

## How should tractography go forward? A Tractometer evaluation of local reconstruction and tracking

Jean-Christophe Houde<sup>1</sup>, Emmanuel Caruyer<sup>2</sup>, Alessandro Daducci<sup>3</sup>, and Maxime Descoteaux<sup>1</sup>

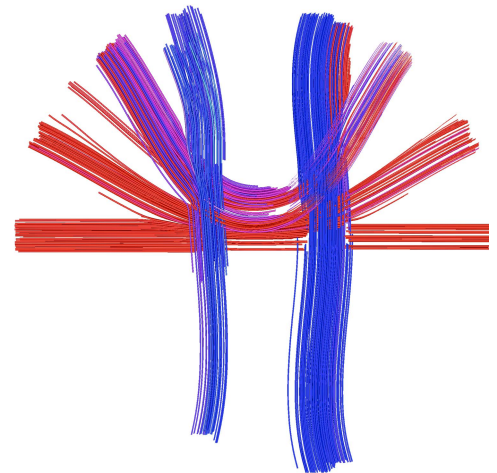
<sup>1</sup>Computer Science Department, Université de Sherbrooke, Sherbrooke, Québec, Canada, <sup>2</sup>Department of Radiology, University of Pennsylvania, Pennsylvania, United States, <sup>3</sup>Signal Processing Lab (LTS5), École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland

**Target audience** – This abstract is of interest to all researchers developing diffusion MRI (dMRI) processing methods for tractography and connectivity analysis. The presented method and related open questions apply to the overall dMRI processing pipeline, including raw image denoising, local diffusion modeling, tractography, tractogram post processing and connectivity analysis steps.

**Purpose** – This abstract presents the Tractometer evaluation system [1] ([www.tractometer.org](http://www.tractometer.org)) and how it was used to evaluate the entries of the ISBI 2013 HARDI Reconstruction Challenge [2] to find the best local reconstruction methods of the contest. This abstract raises questions that researchers should take into consideration when choosing and developing new dMRI methods aimed at improving tractography.

**Methods** – The Tractometer is a software evaluation system that wraps common dMRI processing tools, tests these tools with various parameter combinations against known ground-truth(s), and scores the results using global connectivity metrics. Supported tools include TrackVis [3], MRtrix [4], Dipy [5], Mitk Diffusion [6], as well as in-house tools containing unpublished methods. The tests are currently run on the Fibercup [7] and ISBI 2013 HARDI Reconstruction Challenge [2] ground-truth datasets. Connectivity metrics are computed using regions of interest (ROIs) defined at the end of each bundle, which allow the definition of pairs of ROIs representing various types of connections and bundles. Metrics include: valid connections (VC): streamlines connecting 2 ROIs known to be connected in the dataset; valid connections with an invalid path (VCIP): streamlines connecting 2 valid ROIs through an invalid path; invalid connections (IC): streamlines connecting 2 ROIs that are not connected in the dataset; no connection (NC): streamlines which are not connecting 2 ROIs. From these metrics, the number of correctly reconstructed bundles, called Valid bundles (VB), as well as the number of incorrect bundles, called Invalid Bundles (IB), are computed. For the ISBI 2013 HARDI Reconstruction challenge, 27 reconstructed local models were submitted. Tractography tests were run using the Tractometer. Once scored, a winner was found for each of the 3 categories (DTI, single-shell HARDI, DSI acquisition) by: 1) determining a selection criteria based on the mean and standard deviation of VB, IB, VC and IC, 2) considering the frequency and position of first occurrence of this result in all results matching the selection criteria.

**Results and discussion** – For this evaluation, combining all parameters for the submitted methods yielded 19,806 results, with mean VB: 19.35 and mean IB: 76 (recall that there are 27 VB in the phantom). Using the previously mentioned method, one submission was chosen as the winner for each category. While it was possible to draw some conclusions from the challenge (eg: denoising improves almost all the reconstruction methods), using the Tractometer system to find winners raised a few important questions. One interesting observation is that the winners of the contest, which performed best according to global metrics, did not normally get the best results when looking at local angular error (AE) metrics. Inversely, methods that had the best AE scores did not perform as well when evaluated using the Tractometer global metrics. This raises the question: how important is it to have a good AE, when considering the tractography pipeline as a whole? Most local model development is done on simulations of single and crossing fibers configurations. Those models are then evaluated using only the AE score. But what is the impact on tractography when transitioning from a 45 degrees to a 43 degrees crossing, for example, knowing that the whole tractography process tends to smooth small local variations? In this case, is AE a good quality metric for local model development? Another puzzling issue was the influence of the selection criteria to determine the winners. Changing only one of the criteria (eg: removing the constraint on IB number) could have totally changed the final results. Then, we must ask ourselves: what are the most important features of a good dMRI tractography pipeline? This will always depend on how this result is used. A neurosurgeon might prioritize having the best VB possible, at the cost of increased IC, to make sure that the area around a tumor is really well explored, whereas someone creating a structural connectome database would probably want to have a very high VB, with close to zero IB. However, choosing parameters to get such VB/IB trade-off would most certainly result in missing some important VB. Hence, in the era of connectomics, the community needs to take into account IB and errors of tractography more seriously. These errors are often overlooked because the final analysis is done in conjunction with anatomical priors that can filter and hide them. While this works correctly when using streamlines bundles as regions of interest for further analysis, IB are a plague in connectomics studies. They add a large amount of errors in the connectivity matrices, and are hard to filter out and detect automatically. For example, on a 64 directions, noiseless acquisition of the ISBI 2013 HARDI reconstruction challenge phantom, modeled using Constrained Spherical Deconvolution [8] of order 8, the best probabilistic tractography output found 26 VB out of the 27 possible, while also finding 93 IB! This is a ratio of 3.57 IB for each VB. Over all techniques that recovered 27/27 VB, the lowest IB count was 45!



**Fig 1:** A subset of the valid bundles computed on the ground truth peak directions of the ISBI 2013 HARDI reconstruction challenge

Taking all these points together, the community must ask itself the following questions: For the tractography end user, what is really important? More precise local reconstructions? Tracking algorithms with the most valid bundles? The least number of invalid bundles? Tracking with better bundle white matter coverage? Is angular error the best quality metric for local model development? Could we develop methods where the behavior is tunable towards the desired use of the final result? How should we deal with IB and errors in tractography? And most critically, how can we evaluate tractography pipelines on real data?

**Conclusion** – The Tractometer system has been developed to quantitatively evaluate dMRI processing pipelines. Using the newly defined connectivity metrics, it was possible to evaluate the entries of the ISBI 2013 HARDI Reconstruction challenge. However, using the Tractometer in real life raised important questions, whose answers will shape the choice and design of future dMRI processing methods.

**References** – [1] Côté et al., MedIA 2013. [2] [http://hardi.epfl.ch/static/events/2013\\_ISBI/](http://hardi.epfl.ch/static/events/2013_ISBI/) [3] TrackVis: <http://www.trackvis.org/> [4] Tournier et al., IJIST 2012. [5] Dipy: <http://www.dipy.org> [6] Fritzsche et al., MIM 2012. [7] Fillard et al., Neuroimage 2011. [8] Tournier et al., Neuroimage 2007.