Data in Brief 38 (2021) 107429

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

The diffusion-simulated connectivity (DiSCo) dataset

Jonathan Rafael-Patino[a,1], Gabriel Girard[a,b,c,1,*], Raphaël Truffet[d], Marco Pizzolato[a,e], Emmanuel Caruyer[d], Jean-Philippe Thiran[a,b,c]

[a] Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Switzerland
[b] CIBM Center for Biomedical Imaging, Switzerland
[c] Radiology Department, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland
[d] Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn ERL Rennes, U-1228, France
[e] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

ARTICLE INFO

Article history:
Received 9 April 2021
Revised 8 September 2021
Accepted 22 September 2021
Available online 25 September 2021

Keywords:
Diffusion MRI
Monte Carlo
Simulation
Numerical phantom
Structural connectivity
Tractography
Microstructure

ABSTRACT

The methodological development in the mapping of the brain structural connectome from diffusion-weighted magnetic resonance imaging (DW-MRI) has raised many hopes in the neuroscientific community. Indeed, the knowledge of the connections between different brain regions is fundamental to study brain anatomy and function. The reliability of the structural connectome is therefore of paramount importance. In the search for accuracy, researchers have given particular attention to linking white matter tractography methods – used for estimating the connectome – with information about the microstructure of the nervous tissue. The creation and validation of methods in this context were hampered by a lack of practical numerical phantoms. To achieve this, we created a numerical phantom that mimics complex anatomical fibre pathway trajectories while also accounting for microstructural features such as axonal diameter distribution, myelin presence, and variable packing densities. The

* Corresponding author at: EPFL STI IEL LTS5, ELD 232 (Bâtiment ELD), Station 11, CH-1015 Lausanne.
E-mail addresses: jonathan.patinolopez@epfl.ch (J. Rafael-Patino), gabriel.girard@epfl.ch (G. Girard), raphael.truffet@irisa.fr (R. Truffet), mapiz@dtu.dk (M. Pizzolato), emmanuel.caruyer@irisa.fr (E. Caruyer), jean-philippe.thiran@epfl.ch (J.-P. Thiran).
Social media: (J. Rafael-Patino), (G. Girard), (R. Truffet), (M. Pizzolato), (E. Caruyer), (J.-P. Thiran)
[1] These authors contributed equally.

https://doi.org/10.1016/j.dib.2021.107429

substrate has a micrometric resolution and an unprecedented size of 1 cubic millimetre to mimic an image acquisition matrix of $40 \times 40 \times 40$ voxels. DW-MRI images were obtained from Monte Carlo simulations of spin dynamics to enable the validation of quantitative tractography. The phantom is composed of 12,196 synthetic tubular fibres with diameters ranging from 1.4 µm to 4.2 µm, interconnecting sixteen regions of interest. The simulated images capture the microscopic properties of the tissue (e.g. fibre diameter, water diffusing within and around fibres, free water compartment), while also having desirable macroscopic properties resembling the anatomy, such as the smoothness of the fibre trajectories. While previous phantoms were used to validate either tractography or microstructure, this phantom can enable a better assessment of the connectome estimation's reliability on the one side, and its adherence to the actual microstructure of the nervous tissue on the other.

Specifications Table

Subject	Medical imaging
Specific subject area	Structural connectivity estimation from diffusion-weighted magnetic resonance imaging (DW-MRI)
Type of data	Image
How data were acquired	The in silico DW-MRI data was obtained using a Monte Carlo simulations of spin dynamics. The synthetic substrate was designed and optimized to mimic complex white matter axonal configurations.
Data format	raw
Parameters for data collection	The in silico DW-MRI was simulated for 360 measurements, using four b-values ($b = 1000$, 1925, 3094, 13,191 s/mm^2). Additionally, four non-diffusion-weighted images are included. The echo time was set to 0.0535 s. The DW-MRI dataset was then corrupted with various levels of Rician noise.
Description of data collection	The mesh of the substrate of 1 cubic millimetre was divided into 64,000 voxels prior to the Monte Carlo simulation. The Monte Carlo simulation was performed independently for each voxel, and summed to form the final images. One particle per micrometre cube was initiated for a total of 10^9 Monte Carlo particles.
Data source location	Ecole polytechnique fédérale de Lausanne (EPFL) Lausanne Switzerland
Data accessibility	public repository: Repository name: Mendeley Data Data identification number: 10.17632/fgf86jdfg6.1 Direct URL to data: https://data.mendeley.com/datasets/fgf86jdfg6/1

Value of the Data

- A novel DW-MRI in silico dataset with both microstructure and macrostructure complexity, obtained from Monte Carlo simulations of spins dynamics, to improve the estimation of the quantitative structural connectivity.
- The dataset can benefit researchers targeting method development for quantitative structural connectivity estimation from diffusion-weighted MRI.
- This novel numerical phantom, with a ground truth quantitative connectivity, was designed for the testing and validation of the DW-MRI processing pipeline, from signal pre-processing,

algorithms for fibre orientation distributions estimation and tractography algorithms, to tractography filtering and algorithms for the calculation of connectivity strength.
- This phantom has increased complexity from previously available dataset, and may help improving the reconstruction methods of the quantitative structural connectome. In particular, in this large and complex substrate, the volume of the axon-like structure is preserved across their trajectories, allowing for microstructure-informed tractography methods.

1. Data Description

The four-dimensional (4D) DW-MRI signal obtained from Monte Carlo simulation is available in the file 'DiSCo_DWI.nii.gz' in the standard Nifti format. The image is of dimension $40 \times 40 \times 40 \times 364$, where the 4th dimension corresponds to the DW-MRI gradient parameters (b-values and b-vectors). The files 'DiSCo_DWI_ RicianNoise-snr*.nii.gz' are the same DW-MRI images but corrupted with Rician noise at a signal to noise ratio (SNR) of 10, 20, 30, 40 and 50. The file 'DiSCo_gradients.bvals' lists the b-values (s/mm^2) of the 364 DW-MRI simulated measurements. The files 'DiSCo_gradients_ dipy.bvecs', 'DiSCo_gradients_fsl.bvecs' and 'DiSCo_gradients_mrtrix.b' list the corresponding gradient orientation formatted for input to the DIPY [1], FSL [2] and MRtrix3 [3] software library, respectively.

The file 'DiSCo_ROIs.nii.gz' contains the 3D location of the 16 ROIs of the phantom. Voxels of the ROIs have a corresponding value of 1 to 16, the rest of the voxels of the image have a zero value. The file 'DiSCo_Intra_Strand_Volume_Fraction.nii.gz' maps the fraction of Monte Carlo particles initiated inside the inner tubular mesh of the substrate over the total number of Monte Carlo particles used to generate the signal. The file 'DiSCo_mask.nii.g' identifies all voxels with one or more particles initiated inside the inner tubular mesh of the substrate.

The files 'DiSCo_Strands_Trajectories.trk' and 'DiSCo_Strands_Trajectories.tck' contain the centerline trajectories of the 12,196 strands in the Trackvis [4] and MRtrix3 [3] formats, respectively. The files 'DiSCo_Strands_Diameters.txt' and 'DiSCo_Strands_ROIs_Pairs.txt' list the corresponding inner diameters and the two ending ROIs labels of each strand, respectively. The file 'DiSCo_Connectivity_Matrix_Strands_Count.txt' contains the 16×16 array where each row and column correspond to one of the 16 ROIs. The values of the array elements correspond to the number of strands connecting each pair of ROIs. The file 'DiSCo_Connectivity_Matrix_Cross-Sectional_Area.txt' contains the same array but with the weights corresponding to the sum of the cross-sectional areas of the strands interconnecting the ROIs.

Finally, the file 'DiSCo_mesh.ply' contains the mesh of the substrate used for the Monte Carlo simulation of the DW-MRI signal in the PLY polygon format.

2. Experimental Design, Materials and Methods

In this work, we have designed a novel dataset of simulated Diffusion-Weighted MRI (DW-MRI) images from a numerical phantom to foster the development of diffusion tractography and connectivity methods. The DiSCo (Diffusion-Simulated Connectivity) phantom is composed of 12,196 tubular fibres (strands), with gamma-distributed inner diameters ranging from 1.4 µm to 4.2 µm, connecting 16 distant Regions of Interest (ROIs). The strands form different white matter configurations (e.g., kissing, branching) intersect at different crossing angles, split and group after leaving and ending in the ROIs. The simulation substrate has a micrometric resolution and an unprecedented size of 1 cubic millimetre to mimic an image acquisition matrix of $40 \times 40 \times 40$ voxels (64,000 voxels).

After the Monte Carlo simulation of the DW-MRI signal, the resulting image header was set to a voxel size of 1.0 mm isotropic (from 25 µm isotropic), for a final image size of $4 \times 4 \times 4$ cm^3, compatible with conventional diffusion tractography methods. The simulated images capture the microscopic properties of the tissue (e.g. fibre diameter, water diffusing within and around fibres, free water compartment), while also having desirable macroscopic properties resembling

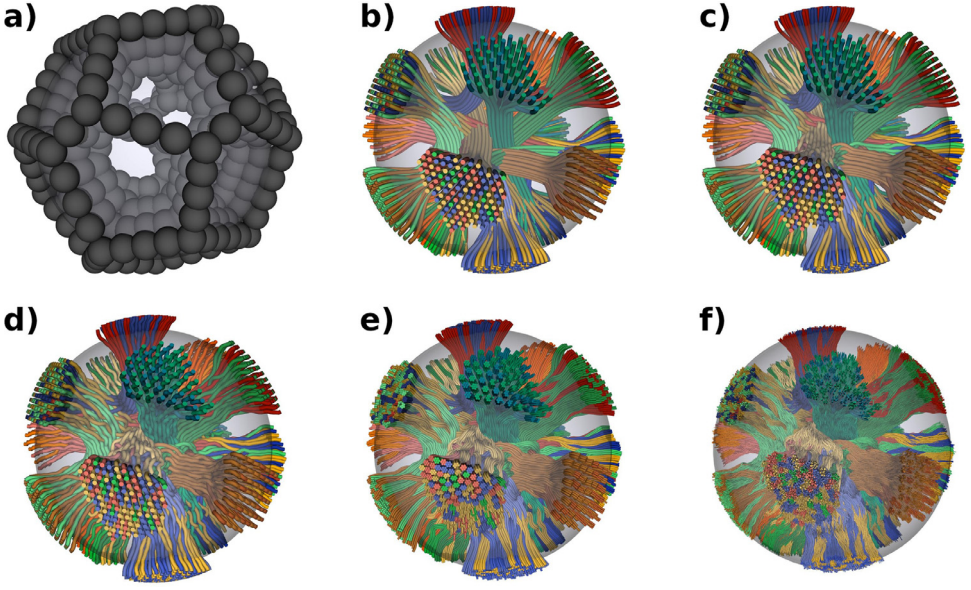

Fig. 1. a) Spherical constraints located around each ROIs, used to force the strands trajectories to converge toward the centre of the phantom. b) Initial strands trajectories. c, d) Strands trajectories after 1 and 4 optimisation steps of the Numerical Fiber Generator, respectively. e) Subdivided strands following hexagonal packing. f) Strand trajectories after the optimisation process, with gamma-distributed diameters. The colours of the strands correspond to a pair of connected ROIs.

the anatomy, such as the smoothness of the fibre trajectories. The values of the 16×16 connectivity matrix correspond to the cross-sectional area of the strands between pairs of ROIs.

2.1. Phantom design

The initial connectivity matrix weights of the DiSCo phantom were randomly generated for 16 ROIs. By controlling for sparsity of the resulting matrix, we obtained 25 non-zero weights, for a total of 120 possible connections among the ROIs. These weights were used to initiate a proportional set of strands of 15 µm in diameter interconnecting the ROIs going toward the centre of the sphere (789 strands, see Fig. 1b). The 16 ROIs were generated on the surface of a sphere of 1 mm in diameter using the spherical Voronoi algorithm [5]. The strands trajectories were then optimised using the Numerical Fiber Generator [6]. To force strands convergence toward the centre of the phantom, we added three layers of spherical constraints of 0.1 mm in diameter along the edges of each ROIs (see Fig. 1a).

The cost function had energy terms controlling for strands curvature and strands length. Moreover, the cost function increases when strands overlap with other strands or with spherical constraints. Multiple optimisation iterations were performed to slowly increase the cost of overlapping strands (see Fig. 1c,d). Each strand was then subdivided into strands of 7.5 µm in diameter following hexagonal packing (5523 strands, see Fig. 1e). Strands were then optimized to reduce overlaps, length and curvature. Finally, each strand was subdivided into cylinders which diameters following a gamma distribution $\Gamma(\kappa, \theta)$, with shape, $\kappa = 0.5$, and scale $\theta = 0.007$ (minimum diameter of 2 µm and maximum diameter of 6 µm). Cylinders were iteratively sampled and placed within each strand cross-sectional surface until a density of 0.7 was reached (12,196 strands, see Fig. 1f). Sampled cylinders not fitting within the surface were discarded. The

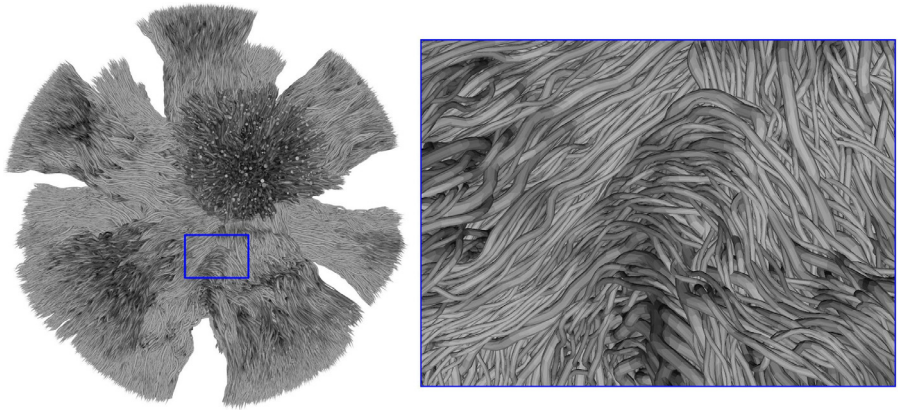

Fig. 2. Mesh of the 12,196 strands used as input to the MC/DC diffusion simulator.

Numerical Fiber Generator optimisation procedure was performed for an additional iteration to further reduce overlapping strands and use the space left by the sampling procedure.

The final set of strands (trajectories and diameters) were used to generate a mesh of the substrate with the Blender software (see Fig. 2). For each strand, a tubular mesh was generated to represent the outer surface of the axon-like structure. An additional inner tubular mesh representing the inner surface of the axon-like structure was generated following the same trajectory, but with a diameter of 0.7 times the outer diameter. The mesh was used as input to the open-source Monte Carlo Diffusion and Collision Simulator (MC/DC) proposed by Rafael-Patino et al. [7].

2.2. Diffusion-weighted MRI signal generation using Monte Carlo simulations of spins dynamics

All the signals were computed using the sum of the accumulated phase shift approximation implemented in the MC/DC simulator [7]. The simulator was specifically adapted to handle substrates with millions of triangles by using a tree-based spatial partitioning of the substrates in conjunction with predefined diffusion conditions; for example, by adjusting the collision maximum distance to the particle's theoretical maximum displacement. Additionally, particles initiated within closed meshes, such as intra-axonal space, can further prune the collision trees by discarding any triangles outside the defined closed domain. We adjusted the parameters of those procedures to improve performance under the simulation conditions and substrate size previously specified, allowing us to compute the whole volume simulation in a reasonable amount of time.

To obtain an accurate parameter estimate of the ensemble average, simulations must contain as many particles as possible to effectively lower the simulation noise. To ensure a high and uniform density in all the 64,000 computed voxels, we used a highly dense regular particle placement with a total of 10^9 particles to achieve a density of one particle per cubic micrometre. The total diffusion time was set to 53.5×10^{-3} s matching the maximum protocol's echo time. The time between each step δt was set to 5.35×10^{-7} s.

The simulated signal was computed separately by labelling all particles inside the inner-mesh as "intra-axonal," those inside the outer-mesh but outside the inner mesh as "myelin," and those outside the outer-mesh as "extra-axonal." The unrestricted diffusion coefficient of the Monte Carlo particles was set to 0.6×10^{-3} mm^2/s. Particles initiated in the "extra-axonal" and "intra-axonal" compartments were used for the DW-MRI signal generation. The simulation parameters for both compartments were the same.

Although the substrate was optimized to minimize strand overlaps, some overlaps remained. To account for the effect of overlapping strands in the simulated DW-MRI signal, the "intra-axonal" signal was generated solely using the mesh of the strand into which the particle was initiated. If a particle is initiated at a point where two strands overlap and thus within more than one strand, one of the strands was randomly chosen to simulate the "intra-axonal" diffusion process.

The DW-MRI protocol includes 360 diffusion-weighted images and 4 non-diffusion-weighted images ($b = 0$). The measurements are distributed on 4 b-shells (b =1000, 1925, 3094, 13,191 s/mm^2). Those correspond to the 3 b-shells of the ActiveAx protocol [8] with an additional shell at $b = 1000$ s/mm^2. The echo time was set to 0.0535 s. Each shell is sampled using 90 uniformly distributed gradient directions on the sphere. The resulting DW-MRI signal was corrupted using various levels of Rician noise using the Diffusion Imaging in Python (DIPY) library [1]. For each voxel of the image, the 1-dimensional diffusion signal S was corrupted following

$$S = \sqrt{(S + N_1)^2 + N_2^2}, \tag{1}$$

where N_1 and N_2 are 1-dimensional vectors with value sampled from a zero-mean Gaussian distribution. The standard deviation was set to the non-diffusion-weighted ($b = 0$) signal intensity divided by the target SNR.

Ethics Statement

No human or animal data were used or collected in this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT Author Statement

Jonathan Rafael-Patino: Conceptualization, Methodology, Investigation, Writing – original draft; **Gabriel Girard:** Conceptualization, Methodology, Investigation, Writing – original draft; **Raphaël Truffet:** Conceptualization, Methodology, Writing – review & editing; **Marco Pizzolato:** Conceptualization, Methodology, Writing – review & editing; **Emmanuel Caruyer:** Conceptualization, Methodology, Investigation, Writing – review & editing, Funding acquisition; **Jean-Philippe Thiran:** Conceptualization, Writing – review & editing, Funding acquisition.

Acknowledgments

We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG). The calculations have been performed using the facilities of the Scientific IT and Application Support Center of EPFL. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. This project has received funding from the Swiss National Science Foundation under grant number 205320_175974 and Spark grant number 190297. Marco Pizzolato acknowledges the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 754462. This research project is part

of the MMINCARAV Inria associate team program between Empenn (Inria Rennes Bretagne Atlantique) and LTS5 (École Polytechnique Fédérale de Lausanne - EPFL) started in 2019. Raphaël Truffet's Ph.D. is partly funded by ENS Rennes.

References

[1] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, I. Nimmo-smith, Dipy Contributors, Dipy, a library for the analysis of diffusion MRI data, Front. Neuroinform. 8 (2014) 1–5. http://www.frontiersin.org/Journal/10.3389/fninf.2014.00008/abstract

[2] S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R.K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, P.M. Matthews, Advances in functional and structural MR image analysis and implementation as FSL, NeuroImage 23 (Suppl 1) (2004) S208–19. http://www.ncbi.nlm.nih.gov/pubmed/15501092

[3] J.D. Tournier, R. Smith, D. Raffelt, R. Tabbara, T. Dhollander, M. Pietsch, D. Christiaens, B. Jeurissen, C.H. Yeh, A. Connelly, MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation, NeuroImage 202 (2019) 116137, doi:10.1016/j.neuroimage.2019.116137.

[4] R. Wang, T. Benner, A.G. Sorensen, V.J. Wedeen, Diffusion toolkit: a software package for diffusion imaging data processing and tractography, Intl Soc Mag Reson Med, Berlin, Germany, 2007.

[5] M. Caroli, P.M.M. de Castro, S. Loriot, O. Rouiller, M. Teillaud, C. Wormser, Robust and efficient delaunay triangulations of points on or close to a sphere, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), in: LNCS, 6049, 2010, pp. 462–473, doi:10.1007/978-3-642-13193-6_39.

[6] T.G. Close, J.-D. Tournier, F. Calamante, L.A. Johnston, I. Mareels, A. Connelly, A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms, NeuroImage 47 (4) (2009) 1288–1300, doi:10.1016/J.NEUROIMAGE.2009.03.077.

[7] J. Rafael-Patino, D. Romascano, A. Ramirez-Manzanares, E.J. Canales-Rodríguez, G. Girard, J.-P. Thiran, Robust Monte–Carlo simulations in diffusion-MRI: effect of the substrate complexity and parameter choice on the reproducibility of results, Front. Neuroinform. 14 (8) (2020) 8, doi:10.3389/fninf.2020.00008.

[8] D.C. Alexander, P.L. Hubbard, M.G. Hall, E.a. Moore, M. Ptito, G.J.M. Parker, T.B. Dyrby, Orientationally invariant indices of axon diameter and density from diffusion MRI, NeuroImage 52 (4) (2010) 1374–1389, doi:10.1016/j.neuroimage.2010.05.043.