



# Anchor-Constrained Plausibility (ACP): A Novel Concept for Assessing Tractography and Reducing False-Positives

Peter F. Neher<sup>1</sup>(✉), Bram Stieltjes<sup>2</sup>, and Klaus H. Maier-Hein<sup>1,3</sup>

<sup>1</sup> Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany

[p.neher@dkfz.de](mailto:p.neher@dkfz.de)

<sup>2</sup> Radiology and Nuclear Medicine Clinic, University Hospital Basel, Basel, Switzerland

<sup>3</sup> Section for Automated Image Analysis, Heidelberg University Hospital, Heidelberg, Germany

**Abstract.** Diffusion tractography suffers from a difficult sensitivity-specificity trade-off. We present an approach that leverages knowledge about anatomically well-known tracts (anchor tracts) in a tractogram to quantitatively assess the remaining tracts (candidate tracts) according to their plausibility in conjunction with this context information. We show that our approach has the potential for greatly reducing the number of false positive tracts in fiber tractography while maintaining high sensitivities using phantom experiments (AUC 0.91). To investigate the applicability of the approach *in vivo*, we analyze 110 subjects of the Human Connectome Project young adult study. We demonstrate how the approach may be used for structured analysis of *in vivo* tractography and show supporting evidence for tracts previously discussed in the literature, while potentially sparking discussions about the role of others.

## 1 Introduction

The problem of false positives in fiber tractography is one of the grand challenges in the research area of diffusion-weighted magnetic resonance imaging (dMRI). Facing fundamental ambiguities especially in bottleneck situations, tractography generates huge numbers of theoretically possible candidate tracts [1, 2]. Only a fraction of these candidates is likely to correspond to the true fiber configuration, posing a difficult sensitivity-specificity trade-off. For example for the field of connectomics, which traditionally focuses on the high sensitivity regime, a recent study showed that specificity is crucial and twice as important as sensitivity when performing certain network analyses [3].

Current methods address the issue of false-positive tracts either by focusing exclusively on well-known fiber bundles using prior knowledge [4, 5] or by using tract filtering techniques based on the image signal [6, 7]. Currently, the link

between these two choices of purely data driven and prior knowledge based approaches is missing.

We propose a novel concept that rigorously exploits prior knowledge about the existence of anatomically known tracts (anchor tracts) to reduce the degrees of freedom of a successive data-driven filtering of the remaining candidate tracts: anchor-constrained plausibility (ACP). This approach is based on the hypothesis that information about the presence or absence of each anchor influences the plausibility of the candidates and thereby reduces the ambiguities in the problem. We demonstrate the potential of this concept to better handle the tractography sensitivity-specificity trade-off in a series of phantom experiments. Since quantitative *in vivo* evaluations of false-positive reduction rates would require a *ground truth* which does not exist, we concentrate on assessing the capabilities of ACP in enabling a structured and objective analysis of tractograms. Therefore we analyzed ACP scores in 110 subjects of the Human Connectome Project (HCP) young adult study and discuss the results in light of existing neuroanatomical knowledge, providing detailed data-driven insights into what we might be missing when focusing only on anatomically known tracts.

## 2 Methods

Essentially, our method scores the candidate tracts by assessing their contribution to the signal, subject to constraints imposed by the anchor tracts. The process consists of three steps:

**Preprocessing:** The input tractogram is filtered using tissue segmentations to discard streamlines that terminate inside the white matter (WM) or that enter the corticospinal fluid (CSF). Based on prior knowledge, anchor tracts are identified and extracted from the filtered tractogram. A variety of open-source tract selection methods such as *TractQuerier* [4], *RecoBundles* [5], *AFQ* [8] or the recently presented *TractSeg* [9] ([github.com/MIC-DKFZ/TractSeg](https://github.com/MIC-DKFZ/TractSeg)) are available for this step. The remaining streamlines are then clustered into individual bundles that represent the candidate tracts. This is achieved in a reproducible way by using cortex parcellations, assigning the streamlines to bundles according to their endpoint labels. Alternatively, e.g. in phantom images, clustering can be employed. Here, we used *QuickBundles* for this purpose [10].

**Residual Calculation:** It is now assessed which parts of the image can be explained by the anchor tracts, using a method similar to *LiFE* [7] but on a fixel [11] instead of a raw signal basis. This is done by fitting a scalar weight for each anchor streamline by minimizing the mean squared error (MSE) between the fiber orientation distribution function (fODF) peak magnitudes calculated from the input image and the corresponding streamline fixels:  $\text{MSE}(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2/n$ , where  $\mathbf{x}$  is the streamline weight vector,  $\mathbf{A}$  is the fixel magnitudes matrix with one column per streamline and one row per fODF peak direction and voxel and  $\mathbf{b}$  is the peak magnitudes vector representing the fODF image. The residual vector  $\mathbf{r}$  of this sparse system contains the fractions of the fODF peak magnitudes that cannot be explained by the anchor tracts:  $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$ , with all negative elements set to zero to retain only the unexplained parts of  $\mathbf{b}$ .

**Candidate Scoring:** Now it is analyzed which parts of the residual vector of the previous step can be explained by the candidate tracts. To this end the error  $\text{MSE}(\mathbf{y}) = \|\mathbf{B}\mathbf{y} - \mathbf{r}\|^2/n$  of a second linear system is minimized, where  $\mathbf{B}$  represents the candidate streamlines. As in *LiFE*, we define the score of each candidate tract as the root MSE it reduces:  $\sqrt{\text{MSE}(\mathbf{y}_i)} - \sqrt{\text{MSE}(\mathbf{y})}$ , where  $\mathbf{y}$  is the weight vector of all candidate streamlines and  $\mathbf{y}_i$  is the modified weight vector with entries corresponding to candidate tract  $i$  set to zero. Albeit determined by the weights of the individual candidate streamlines, this score is tract- and not streamline-specific. The procedure follows the intuition that it is plausible to assume a candidate’s existence if it can explain parts of the signal that are not explained by any known tract. On the other hand, candidate tracts that are exclusively composed of parts of the anchor tracts – which is a typical cause for false-positives [1] – receive a lower score. The score is interpreted as the candidate’s support by the data, given boundary conditions in form of anchor tracts (prior knowledge). It therefore represents a plausibility score for assuming a tract’s existence under consideration of the yet unexplained parts of the signal.

### 3 Experiments and Results

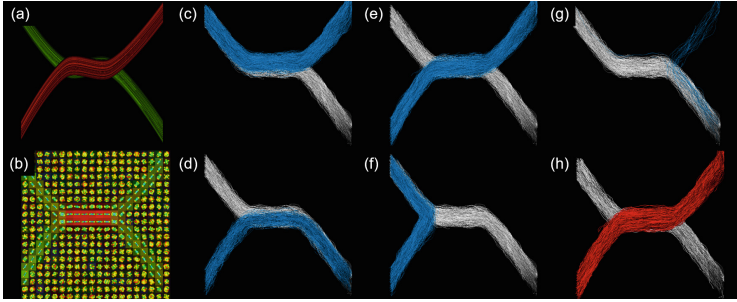
#### 3.1 In Silico Experiments and Results

We performed three phantom experiments with different degrees of complexity. Experiment 1 is intended as an illustration of the proposed method. The purpose of the other two experiments is to assess the capabilities of ACP in context of the sensitivity-specificity trade-off described in Sect. 1. For each phantom dataset, one test-tractogram was obtained using probabilistic constrained spherical deconvolution (CSD) tractography [12]. Since the ground truth is known in these cases, the anchor tracts could be simply defined using the binary masks of the respective ground truth tracts.

**Experiment 1:** Figure 1 illustrates the principle of the ACP analysis on an example consisting of two crossing fibers simulated with *Fiberfox* [13]. The Invalid Bundle Ratio (IVR) of the original tractogram was 3.8 (the invalid tracts outnumbered the valid tracts by a factor of 3.8). The experiment was performed once with each ground truth tract as anchor. For both configurations, this correctly resulted in the candidates corresponding to the second ground truth tract receiving the highest scores.

**Experiment 2:** For this experiment we employed a simulated replication of the *FiberCup* phantom consisting of seven individual tracts mimicking a coronal slice through the brain [13]. The IVR of the original tractogram was 1.9. The experiment was repeated five times with four out of seven randomly selected anchor tracts in each repetition. The three candidate tracts corresponding to the ground truth tracts were ranked highest in all repetitions (see Fig. 2a).

**Experiment 3:** The main phantom experiment is based on the brain-like phantom simulated with *Fiberfox* used in the ISMRM Tractography Challenge 2015 [1, 14]. The IVR of the original tractogram was 7.7. The experiment was repeated fifty times with 50% of the ground truth bundles extracted from the

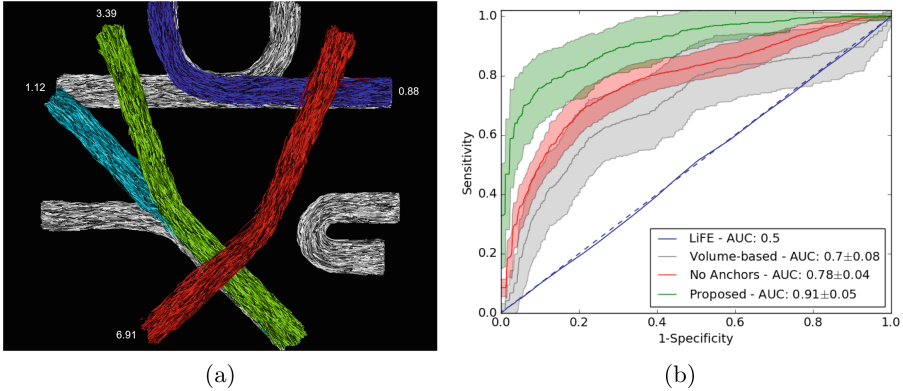


**Fig. 1.** Experiment 1, illustrating the proposed method. (a) and (b) show the ground truth: two crossing fiber tracts and their corresponding simulated ODF representation. (c-g) show exemplary candidate tracts (blue) obtained from the probabilistic tractogram using streamline clustering with the selected anchor tract (white). (h) shows the candidate tract (red) that was correctly ranked highest by the proposed method together with the anchor tract (white).

input tractogram randomly selected as anchor tracts in each repetition. Due to its large extent and dominance, the Corpus Callosum was always included in the set of anchor tracts. For comparison, another fifty repetitions were performed without any anchor tracts. Additional benchmarks were obtained using a volume-based ranking of the candidate tracts, as well as a streamline-weight-based ranking obtained with *LiFE* [7]. The resulting ROC curves are shown in Fig. 2b. The proposed method ( $AUC = 0.91$ ) performed significantly better than the benchmarks without anchor tracts ( $AUC = 0.78$ , t-test:  $p = 1.7^{-25}$ ) and with volume-based scoring ( $AUC = 0.7$ , t-test:  $p = 1.5^{-29}$ ). The *LiFE* streamline-weight-based ranking performed similar to random guessing ( $AUC = 0.5$ ).

### 3.2 In Vivo Experiments and Results

*In vivo* experiments were performed on 110 subjects of the HCP young adult study. For each subject we performed probabilistic CSD tractography with and without anatomical constraints (MRtrix) as well as deterministic peak tractography with anatomical constraints (MITK Diffusion). We used multiple tractography methods and joined the results for increased sensitivity and to mitigate tractography biases. From these whole brain tractograms we extracted 63 anchor tracts using overlap and streamline shape criteria, similar to *RecoBundles* [5], on basis of the reference tracts of the same subjects published by Wasserthal et al. [15]. Successively we generated the candidate tracts from the remaining streamlines by grouping them based on their endpoint locations with respect to the *FreeSurfer* Desikan-Killiany atlas cortex parcellations readily available for all HCP subjects [16]. To remove spurious streamlines from these tracts, a simple tract density based filtering was applied. Furthermore, streamlines that connect the same start and end label (loops) as well as very sparse tracts containing less than 50 streamlines were excluded from the subsequent analysis. This process



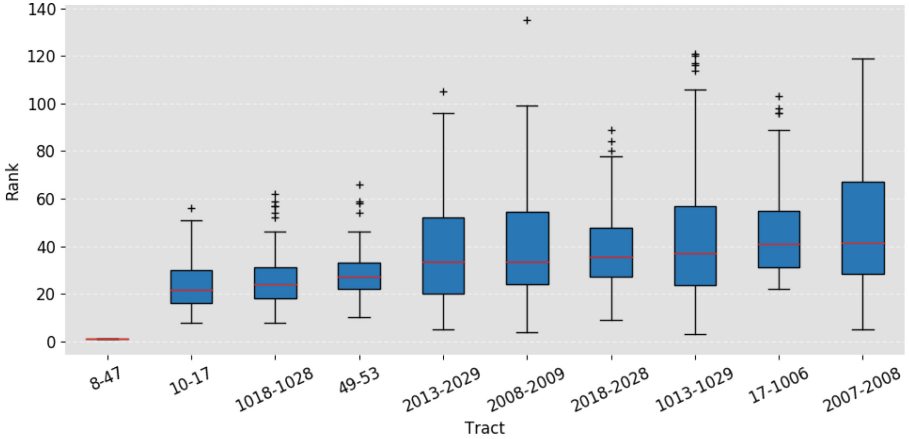
**Fig. 2.** (a) Experiment 2: Exemplary results on the simulated FiberCup dataset. The four highest ranked candidate tracts (colored) are labeled with their ACP score. The anchor tracts are colored white. The three real tracts received the highest scores. (b) Experiment 3: ROC curves on the ISMRM 2015 Tractography Challenge phantom obtained with the proposed approach (green), the same approach without anchor tracts (red), simple tract-volume-based ranking (gray) and *LiFE* streamline-weight-based ranking (blue).

resulted in an average number of 416 candidate tracts per subject that were included in the subsequent ACP analysis. In the results presented in the remainder of this section, we only included tracts that were detected in at least 90% of all subjects, resulting in 151 reproducible candidate tracts. These candidates are ranked per subject according to their ACP score.

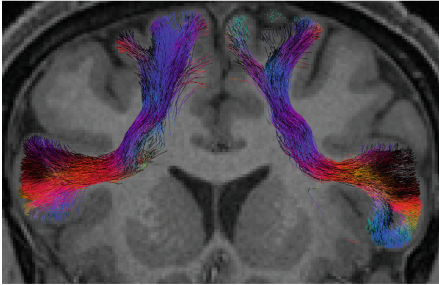
42% of the reproducible candidate tracts consisted of cortical U-fibers, i.e. tracts that connect neighboring gyri. Out of the top-ten ranking candidates, nine are U-fibers. This confirms the intended behaviour of our approach: U-fibers were not included in the set of anchor tract, but are known to exist. This is well reflected by their high ranking.

The top-ten ranking non-U-fiber candidates (see Fig. 3a) include several well known tracts such as the frontal aslant tract (FAT) (left and right hemisphere, parcellation labels 1018–1028 and 2018–2028, see Fig. 3b), tracts connecting the hippocampus and the thalamus (left and right hemisphere, 10–17 and 49–53), which are arguably part of the Fornix but missing from the anchor tracts, as well as connecting the hippocampus and the entorhinal cortex (left hemisphere, 17–1006), which might also include parts of the lower cingulum or the stria terminalis. Furthermore ranked in the top-ten, albeit with a larger ranking variance across subjects compared to the aforementioned tracts, are vertical tracts between the lingual gyrus in the occipital lobe and the superior parietal lobule (left and right hemisphere, labels 1013–1029 and 2013–2029) as well as fibers from the inferior parietal to the inferior temporal (right hemisphere, labels 2008–2009) and fusiform gyrus (right hemisphere, labels 2007–2008). The corresponding contralateral tracts that did not reach the top-ten are still ranked relatively

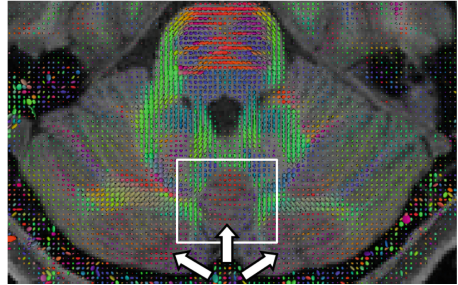
high (top-twenty). The overall highest ranked tract across all subjects consists of fibers connecting the left and right cerebellar cortex (parcellation labels 8 and 47), jumping from one cerebellar hemisphere to the other. In this case, ACP ranking was helpful in identifying a systematic tractography artifact that arises from the strong left-right anisotropy in this region and the limited image resolution. The gap between the two hemispheres, which are tightly pressed against each other, is not adequately resolved and makes the region appear as continuous tissue (see Fig. 3c).



(a) Top-ten non-U-Fiber candidate tracts



(b) FAT ("1018-1028" & "2018-2028")



(c) Cerebellum (causing "8-47")

**Fig. 3.** (a) Top-ten ranked non-U-fiber candidate tracts. The tracts are named according to their parcellation labels described in the text. (b) Coronal view of left and right frontal aslant tract (FAT) of a random subject. (c) Axial view on the cerebellum T1 with overlaid tensor glyphs. The arrows indicate areas with relatively high left-right anisotropy ( $FA > 0.15$ ). The area in the white box caused systematic false positives in the tractograms (see text).

## 4 Discussion and Conclusion

We proposed a novel concept, anchor-constrained plausibility analysis (ACP), that derives quantitative candidate plausibility scores by jointly assessing tract-based signal contribution levels and prior knowledge in the form of anchor tracts. We evaluate the concept in multiple phantom experiments, showing that this approach has the potential to greatly improve the sensitivity-specificity trade-off in tractography, which is a central issue of current tractography pipelines [1, 3]. Our *in vivo* experiments in a cohort of 110 subjects of the HCP project showed that the presented approach yields valuable information for a structured and objective analysis of tractography results.

Even though there is no *ground truth* for the *in vivo* evaluation, the experiments yielded several interesting insights. First, it was reassuring that well-known tracts which were not included as anchor tracts, i.e. the cortical U-fibers or FAT, received high plausibility scores. Second, ACP scoring turned out to be helpful in assessing the quality of the existing anchor tracts and the tractogram in general: parts of the Fornix and smaller connections between the hippocampus and the entorhinal cortex were missing in the reference and consistently popped up as tracts with high ACP scores [17]. Another tract that was systematically scored high turned out to be a systematic artifact of tractography, which we were previously not aware of. Third, ACP scores could play a role in ongoing discussions on brain anatomy. The high ranked vertical association tracts, for example, connecting the lingual gyrus in the occipital lobe and the superior parietal lobule, seem to be associated to the structure identified as vertical occipital fasciculus (VOF) by Yeatman et al. [18]. Other high ranked tracts could not directly be associated with known anatomy, such as connections between inferior parietal and inferior temporal gyrus, which visually reminds of a U-fiber bundle, or between the inferior parietal and fusiform gyrus. Both examples seem to be unrelated to prominent functional connections, but they are consistently important to explain the image data.

In all these considerations, though, it is important to keep in mind that a low score does not necessarily indicate that the respective tract is a false positive but only that its existence is not essential for explaining the measured data. Vice versa, a tract with high score is required to explain the data and it is therefore often plausible to assume its existence. Nevertheless, in some cases, as demonstrated for the cerebellum, additional factors such as the limited image resolution, missing streamlines in the tractogram and prior anatomical knowledge have to be taken into account to assess a tract's overall plausibility.

All methods described in this work are available online in the open-source Medical Imaging Interaction Toolkit ([mitk.org/wiki/DiffusionImaging](http://mitk.org/wiki/DiffusionImaging)), MRtrix ([mrtrix.org](http://mrtrix.org)) or Dipy ([nipy.org/dipy](http://nipy.org/dipy)). Future work will investigate the proposed approach in conjunction with connectomics analyses where recent studies have highlighted the disruptive impact of insufficiently specific tractography on global network measures [3], as well as the influence of the proposed filtering on the relationship between the structural and functional connectome.

**Acknowledgements.** This work was supported by the German Research Foundation (DFG) grant numbers MA 6340/10-1 and MA6340/12-1. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54-MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

## References

1. Maier-Hein, K.H., et al.: The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* **8**(1), 1349 (2017)
2. Sinke, M.R.T., et al.: Diffusion MRI-based cortical connectome reconstruction: dependency on tractography procedures and neuroanatomical characteristics. In: *Brain Structure and Function* (2018)
3. Zalesky, A., et al.: Connectome sensitivity or specificity: which is more important? *Neuroimage* **142**, 407–420 (2016)
4. Wassermann, D., et al.: The white matter query language: a novel approach for describing human white matter anatomy. *Brain Struct. Funct.* **221**(9), 4705–4721 (2016)
5. Garyfallidis, E.: Recognition of white matter bundles using local and global streamline-based registration and clustering. *Neuroimage* **170**, 283–295 (2017)
6. Sherbondy, A.J., et al.: ConTrack: finding the most likely pathways between brain regions using diffusion tractography. *J. Vis.* **8**(9), 15.1–16 (2008)
7. Pestilli, F., et al.: Evaluation and statistical inference for human connectomes. *Nat. Methods* **11**(10), 1058–1063 (2014)
8. Yeatman, J.D., et al.: Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS ONE* **7**(11), e49790 (2012)
9. Wasserthal, J., et al.: TractSeg - fast and accurate white matter tract segmentation. *Neuroimage* **183**, 239–253 (2018)
10. Garyfallidis, E., et al.: Quickbundles, a method for tractography simplification. *Front. Neurosci.* **6**, 175 (2012)
11. Raffelt, D.A., et al.: Connectivity-based fixel enhancement: whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres. *NeuroImage* **117**, 40–55 (2015)
12. Tournier, J.D., et al.: MRtrix: diffusion tractography in crossing fiber regions. *Int. J. Imaging Syst. Technol.* **22**(1), 53–66 (2012)
13. Neher, P.F., et al.: Fiberfox: facilitating the creation of realistic white matter software phantoms. *Magn. Reson. Med.* **72**(5), 1460–1470 (2014)
14. Maier-Hein, K., et al.: Tractography Challenge ISMRM 2015 Data (2015). type: dataset. <https://doi.org/10.5281/zenodo.572345>
15. Wasserthal, J., et al.: High quality white matter reference tracts (2017). type: dataset. <https://doi.org/10.5281/zenodo.1088278>
16. Desikan, R.S., et al.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**(3), 968–980 (2006)
17. Nieuwenhuys, R., et al.: *The Human Central Nervous System: A Synopsis and Atlas*. Springer, Heidelberg (1988). <https://doi.org/10.1007/978-3-540-34686-9>
18. Yeatman, J.D., et al.: The vertical occipital fasciculus: a century of controversy resolved by in vivo measurements. *PNAS* **111**(48), E5214–E5223 (2014)