The Microstructural Features of the Diffusion-Simulated Connectivity (DiSCo) Dataset
Jonathan Rafael-Patino, Gabriel Girard, Raphaël Truffet, Marco Pizzolato, Jean-Philippe Thiran, Emmanuel Caruyer

To cite this version:
Jonathan Rafael-Patino, Gabriel Girard, Raphaël Truffet, Marco Pizzolato, Jean-Philippe Thiran, et al.. The Microstructural Features of the Diffusion-Simulated Connectivity (DiSCo) Dataset. CDMRI 2021 - International Workshop on Computational Diffusion MRI, Oct 2021, Strasbourg, France. pp.159-170, 10.1007/978-3-030-87615-9_14. inserm-03366318

HAL Id: inserm-03366318
https://inserm.hal.science/inserm-03366318
Submitted on 5 Oct 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.
The Microstructural Features of the Diffusion-Simulated Connectivity (DiSCo) Dataset

Jonathan Rafael-Patino\textsuperscript{1,2,4}, Gabriel Girard\textsuperscript{1,2,3,}, Raphaël Truffet\textsuperscript{4}, Marco Pizzolato\textsuperscript{5,1}, Jean-Philippe Thiran\textsuperscript{1,2,3}, and Emmanuel Caruyer\textsuperscript{4}

\textsuperscript{1}Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
\textsuperscript{2}Radiology Department, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland
\textsuperscript{3}Center for BioMedical Imaging (CIBM), Lausanne, Switzerland
\textsuperscript{4}Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn ERL U-1228, F-35000, Rennes, France
\textsuperscript{5}Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

Abstract. We present a detailed description of the structural characteristics of the MICCAI 2021 Diffusion Simulated Connectivity (DiSCo) Challenge synthetic dataset. The DiSCo dataset are one of a kind numerical phantoms for the simulation of the diffusion-weighted images (DWIs) via Monte-Carlo diffusion simulations. The microscopic and macroscopic complexity of the synthetic substrates allows the evaluation of processing pipelines for the estimation of the quantitative structural connectivity. The diffusion-weighted signal in each voxel of the DWIs is obtained from Monte-Carlo simulations of particle dynamics within a substrate of an unprecedented size of 1mm\textsuperscript{3}, allowing for an image matrix size up to 40x40x40 voxels (isotropic voxel sizes of 25\textmu m). In this paper, we provide a characterization of the microstructural properties of the DiSCo dataset, which is composed of three numerical phantoms with comparable microstructure. We report the ground-truth tissue volume fractions (“intra-axonal”, “extra-axonal”, “myelin”), the fibre density, the bundle density and the fibre orientation distributions (FODs). We believe that this characterization will be beneficial for validating quantitative structural connectivity processing pipelines, and that could eventually find use in microstructural modelling based on machine learning approaches.

Keywords: Monte-Carlo Simulations · DW-MRI · phantoms · tractography · microstructure

1 Introduction

For the last two decades, diffusion-weighted magnetic resonance imaging (DW-MRI) has been an active area of research, with numerous contributions to the de-
development of structural connectivity analyses. However, it is difficult to quantify the effect of a particular element of the DW-MRI data processing pipeline, like noise reduction methods [29, 9], local reconstruction methods of the angular diffusion information [30, 26], or tractography algorithms [27, 12], on the structural connectivity results. Furthermore, in order to obtain a quantitative comparison of these methods, the use of tracers on animal models [18], or post-mortem dissection, or cortical electro-stimulation is required [19]. These techniques are time-consuming and moderately to highly invasive, and they do not provide a systematic ground truth mapping of the axonal fibre pathways.

To overcome such challenges, some physical phantoms have been developed [19, 16], providing a convenient way to evaluate DW-MRI image processing methods in a more quantitative manner. However, these phantoms’ fibre geometries and microstructural features are typically much simpler than those found in the brain. Moreover, the precise structural measurements of the manufactured phantom may be not fully known, defeating the purpose of using such phantoms. Numerical phantoms are of particular interest in this context and have become the standard de facto for evaluating novel DW-MRI signal processing methods [7, 21, 4, 14, 1, 13, 6]. The realism of phantoms is a fundamental aspect. Generally, it is possible to think of two levels of realism connected to numerical phantoms for DWIs. One, macroscopic, has to do with the fidelity to the known key features of the tissue organization, such as the complex and convoluted trajectories of white matter fibres and their configuration. The other, microscopic, is the fidelity of the numerical phantoms to the potential properties of the tissue microstructure such as its composition — axons, myelin, etc. —, geometrical features — axonal radii —, and physicochemical characteristics that are relevant for characterizing the tissue magnetization — such as the transverse relaxation time.

Freely available software have been developed and released [7, 17, 8] to create numerical phantoms for validating structural connectivity pipelines. For instance, Phantomas [7] and Fiberfox [17] allow the creation of complex DW-MRI signal from user-defined fibre configurations and diffusion parameters. Additionally, the Numerical Fibre Generator (NFG) [8] framework generates numerical structures randomly, resulting in an intricate set of fibre bundles from which DW-MRI images are generated. While these methods are capable of generating DWIs from substrates containing a large number of bundles of axonal fibers, they fall short on the microscopic realism that is necessary for evaluating a more quantitative structural connectivity.

The fidelity to the microstructural properties of the white matter tissue can be achieved with Monte-Carlo Diffusion Simulation (MCDS). In contrast with the approaches mentioned before, MCDS does not require an explicit model of the diffusion signal. Instead, MCDS requires a precise physical representation of the tissue geometry in the form of a 3D mesh substrate used to generate the dynamics of virtual water particles diffusing within and interacting with the substrate’s barriers. MCDS is known for being computationally expensive and time-consuming. Moreover, it requires careful setup of the simulation parame-
ters, design of the 3D mesh substrates, and handling of the particle interactions. In recent years a notable effort has been made to introduce state-of-the-art methods to obtain faster and more robust simulations [21, 31], as well as state-of-the-art frameworks to create complex mesh substrates [13, 6]. However, the computational expensiveness of these methods has still limited its use to single-voxel simulations, away from the demands of connectome validation studies.

In the context of macro- and microscopically realistic simulations, and in an effort to provide means for jointly evaluating local reconstruction, tractography, and connectivity methods, we developed the DiSCo dataset, a Monte-Carlo based dataset of unprecedented complexity and volumetric size. The numerical DiSCo phantoms are large enough (1mm³) to test tractography and connectivity methods, while also having rich microstructural properties suitable for testing tissue biophysical modeling and orientation estimation methods. In this work, we present a detailed analysis of the MICCAI 2021 DiSCo challenge numerical phantoms, reporting ground-truth microstructural maps at various resolutions, such as the voxel-wise fibre orientation distributions, the compartmental volume fractions and fibre density, and the mean axon diameter distribution.

2 Methods

The three phantoms shown in Figure 1 (coined as DiSCo1, DiSCo2 and DiSCo3), were constructed following the procedure described on [20] using 16 randomly generated regions of interest (ROIs). The ROIs are then used to generate a connectome with desired properties, like sparsity, weight randomness and non-self connections. The main differences between the phantoms arise from the randomly generated ROIs and from the set of randomly generated non-zeros weights defining the weighted connection between them. However, due to the strands optimization procedure based on the NFG to pack and interdigitate the generated strands connecting the ROIs, structural differences are introduced in terms of the resulting number and orientations of the axons’ bundles per voxel, effective diameter distribution, and compartmental volume fractions. Some of these differences are known a priori from the ground truth information used for the design of the phantoms, however, due to the complexity of the resulting substrate, some other features need to be estimated after the phantom has been produced.

The phantoms contain three water tissue compartments, intra-axonal, extra-axonal and myelin. The signal was simulated separately for each compartment using the MC/DC simulator [21] using the settings described in [20]. The myelin compartment was simplified as a non-diffusing compartment with water fraction proportional to the myelin volume. All the maps we report were computed using the strands’ information generated from the final meshing procedure [20] in which an inner and outer layer was added as follows. The strands are defined by its center-line and the cross-sectional area, which are used to construct
Fig. 1. Meshes of the three phantoms (DiSCo1, DiSCo2, DiSCo3) obtained following the strand optimization procedure [20]. The strands have their endpoints on the surface a sphere and trajectories propagating inside the sphere. Each strand interconnects two of the 16 ROIs.

the outer and inner mesh given the strands trajectories. The outer mesh is defined using the strands cross-sectional diameter, from which an inner mesh is generated using a down-scaled diameter by a 0.7 factor (considered as the g-ratio). The bundles are then defined as the set of strands that starts and ends in the two specific ROIs.

2.1 Volume fraction estimation

The compartmental volume fractions were computed via the Monte-Carlo sampling procedure of the diffusion simulations. In order to do so, we tracked the position of each individual $i$-th particle at time 0, $p_{i,0}$, and evaluated to which compartmental domain $\Omega \in \mathbb{R}^3$ that position belongs. In particular, we defined the intra-axonal compartment, $\Omega_i; \Omega \subset \Omega_i$, as any enclosed domain with no other substrate elements inside; the outer axonal-space ($\Omega_o$) — related to a specific subspace $\Omega_i$ — was defined then as any enclosed domain fully containing the subspace of the intra-axonal subspace $\Omega_i$. With this, the compartmental myelin volume fraction can be defined as the space in between those two, $\Omega_m = \Omega_o - \Omega_i$. Finally, we defined the extra-axonal compartment as anything else outside the outer compartment ($\Omega_e = \Omega - \Omega_o$). The final volume fractions maps were computed by uniformly sampling the substrate space $\Omega$ with a particle density of one particle per $\mu m^2$. The volume fraction maps were computed by subdividing the averages into the voxel regions using the maximum resolution grid of $25 \times 25 \times 25 \mu m^2$.

2.2 Fibers information maps

The ground truth fibre orientation distribution functions (FODs) were computed using the strand trajectories and the cross-sectional areas. The FOD in a particular voxel is estimated from a collection of directions, representing the variability of the fibre directions within that voxel. This accounts for the different fiber bundles potentially passing through but not wholly contained in the voxel,
the diameters of the fibers, and for the angular dispersion of bending strands. In order to translate this discrete representation of the FOD into a continuous representation, we used kernel density estimation (KDE), using a symmetric Von-Mises Fisher kernel, defined as

\[ v_\mu(\omega) = c_\kappa \exp(\kappa |\mu^T \omega|), \quad \text{where} \quad c_\kappa = \frac{\kappa}{4\pi(\exp(\kappa) - 1)}, \quad (1) \]

where \( \kappa \) is the concentration parameter, and \( \mu \) is the axis. This function has already been used in the context of diffusion MRI modelling \[15, 32, 7\]. The FOD in each voxel is obtained by summing the kernel aligned with \( \mu \) for all fibre segments intersecting the voxel. Moreover, each kernel is weighted by the length of the segment and by its cross-sectional area to account for various fiber volumes. The diameter of the circumference defining the cross-sectional area is used to compute the effective axon diameter distribution map per voxel. The number of strands and bundles per voxel were also computed for three nominal resolutions (25 \( \mu m \), 50 \( \mu m \) and 100 \( \mu m \)) using the weighted approach of the FOD explained before being separated by strand or by ROI bundle.

2.3 Peaks extraction

Peaks were extracted from the FODs for each voxel size (25 \( \mu m \), 50 \( \mu m \) and 100 \( \mu m \)) using Dipy \[12\]. Peaks were kept only if the FOD value in the peak orientation was equal or more than 20% of the FOD maximum (relative_peak_threshold=0.2). The minimum separation angle between peaks was set to 25\(^\circ\) (if multiple peaks are identify within a 25\(^\circ\) angle, the peak with the highest FOD value is kept).

2.4 Tensor-based metrics

The diffusion tensors \[5\] and the corresponding fractional anisotropy (FA) and mean diffusivity (MD) maps were computed using the re-weighted least squares method implemented in MRtrix3 \[28\]. They were estimated using the full noiseless DW-MRI signal \[20\].

3 Results

3.1 Compartmental volume fraction

Figure 2 shows a cut section of the estimated ground-truth map of DiSCo1 for the three resolutions. The extra-axonal space is shown in the first row, which also contains the free water outside the main phantom corpus. The maps show the close relationship between the intra-axonal and the myelin fraction (bottom row). For the highest resolution, the combined volume fraction of these compartments is about 52% in the highly dense areas near the main center area of the phantom. Such value may result in a less hindered extra-axonal compartment compared to that expected in real tissue. In the lower resolution, this value can be even smaller since the partial volume is present in most of the voxels.
**Fig. 2.** Ground truth volume fraction map of DiSCo1 of the extra-axonal compartment (top), intra-strand compartment (middle) and myelin layer compartment (bottom). The voxel size of the image voxel size was set to 25 \( \mu m \) (left), 50 \( \mu m \) (center) and 100 \( \mu m \) (right) isotropic.

Figure 3 shows the histogram of the volume fractions on the three phantoms and for the three resolutions. The effect of the partial volume in the compartments’ volume fractions is noticeable especially starting from 50\( \mu m \) isotropic resolution.

**Fig. 3.** Histograms of the fraction of inner strand fraction for the phantoms DiSCo1 (red), DiSCo2 (green) and DiSCo3 (blue). The voxel size of the image voxel size was set to 25 \( \mu m \) (left), 50 \( \mu m \) (center) and 100 \( \mu m \) (right) isotropic.
The Microstructure of the DiSCo dataset

3.2 FODs and number of streamlines as a function of resolution

The top-row images of Figure 5 show the voxel-wise count of the number of strands for a section of the DiSCo1 phantom. As expected, the number of fibers is higher as the resolution decreases. At the highest resolution the maximum number of strands in a single voxel is 82 and the maximum number of bundles is 5. Conversely, at the lowest resolution the maximum number of strands in a single voxel is 1136, and 18 is the maximum number of different bundles. At the highest resolution the voxel-wise mean diameter ranges from 1.3 \( \mu m \) to 4.5 \( \mu m \), centered at 2.25 \( \mu m \), which is comparable to the range of values at the other two resolutions.

The ground truth orientations and number of peaks of a cross section of DiSCo1 is shown in Figure 4 for the various resolutions. Bundles close to the ROIs are notably more homogeneous than those in crossing areas, which can also be noted in the FA maps in Figure 6. The number of peaks in a single voxel is shown in the second row; notably, some highly dense voxels contained a total of 8 peaks in the FOD beyond the set threshold (see Methods section).

![Fig. 4. Ground truth fibre orientation distribution functions (top) and corresponding peaks (bottom). The peaks are overlaid onto the peak count map. The voxel size of the image voxel size was set to 25 \( \mu m \) (left), 50 \( \mu m \) (center) and 100 \( \mu m \) (right) isotropic.](image)

3.3 MD and FA maps

The Diffusion Tensors (DT) derived maps are shown in Figure 6. In the top row, the resulting DT maps are shown. The effect of partial volume in the lowest resolution is particularly evident in the DT maps, where single bundles near the ROIs may look fully anisotropic and thus have higher FA (as shown in the second row). The mean diffusivity is shown in the bottom row. From these maps, it is
possible to observe that in the correspondence of the crossing area, the mean diffusivity is still remarkably low and homogeneous despite having a high extraxonal volume fractions and tortuous structure.

4 Discussion and conclusion

We presented quantitative maps of the microstructural properties representative of the DiSCo phantoms. These maps show the complexity achieved in the three main computed resolutions and provide a novel and multiplex microstructural environment for testing and validating connectomics and microstructural techniques. For instance, besides the context of connectomics analysis, which was the focus of the DiSCo 2021 Challenge, these phantoms can be used for validating dispersion based techniques [32], multi-tensor approaches [23, 25], axons diameter mapping [2], acquisition strategies for tractography [24], and tractogram filtering methods [10]. Secondly, these phantoms can be used as well to test or train DWI-based super-resolution approaches [3] given the availability of the three different resolutions presented here. However, from our experiments, we noted that the fidelity to the microstructure at the lowest resolution might be too poor and suffer from excessive partial volume effects. Another important
factor to consider is the availability of two additional phantoms which can be used as test and validation datasets as classically needed for machine learning approaches. We verified in our experiments that the framework can create distinct connectomes while preserving the microstructural coherence, like the volume preservation, and achieving diffusion characteristics as those expected in real tissue. The end-to-end construction and simulation of each phantom was achievable in about one week, of which the substrate optimization procedure took about 5 to 6 days to complete. Finally, given the mesh information and the capability of handling the simulation independently for each of the three compartments, in the near future, we expect to be able to enhance the phantoms realism by including the transverse relaxation effects for each compartment individually. This will provide, for instance, an additional signal contrast to myelin and can be helpful for validating the biophysical modeling of the microstructural, including simulation-assisted machine learning approaches to it [22], and for the validation of methods that jointly use diffusion and relaxation information to detect and characterize pathology [11].

To summarize, we have shown an overview of the microstructural properties of the DiSCo dataset that are part of the MICCAI 2021 DiSCo Challenge. All of the
computed maps, mesh information, and DWIs are to be made available publicly after the challenge event. We believe that these maps will boost the validation of connectomics and microstructure modeling. In addition, the phantoms can be reused to simulate more advanced protocols and even add new sources of contrast by tailoring the substrates and the biophysical properties to the specific research needs.

Acknowledgments

We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, 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). 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 MMIN-CARAV Inria associate team program between Empenn (Inria Rennes Bretagne Atlantique) and LTS5 (EPFL) started in 2019. Raphaël Truffet’s PhD is partly funded by ENS Rennes.

References


20. Rafael-Patino, J., Girard, G., Truffet, R., Pizzolato, M., Caruyer, E., Thiran, J.P.: The Diffusion-Simulated Connectivity (DiSCo) Dataset. Data in Brief (July 2021)