MRI individual subject space, to visualize the physical location of the brain abnormalities.

For the examples shown in our study, we have used some selected illustrative diffusion MRI metrics (i.e. FA, bundle AFD, TW-FA and TW-FOD); it is however straightforward to extend the *TractLearn* framework to include other metrics, including exploiting the flexibility of the TWI



**Fig. 5.** Radar plots of *TractLearn* results for patient 2, showing a low number of voxels presenting with significant altered Z scores in each bundle, as identified using *TractLearn* based on the session 1.

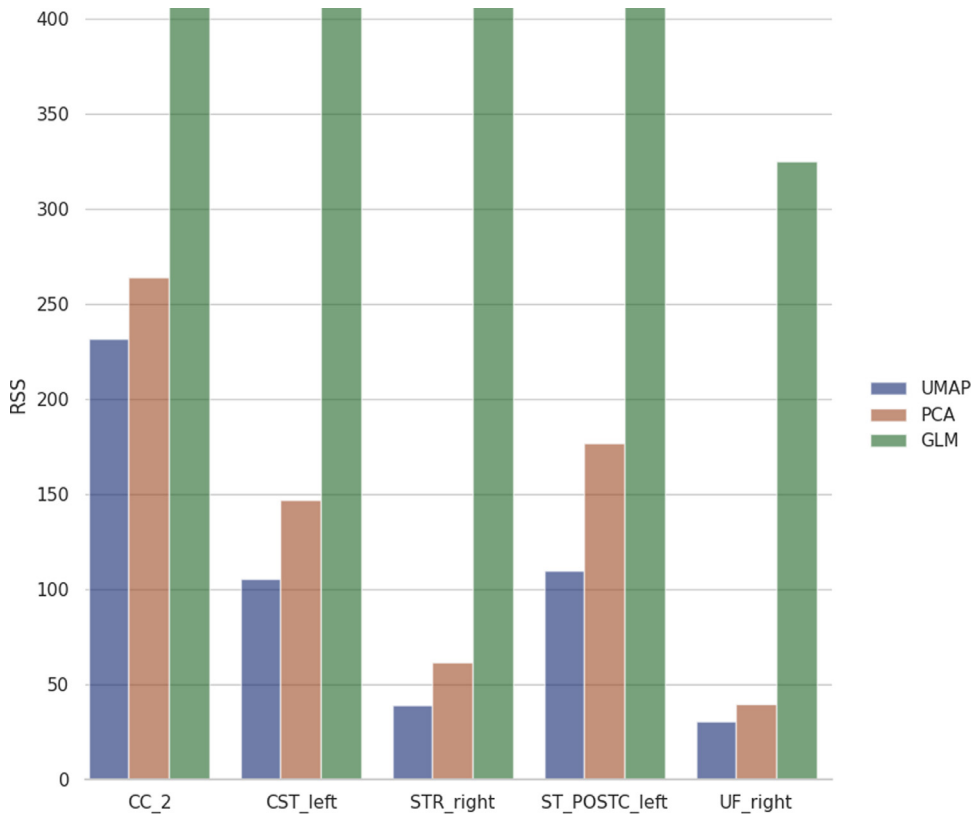
approach, which allows to combine a tractogram with any other (even non-MRI) co-registered metric (Calamante, 2017). *TractLearn* only needs a collection of quantitative diffusion MRI metrics for each bundle as input.

#### Benefits of manifold learning

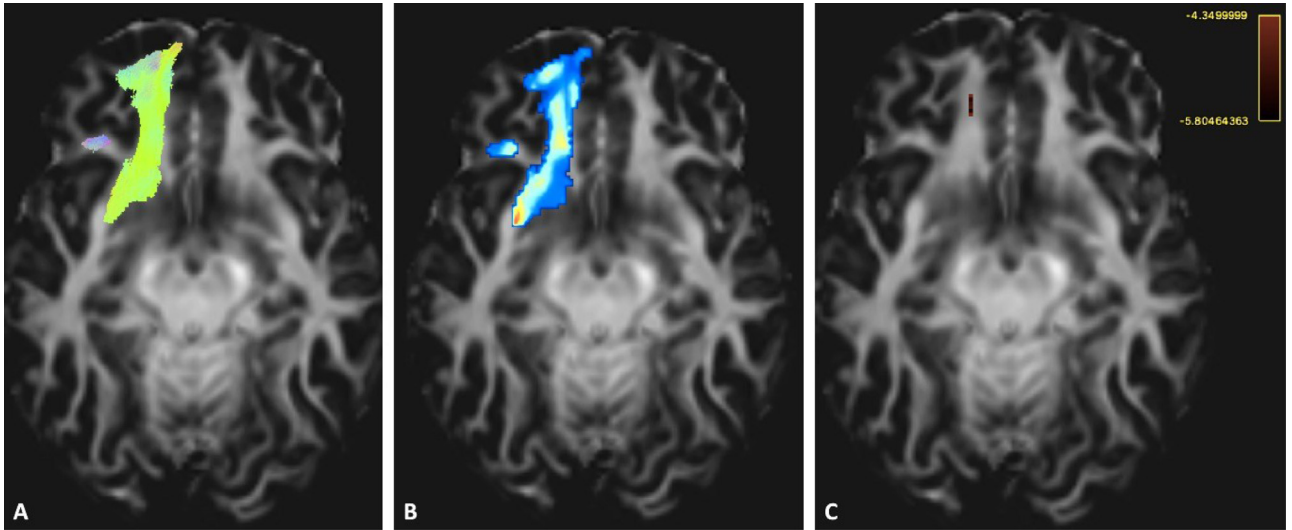
A common problem in DWI studies is the difficulty to detect abnormalities at the single subject level (e.g. for precision medicine), and instead having to rely more commonly on group-based studies. The possibility to detect subtle fiber architecture alterations at the voxel level in individuals through direct analysis of multiple MRI contrasts requires advanced tools with the ability to manage high-dimensional vector space. Typical clinical studies often include a limited number of patients; the natural mathematical space of the quantitative values at the fascicle level then needs to be constrained to control for the mismatch between the number of samples (here the number of controls/patients) and the number of features (here the number of voxels in each bundle). This problem, which occurs in the context of HDLSS data, is known as the curse of dimensionality, and we therefore need to reduce the space, for example to be able to apply Z-score algorithms on quantitative analysis.

Dimensionality reduction seeks to produce a low dimensional representation of high dimensional data that preserves the original structure. Dimensionality reduction algorithms offer a solution to important problems in data science. They allow visualization of complex data and they can be applied as potential pre-processing step for machine learning. They are classically divided into two categories: those that seek to preserve the original distance structure within the data, and those that favor the preservation of local distances over global distance. The second category of algorithms includes for example t-SNE, the state-of-the-art in dimensionality reduction for visualization (Maaten and Hinton, 2008). The advantage of *UMAP* over t-SNE is a better preservation of the original structure (here the diffusion MRI metrics coming from brain bundles) (McInnes et al., 2018) while being as powerful for visualization purpose.

We also proposed a generalization of the classical Z-score where  $f(x)$  corresponds to the mean value. Indeed, we have modeled the population by a regression function instead of a unique sample. In other words, all healthy controls do represent the population of reference instead of a unique individual. The first consequence is that each new studied subject in a previously learned manifold will be projected as close as possible to one of the control subjects (i.e. those with diffusion MRI metrics values closest to the new subject) (see supplementary figure S1). In the classi-



**Fig. 6.** SSR bar plots in 5 randomly selected brain bundles (Corpus Callosum part 2, left corticospinal tract, right superior thalamic radiation, left striato-postcentral and right uncinate fascicle). While all bundles showed a big difference between GLM SSR and PCA/UMAP SSR values; the differences between PCA and UMAP varies across bundles. Here for example the difference appeared more important for the left striato-postcentral bundle than for the right uncinate fascicle.



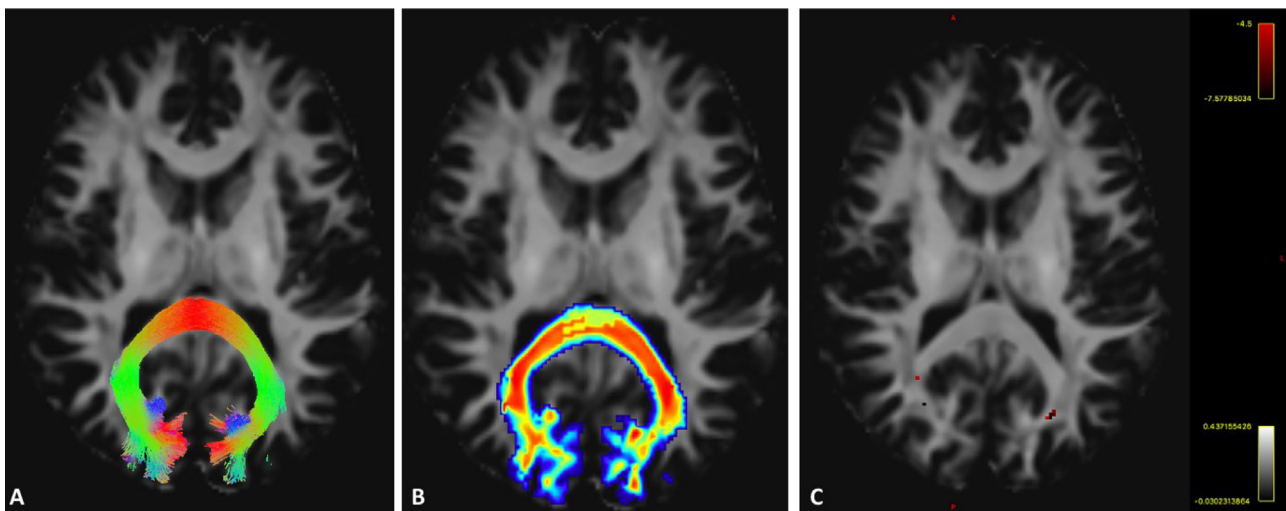
**Fig. 6.** Back projection of the manifold residual in the template space, to localize the brain abnormality in Patient 1. Here is the example of the first part of the Striato-fronto-orbital (ST-FO right) with the track file identified by TractSeg (A), the track-weighted maps based on the TW-FOD contrast (B) and the back projection of the Z Score (C). The Z-score threshold was set to  $-4.35$ ; corresponding to a Bonferroni correction for a bundle of 7444 voxels. It has allowed the identification of voxels with significantly different TWI contrast, in comparison with the control group (here retest session).

cal Euclidean framework, there is no local mean as proposed but global mean, thus all tested subjects will have the same reference. This automatically leads to a higher standard deviation in diffusion MRI metrics comparison, which in turn will hinder the possibility to detect quantitative abnormalities as the residual  $\varepsilon$  does contain the standard deviation added to the potential pathological effect.

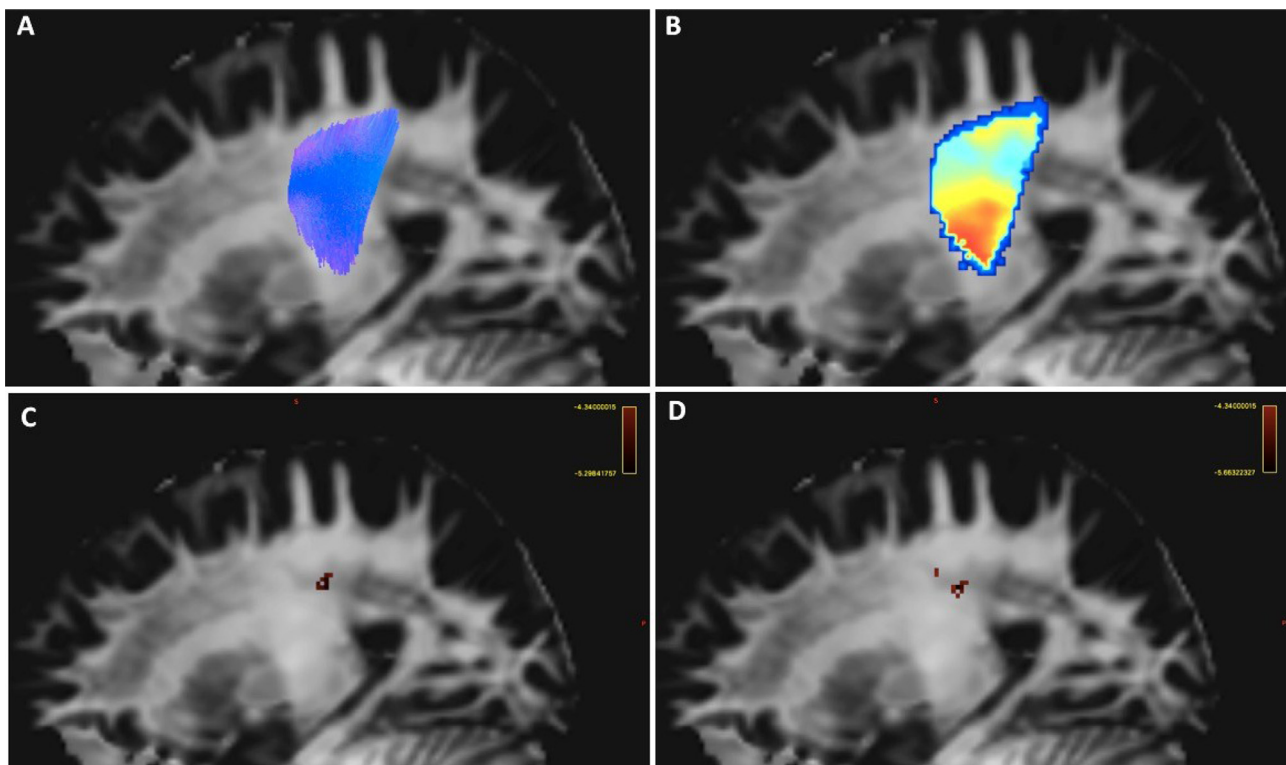
Tilquin et al. (2019) have already proposed to learn the manifold spanned by the normal controls using non-linear dimensionality reduction techniques. In their work, based on T1-weighted imaging, the image of a subject is projected on the control group manifold allowing a com-

parison of the reconstruction with the subject's original neuroimaging data. The objective of these projection techniques is to detect abnormal patterns by way of statistical tests on the residuals. In this prior work, the importance of non-linear modeling of the manifold in the reduced-dimension subspace was highlighted, as well as robustness to abnormalities detection on a larger dataset, yet without retest session to assess the reliability, as we were able to include in our study.

Manifold learning has also been previously used in DWI. For example, the estimation of the embedded reduced space for DTI-based tensors has been previously described using Manifold Learning (Khurd et al.,



**Fig. 7.** Another example from the corpus callosum part 7 (Patient 1) with the track file identified by TractSeg (A), the track-weighted maps based on the AFD contrast (B) and the back projection of the Z Score (C). The Z score has allowed the identification of voxels with significantly different AFD contrast ( $p < 0.05$  Bonferroni corrected for colored voxels), in comparison with the control group. The Z-score threshold was set to  $-4.73$ ; corresponding to a Bonferroni correction for a bundle of 21,303 voxels.



**Fig. 8.** Example of the patient 4 left Superior thalamic radiation (A) and the subsequent track-weighted fractional anisotropy contrast (TW-FA, B). The image C illustrates the location of a local association of altered voxels (more than 15 voxels as identified using the radar plots) based on the test session manifold. The image D illustrates the altered voxels (here represented as blue voxels) based on Z-score calculated from the manifold of the retest session. The Z-score threshold was set to  $-4.34$ ; corresponding to a Bonferroni correction for a bundle of 4100 voxels.

2007). More recently, manifold learning has also been used on DWI to map the white matter fiber in a controlled subspace with an adapted model of fingerprint called “fingerprint” and on multimodal MRI acquisition including DWI data (Kumar et al., 2018; Kumar et al., 2017); these studies have shown interesting relationship between compact fingerprint and genetic biomarkers on a large population data. The novelty of our study is to propose and assess a framework to identify altered voxels based on diffusion MRI metrics using a manifold Riemannian frame-

work and state-of-the-art tractography analysis methods. We propose to replace the GLM by a manifold reference for the healthy control group, improving diagnostic performance in terms of disease detection at the patient-specific level.

The possibility to back project the manifold information (here the Z-score testing) in the subject space help to precisely locate a disease process or pathophysiology in a patient-specific basis. Taken into account that the DWI data used here is consistent with a clinically suit-

able protocol (e.g. total acquisition time of DWI sequence <10 min), it can be anticipated that comparable level of sensitivity to detect brain alterations is possible with other similar acquisition protocols. It should be emphasized that patient-specific abnormalities were here detected either using the first learned manifold (built from the test session of the healthy controls) or using the second learned manifold (built from the retest session ~1 year later). The global effect on voxels diffusion MRI metrics owing to the acquisition condition (test or retest) can be visualized in the reduced subspace. Regardless of the learned manifold chosen, the resulting number and location of abnormalities showed good correlation using the zero-one-loss test. It should be noted that Cohen's Kappa should be avoided as performance measure due to the unbalanced nature of our dataset (Delgado and Tibau, 2019) (i.e. an important mismatch between the high number of bundles with no voxel abnormalities and the low number of altered bundles).

This suggests that, in our dataset, the local effect due to local axonal shear lesion or bleeding (associated with mTBI) is larger than the global effects related to the use of either the test or retest session to compute the learned manifold.

It could be speculated that this may also suggest that TractLearn might be more lenient to be able to detect local (disease related) effects over global (acquisition related) effects when using data from various acquisition or scanning protocols. This remains to be tested in future work, for example using datasets coming from multi-center studies.

#### *Riemannian versus Euclidean statistical analysis*

The GLM, based on a Euclidean framework, has been widely used in the medical and neurosciences communities, and has been implemented in popular software tools, such as Statistical Parametric Mapping (SPM, <https://www.fil.ion.ucl.ac.uk/spm/>). While being effective in the case of low-dimensional data comparison for group study, its main limitation relies on its inability to properly model the data distribution without using a smoothing step. The smoothing of Diffusion-weighted data usually decreases the standard deviation of all voxels metrics coming from the GLM, yet here we have shown that it leads to standard deviation much higher than in our framework (as illustrated in the supplemental material). It could be argued that smoothing still occurs within the TWI calculation; this however can be seen as a 'smart-type' (streamline-informed) smoothing (Calamante, 2017).

Imposing the Euclidean topology led to poor estimation, as demonstrated with the mean SSR difference between UMAP and the GLM. In addition, the non-linear topology of the brain bundles was also beneficial, based on the difference between PCA and UMAP mean SSR values, with residuals lower in the latter model when making the difference between the synthetic subject reconstructed from the manifold subspace and the real values coming from DWI acquisition. We have then used UMAP, which relies on non-Gaussian and robust modeling, to capture the variability of the brain bundles diffusion MRI metrics.

Such inabilities to model quantitative data relying on complex topology have been also demonstrated using dimensionality reduction by Tenenbaum et al., 2000), but also for comparing T1-weighted imaging anatomy (Miller, 2004) or high angular Diffusion-weighted data (Goh et al., 2011). In addition, using Riemannian distances allowed us to detect joint statistical variations in a group of voxels.

It should be noted that, even including a low sample data size for controls - with its inherent limitations to have potential anatomical variations displayed in the manifold, we were able to detect local bundle alteration in individual patients with mTBI. Riemannian geometry has already been used in various applications linked to diffusion MRI, such as proposing new nonlinear metrics (Batchelor et al., 2003; Fuster et al., 2014), identifying bundle shape variability (Sun et al., 2017) or improving segmentation of white matter bundles (Hao et al., 2014). To the best of our knowledge TractLearn is the first toolbox that can be employed on usual DWI MRI metrics for individual's abnormality identification exploiting the benefits of a manifold analysis.

#### *The benefit of using twi as a quantitative biomarker*

TWI has proved to be powerful in mapping fascicles alterations in neurodegenerative diseases (Bozzali et al., 2011; Ziegler et al., 2014) or mTBI preclinical imaging (Tan et al., 2016), among others. The TWI contrast intrinsically contains the anatomical information of the fiber architecture (through the track-weighted averaging step), while also allowing a way to combine it with other local quantitative information (e.g. FA and FOD, in the current study). Numerous studies have proven the benefits of extracting FOD-related parameters for group studies, such as for example in the case of AFD in motor neuron disease (Raffelt et al., 2012b) or, more recently, Alzheimer's disease (Mito et al., 2018). FA analyses, based on the tensor model, have been widely done, and they have allowed us to increase our understanding of numerous brain diseases, including traumatic disorders (Ilvesmäki et al., 2014) (Aoki et al., 2012), neurocognitive diseases (Bozzali et al., 2011), and in optic pathways studies (Bender et al., 2014; Mandelstam, 2012). These were some of the reasons for our choice of possible parameters to illustrate TractLearn in the current study.

Here, more brain bundles alterations were detected using the bundle AFD metric.

Our work is, however, limited by the inability to validate our findings with pathological analyses, thus making difficult to compare the real added value of TractLearn over a more classical Euclidean approach. Future work is needed to address the sensitivity and specificity of the abnormalities detected by TractLearn.

As emphasized above, the track-weighted contrast uses a 'smart' smoothing effect (of the FOD amplitudes or FA values, in the current study) along the streamline. This work complements that of Willats et al. (2014), which has demonstrated that the track-weighted contrast had good reproducibility in a sample of 8 subjects. Here we find high ICC values in 68 different brain regions of very different sizes in the context of a relatively large test-retest protocol (one year between sessions, instead of <2 weeks in the work of Willats et al.). We also note that TW-FA was shown to be more reproducible than FA using a paired *t*-test on each bundle, consistent with the work of Willats et al. In contrast, we did not find a significant difference between TW-FOD and AFD ICC values. It should be noted however, that TW-FOD is not constructed based on the bundle AFD map and, therefore, their relationship is not as direct as the FA case (cf. TW-FA is constructed based on the FA map). Furthermore, bundle AFD mapping does also include some smoothing along the bundle (related to the *afdconnectivity* command in MRtrix), and thus the benefits of increased reproducibility with the track-weighted approach in TWI is limited when compared with bundle AFD.

Interestingly, the Zero-one loss function results indicated that FA had higher concordance of detecting an abnormality than TW-FA. This might relate to the very focal abnormalities in the mTBI patients included in our small cohort, as then the streamline-specific smoothing involved with TW-FA could dilute the effect of the abnormality rather than amplify it (Calamante, 2017). Future work with a larger range of patient severities is required to fully characterize the benefits of TWI vs using DTI metrics directly.

#### *Spatial coregistration*

Most group DWI neuroimaging studies are registered to a common template. Here we have used a custom-made FOD template for robust coregistration of each individual FOD maps to the template, and thus obtain accurate matching between white matter structures (Raffelt et al., 2011); the resulting spatial transformation was applied to the TractSeg tractograms for each subject, to ensure correspondence between the identified bundles. This also ensures that the FODs and tracks lie on the same space, in order to correctly compute TW-FOD values (Calamante, 2017). However, we have here limited the anatomical variability of cortical termination by keeping only the top 80% voxels values of each bundle, based on the track-density contrast. Indeed, while

the coregistration based on FOD symmetric diffeomorphic has allow to match major brain bundles, we have noticed that cortical variability made more difficult a perfect matching for the entire bundle. Absence of this step could potentially lead to false positive lesions on the bundles boundaries. In addition, as TractSeg tends to produce bundle overlaying (i.e. some boundaries voxels can be linked to two different bundles), using a thresholding has allowed to precisely locate abnormalities.

### Future perspectives

In medicine, researchers often raise a prior anatomical hypothesis on potential diseased brain bundles from which the quantitative analysis would allow to early detect the prodromal stages. Here we propose not to pass through this step to directly study a large part of the brain fascicle in a semi-supervised way (as the status of each point in the manifold is known before the calculation of the residual, i.e. whether the subject is a healthy control or a patient).

While *TractLearn* was here used for precision medicine (i.e. one vs. a healthy group), in general, it could be also applied on the more classical approach of group comparison studies, such as by replacing Z-score back projection with p-values back projection. The objective would be then to detect brain bundles alteration at the group level taking advantage on the Riemannian framework. With the appropriate statistical models in hand, we may also regress the manifold data directly against one or more independent variables, for example clinical, biological or pathological physiological data.

While it was very reassuring that *TractLearn* provided stable results even with a relatively small healthy cohort (i.e. we demonstrate good test-retest with only 20 healthy subjects to define the healthy manifold), it remains to be investigated how reliability is impacted by factors such as type of disease, location, extent, severity, as well as data quality. Future studies evaluating these factors are guaranteed, which should help further characterize the performance of *TractLearn*.

Finally, while here we have used one scalar value per voxel for each manifold (e.g. a manifold for TW-FOD and a different one for TW-FA, etc.), it might be the case that some pathologies will not be able to be characterized just with one parameter. Similarly, it might be the case that higher order dimensions are needed to represent certain DWI features (e.g. using a manifold of the spherical harmonics amplitudes of the FODs rather than just the TW-FOD). The manifold framework should allow to decompose the DW signal by providing the same algorithm a collection of MRI metrics per voxel simultaneously.

### Conclusion

We have presented *TractLearn*, a unified framework for brain fascicles quantitative analyses by geodesic learning. The possibility to detect abnormalities in individuals in the context of high-dimensional low sample size data hold promise for precision medicine.

### Data and code availability statements

The code for *TractLearn* is available on <https://github.com/geodaistics/TractLearn>.

The code for *UMAP* (McInnes et al., 2018) is also freely available on <https://github.com/lmcinnes/umap> and the TractSeg pre-trained DWI algorithm, openly available at <https://github.com/MIC-DKFZ/TractSeg> (Wasserthal et al., 2018a).

### Declaration of Competing Interest

None for the present work.

### Acknowledgments

The computation server used benefitted from the ERATRANIRMA project, by Agence Nationale de la Recherche [grant ANR-12-EMMA-

0056]. We would like to acknowledge the funding from the Alain Rahmouni grants of the French Society of Radiology. This work was also supported by funding from the National Health and Medical Research Council of Australia and the Australian Research Council.

### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2021.117927](https://doi.org/10.1016/j.neuroimage.2021.117927).

### References

- Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* 20, 870–888. doi:[10.1016/S1053-8119\(03\)00336-7](https://doi.org/10.1016/S1053-8119(03)00336-7).
- Aoki, Y., Inokuchi, R., Gunshin, M., Yahagi, N., Suwa, H., 2012. Diffusion tensor imaging studies of mild traumatic brain injury: a meta-analysis. *J. Neurol. Neurosurg. Psychiatry* 83, 870–876. doi:[10.1136/jnnp-2012-302742](https://doi.org/10.1136/jnnp-2012-302742).
- Archdeacon, T.J., 1994. *Correlation and Regression Analysis: A Historian's Guide*. University of Wisconsin Press.
- Barajas, R.F., Hess, C.P., Phillips, J.J., Von Morze, C.J., Yu, J.P., Chang, S.M., Nelson, S.J., McDermott, M.W., Berger, M.S., Cha, S., 2013. Super-resolution track density imaging of glioblastoma: histopathologic correlation. *AJNR Am. J. Neuroradiol.* 34, 1319–1325. doi:[10.3174/ajnr.A3400](https://doi.org/10.3174/ajnr.A3400).
- Batchelor, P.G., Atkinson, D., Hill, D.L.G., Calamante, F., Connelly, A., 2003. Anisotropic noise propagation in diffusion tensor MRI sampling schemes. *Magn. Reson. Med.* 49, 1143–1151. doi:[10.1002/mrm.10491](https://doi.org/10.1002/mrm.10491).
- Bender, B., Heine, C., Danz, S., Bischof, F., Reimann, K., Bender, M., Nägele, T., Ernmann, U., Korn, A., 2014. Diffusion restriction of the optic nerve in patients with acute visual deficit. *J. Magn. Reson. Imaging JMRI* 40, 334–340. doi:[10.1002/jmri.24367](https://doi.org/10.1002/jmri.24367).
- Bozzali, M., Parker, G.J.M., Serra, L., Embleton, K., Gili, T., Perri, R., Calta-girone, C., Cercignani, M., 2011. Anatomical connectivity mapping: a new tool to assess brain disconnection in Alzheimer's disease. *Neuroimage* 54, 2045–2051. doi:[10.1016/j.neuroimage.2010.08.069](https://doi.org/10.1016/j.neuroimage.2010.08.069).
- Calamante, F., 2017. Track-weighted Imaging methods: Extracting Information from a Streamlines Tractogram. *Magma N. Y. N* doi:[10.1007/s10334-017-0608-1](https://doi.org/10.1007/s10334-017-0608-1).
- Calamante, F., 2016. Super-resolution track density imaging: anatomic detail versus quantification. *AJNR Am. J. Neuroradiol.* doi:[10.3174/ajnr.A4721](https://doi.org/10.3174/ajnr.A4721).
- Calamante, F., Smith, R.E., Tournier, J.-D., Raffelt, D., Connelly, A., 2015. Quantification of voxel-wise total fibre density: investigating the problems associated with track-count mapping. *Neuroimage* 117, 284–293. doi:[10.1016/j.neuroimage.2015.05.070](https://doi.org/10.1016/j.neuroimage.2015.05.070).
- Calamante, F., Tournier, J.-D., Jackson, G.D., Connelly, A., 2010. Track-density imaging (TDI): super-resolution white matter imaging using whole-brain track-density mapping. *Neuroimage* 53, 1233–1243. doi:[10.1016/j.neuroimage.2010.07.024](https://doi.org/10.1016/j.neuroimage.2010.07.024).
- Calamante, F., Tournier, J.-D., Smith, R.E., Connelly, A., 2012. A generalised framework for super-resolution track-weighted imaging. *Neuroimage* 59, 2494–2503. doi:[10.1016/j.neuroimage.2011.08.099](https://doi.org/10.1016/j.neuroimage.2011.08.099).
- Delgado, R., Tibau, X.-A., 2019. Why Cohen's Kappa should be avoided as performance measure in classification. *PLoS One* 14, e0222916. doi:[10.1371/journal.pone.0222916](https://doi.org/10.1371/journal.pone.0222916).
- Delouche, A., Attyé, A., Heck, O., Grand, S., Kastler, A., Lamalle, L., Renard, F., Krainik, A., 2016. Diffusion MRI: pitfalls, literature review and future directions of research in mild traumatic brain injury. *Eur. J. Radiol.* 85, 25–30. doi:[10.1016/j.ejrad.2015.11.004](https://doi.org/10.1016/j.ejrad.2015.11.004).
- Fuster, A., Tristán-Vega, A., Haije, T.C.J.D., Westin, C.-F., Florack, L.M.J., 2014. A novel Riemannian metric for geodesic tractography in DTI. In: *Proceedings of the Computational Diffusion MRI and Brain Connectivity (MICCAI Workshop, Nagoya, Japan, September 22, 2013)*. Presented at the MICCAI 2013 Workshop on Computational Diffusion MRI (CDMRI '13), September 22, 2013, Nagoya, Japan, Springer, pp. 97–104. doi:[10.1007/978-3-319-02475-2\\_9](https://doi.org/10.1007/978-3-319-02475-2_9).
- Goh, A., Lenglet, C., Thompson, P.M., Vidal, R., 2011. A nonparametric Riemannian framework for processing high angular resolution diffusion images and its applications to ODF-based morphometry. *Neuroimage* 56, 1181–1201. doi:[10.1016/j.neuroimage.2011.01.053](https://doi.org/10.1016/j.neuroimage.2011.01.053).
- Hao, X., Zygmunt, K., Whitaker, R.T., Fletcher, P.T., 2014. Improved segmentation of white matter tracts with adaptive Riemannian metrics. *Med. Image Anal.* 18, 161–175. doi:[10.1016/j.media.2013.10.007](https://doi.org/10.1016/j.media.2013.10.007).
- Hauberg, S., Schöber, M., Liptrot, M., Hennig, P., Feragen, A., 2015. A Random Riemannian Metric for Probabilistic Shortest-Path Tractography. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 597–604. doi:[10.1007/978-3-319-24553-9\\_73](https://doi.org/10.1007/978-3-319-24553-9_73).
- Ilvesmäki, T., Luoto, T.M., Hakulinen, U., Brander, A., Ryymin, P., Eskola, H., Iversen, G.L., Ohman, J., 2014. Acute mild traumatic brain injury is not associated with white matter change on diffusion tensor imaging. *Brain J. Neurol.* 137, 1876–1882. doi:[10.1093/brain/awu095](https://doi.org/10.1093/brain/awu095).
- Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *Neuroimage* 103, 411–426. doi:[10.1016/j.neuroimage.2014.07.061](https://doi.org/10.1016/j.neuroimage.2014.07.061).
- Khurd, P., Verma, R., Davatzikos, C., 2007. Kernel-based manifold learning for statistical analysis of diffusion tensor images. *Inf. Process. Med. Imaging Proc. Conf.* 20, 581–593.

- Kumar, K., Desrosiers, C., Siddiqi, K., Colliot, O., Toews, M., 2017. Fiberprint: a subject fingerprint based on sparse code pooling for white matter fiber analysis. *Neuroimage* 158, 242–259. doi:[10.1016/j.neuroimage.2017.06.083](https://doi.org/10.1016/j.neuroimage.2017.06.083).
- Kumar, K., Toews, M., Chauvin, L., Colliot, O., Desrosiers, C., 2018. Multi-modal brain fingerprinting: a manifold approximation based framework. *Neuroimage* 183, 212–226. doi:[10.1016/j.neuroimage.2018.08.006](https://doi.org/10.1016/j.neuroimage.2018.08.006).
- Lyksborg, M., Siebner, H.R., Sørensen, P.S., Blinkenberg, M., Parker, G.J.M., Dagonowski, A.-M., Garde, E., Larsen, R., Dyrby, T.B., 2014. Secondary progressive and relapsing remitting multiple sclerosis leads to motor-related decreased anatomical connectivity. *PLoS One* 9, e95540. doi:[10.1371/journal.pone.0095540](https://doi.org/10.1371/journal.pone.0095540).
- Maaten, L.van der, Hinton, G., 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Mandelstam, S.A., 2012. Challenges of the anatomy and diffusion tensor tractography of the Meyer loop. *AJNR Am. J. Neuroradiol.* 33, 1204–1210. doi:[10.3174/ajnr.A2652](https://doi.org/10.3174/ajnr.A2652).
- McInnes, L., Healy, J., Melville, J., 2018. UMAP: uniform manifold approximation and projection for dimension reduction. *ArXiv180203426 Cs Stat.*
- Miller, M.I., 2004. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *Neuroimage* 23 (1), S19–S33. doi:[10.1016/j.neuroimage.2004.07.021](https://doi.org/10.1016/j.neuroimage.2004.07.021), Suppl.
- Mito, R., Raffelt, D., Dhollander, T., Vaughan, D.N., Tournier, J.-D., Salvado, O., Brodtmann, A., Rowe, C.C., Villemagne, V.L., Connelly, A., 2018. Fibre-specific white matter reductions in Alzheimer's disease and mild cognitive impairment. *Brain J. Neurol.* doi:[10.1093/brain/awx355](https://doi.org/10.1093/brain/awx355).
- Nguyen, L.H., Holmes, S., 2019. Ten quick tips for effective dimensionality reduction. *PLOS Comput. Biol.* 15, e1006907. doi:[10.1371/journal.pcbi.1006907](https://doi.org/10.1371/journal.pcbi.1006907).
- Pannek, K., Mathias, J.L., Bigler, E.D., Brown, G., Taylor, J.D., Rose, S.E., 2011. The average pathlength map: a diffusion MRI tractography-derived index for studying brain pathology. *Neuroimage* 55, 133–141. doi:[10.1016/j.neuroimage.2010.12.010](https://doi.org/10.1016/j.neuroimage.2010.12.010).
- Raffelt, D., Tournier, J.-D., Crozier, S., Connelly, A., Salvado, O., 2012a. Reorientation of fiber orientation distributions using apodized point spread functions. *Magn. Reson. Med.* 67, 844–855. doi:[10.1002/mrm.23058](https://doi.org/10.1002/mrm.23058).
- Raffelt, D., Tournier, J.-D., Fripp, J., Crozier, S., Connelly, A., Salvado, O., 2011. Symmetric diffeomorphic registration of fibre orientation distributions. *Neuroimage* 56, 1171–1180. doi:[10.1016/j.neuroimage.2011.02.014](https://doi.org/10.1016/j.neuroimage.2011.02.014).
- Raffelt, D., Tournier, J.-D., Rose, S., Ridgway, G.R., Henderson, R., Crozier, S., Salvado, O., Connelly, A., 2012b. Apparent Fibre Density: a novel measure for the analysis of diffusion-weighted magnetic resonance images. *Neuroimage* 59, 3976–3994. doi:[10.1016/j.neuroimage.2011.10.045](https://doi.org/10.1016/j.neuroimage.2011.10.045).
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional Networks for Biomedical Image Segmentation. *ArXiv150504597 Cs.*
- Stokum, J.A., Sours, C., Zhuo, J., Kane, R., Shanmuganathan, K., Gullapalli, R.P., 2015. A longitudinal evaluation of diffusion kurtosis imaging in patients with mild traumatic brain injury. *Brain Inj* 29, 47–57. doi:[10.3109/02699052.2014.947628](https://doi.org/10.3109/02699052.2014.947628).
- Sun, Z.Y., Houenou, J., Duclap, D., Sarrazin, S., Linke, J., Daban, C., Hamdani, N., d'Albis, M.-A., Le Corvoisier, P., Guevara, P., Delavest, M., Bellivier, F., Almeida, J., Versace, A., Poupon, C., Leboyer, M., Phillips, M., Wessa, M., Mangin, J.-F., 2017. Shape analysis of the cingulum, uncinate and arcuate fasciculi in patients with bipolar disorder. *J. Psychiatry Neurosci.* JPN 42, 27–36. doi:[10.1503/jpn.150291](https://doi.org/10.1503/jpn.150291).
- Tan, X.L., Wright, D.K., Liu, S., Hovens, C., O'Brien, T.J., Shultz, S.R., 2016. Sodium selenate, a protein phosphatase 2A activator, mitigates hyperphosphorylated tau and improves repeated mild traumatic brain injury outcomes. *Neuropharmacology* 108, 382–393. doi:[10.1016/j.neuropharm.2016.05.001](https://doi.org/10.1016/j.neuropharm.2016.05.001).
- Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323. doi:[10.1126/science.290.5500.2319](https://doi.org/10.1126/science.290.5500.2319).
- Tilquin, F., Faisan, S., Heitz, F., Noblet, V., Blanc, F., Namer, I., 2019. Anomaly Detection in Single Subject vs Group Using Manifold Learning. In: *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, ICASSP 2019, pp. 2867–2871. doi:[10.1109/ICASSP.2019.8682472](https://doi.org/10.1109/ICASSP.2019.8682472).
- Titsias, M.K., Lawrence, N.D., 2004. Gaussian process latent variable models for visualisation of high dimensional data. In: *Proceedings of the 16th International Conference on Advances in Neural Information Processing Systems, NIPS'03*.
- Tournier, J.-D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* 35, 1459–1472. doi:[10.1016/j.neuroimage.2007.02.016](https://doi.org/10.1016/j.neuroimage.2007.02.016).
- Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., Connelly, A., 2019. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* 202, 116137. doi:[10.1016/j.neuroimage.2019.116137](https://doi.org/10.1016/j.neuroimage.2019.116137).
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29, 1310–1320. doi:[10.1109/TMI.2010.2046908](https://doi.org/10.1109/TMI.2010.2046908).
- Veraart, J., Novikov, D.S., Christiaens, D., Ades-Aron, B., Sijbers, J., Fieremans, E., 2016. Denoising of diffusion MRI using random matrix theory. *Neuroimage* 142, 394–406. doi:[10.1016/j.neuroimage.2016.08.016](https://doi.org/10.1016/j.neuroimage.2016.08.016).
- Vik, T., Heitz, F., Charbonnier, P., 2007. Robust pose estimation and recognition using non-Gaussian modeling of appearance subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 901–905. doi:[10.1109/TPAMI.2007.1028](https://doi.org/10.1109/TPAMI.2007.1028).
- Wasserthal, J., Neher, P., Maier-Hein, K.H., 2018a. TractSeg - Fast and accurate white matter tract segmentation. *Neuroimage* 183, 239–253. doi:[10.1016/j.neuroimage.2018.07.070](https://doi.org/10.1016/j.neuroimage.2018.07.070).
- Wasserthal, J., Neher, P., Maier-Hein, K.H., 2018b. TractSeg - Fast and accurate white matter tract segmentation. *Neuroimage* 183, 239–253. doi:[10.1016/j.neuroimage.2018.07.070](https://doi.org/10.1016/j.neuroimage.2018.07.070).
- Wasserthal, J., Neher, P.F., Hirjak, D., Maier-Hein, K.H., 2019. Combined tract segmentation and orientation mapping for bundle-specific tractography. *Med. Image Anal.* 58, 101559. doi:[10.1016/j.media.2019.101559](https://doi.org/10.1016/j.media.2019.101559).
- Willats, L., Raffelt, D., Smith, R.E., Tournier, J.-D., Connelly, A., Calamante, F., 2014. Quantification of track-weighted imaging (TWI): characterisation of within-subject reproducibility and between-subject variability. *Neuroimage* 87, 18–31. doi:[10.1016/j.neuroimage.2013.11.016](https://doi.org/10.1016/j.neuroimage.2013.11.016).
- Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., 2018. Image reconstruction by domain-transform manifold learning. *Nature* 555, 487–492. doi:[10.1038/nature25988](https://doi.org/10.1038/nature25988).
- Ziegler, E., Rouillard, M., André, E., Coolen, T., Stender, J., Balteau, E., Phillips, C., Garraux, G., 2014. Mapping track density changes in nigrostriatal and extranigral pathways in Parkinson's disease. *Neuroimage* 99, 498–508. doi:[10.1016/j.neuroimage.2014.06.033](https://doi.org/10.1016/j.neuroimage.2014.06.033).