

# TOWARDS A DIFFUSION IMAGE PROCESSING VALIDATION AND ACCURACY PREDICTION FRAMEWORK

Francesca Pizzorni Ferrarese<sup>1</sup>, Alessandro Daducci<sup>2</sup>, Meritxell Bach Cuadra<sup>2</sup>,  
Alia Lemkaddem<sup>2</sup>, Cristina Granziera<sup>3</sup>, Jean-Philippe Thiran<sup>2</sup>  
and Gloria Menegaz<sup>1</sup>

<sup>1</sup>Department of Computer Science, University of Verona, Italy

<sup>2</sup>Signal Processing Laboratory(LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

<sup>3</sup>Brain Mind Institute (BMI), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

## ABSTRACT

Validation is the main bottleneck preventing the adoption of many medical image processing algorithms in the clinical practice. In the classical approach, *a-posteriori* analysis is performed based on some objective metrics. In this work, a different approach based on Petri Nets (PN) is proposed. The basic idea consists in predicting the accuracy that will result from a given processing based on the characterization of the sources of inaccuracy of the system. Here we propose a proof of concept in the scenario of a diffusion imaging analysis pipeline. A PN is built after the detection of the possible sources of inaccuracy. By integrating the first qualitative insights based on the PN with quantitative measures, it is possible to optimize the PN itself, to predict the inaccuracy of the system in a different setting. Results show that the proposed model provides a good prediction performance and suggests the optimal processing approach.

**Index Terms**— validation, diffusion MRI, segmentation, registration

## 1. INTRODUCTION

In medical image processing, validation is required in order to transfer the scientific research outcomes to clinically usable solutions. Sources of uncertainties are numerous rising from biological variability and image acquisition and processing methods [1]. From the image modality point of view, Magnetic Resonance Imaging (MRI) is a rapidly expanding field and a widely used medical imaging modality, since it is capable of probing functional activity as well as tissue morphology in the brain. Data analysis in MRI is sophisticated and can be thought of as a “pipeline” of closely connected processing and modeling steps.

In particular, tremendous progress in diffusion MRI technology during the past decade has enabled high-quality Diffusion Weighted (DWI) and Diffusion Tensor Imaging (DTI) of the brain in clinically feasible scanning times, for use in

routine diagnostic evaluation and in clinical research. Moreover, continued improvements in the technology of diffusion MRI are superseding the current state-of-the-art. In particular the main progresses are due to the new mathematical models such diffusion spectrum imaging (DSI) [2], which overcome the shortcomings of the DTI for representing complex white matter architectures. However, while some validation issues have been studied for DTI [3], only little work has been carried on for DSI so far [4]. In fact, its effective contribution is still to be highlighted in a quantitative and exhaustive manner, especially in terms of robustness and reproducibility of the method. Therefore validation studies are required in order to fill the gap between the scientific research and the clinical practice to make advanced technique, as the DSI, usable in medical protocols.

In the classical approach, validation is conceived as an *a-posteriori* analysis performed on the basis of some predefined objective metrics. In this paper, a novel approach following a different perspective is proposed: instead of evaluating *a-posteriori* the effectiveness of a given algorithm, the aim here is to identify the critical steps that are involved in the complete imaging chain and to provide a prediction of the performance with respect to the feature of interest. Previous works have been devoted to the analysis of simple clinical scenarios, that were only accounting of a single processing step, namely image segmentation [5] and registration [6]. In this paper we make a step forward by applying the validation framework developed to a complex DSI processing pipeline, the Connectome Mapping ToolKit, or CMTK [7], developed in our group. In order to get a local and global overview of the outcome of the processing pipeline, and to compare them, we have first analyzed the single processing modules applying the classical evaluation approach, then we have focused on the final pipeline results, according to our new validation framework.

This paper is organized as follows: in Section 2, we overview the problem investigated and the data analyzed, and

we introduce our PN-based framework. In Section 3, the performance of the PN is assessed and compared to the classical approach. In Section 4 we overview the work carried on and the promising results obtained.

## 2. MATERIALS AND METHODS

### 2.1. Data Acquisition

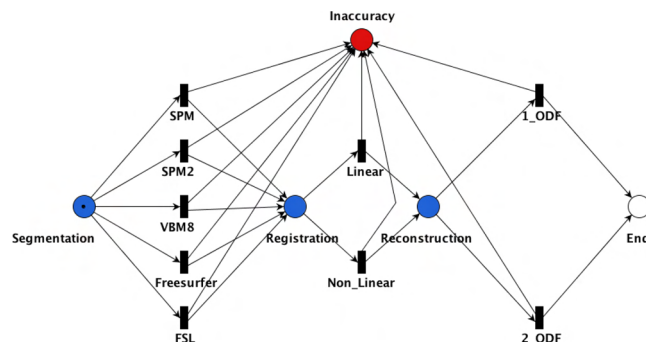
12 healthy subjects (Age:  $57.2 \pm 14.5$ ; female:male=7:5), in addition to T1- and T2-weighted acquisitions, underwent DSI scans twice within a 1 month interval ( $\pm 1$  week). All measurements were performed at 3T (Trio a Tim System, Siemens, Erlangen, Germany) using a 32 channel head coil (TR/TE=6600/138 ms, FoV=212x212 mm, 34 slices, 2.2x2.2x3 mm resolution, 258 diffusion directions,  $b=8000$  s/mm<sup>2</sup>).

### 2.2. Connectome Mapping ToolKit

The CMTK is a Diffusion MR image processing pipeline that consists of a four-step process [7]: (1) segmentation of white matter (WM) and gray matter (GM) and subdivision of the cortex into anatomical regions, (2) morphological-to-diffusion images registration (3) orientation distribution function (ODF) reconstruction and white matter fiber tracking, i.e. tractography, and (4) network construction. We have analyzed the aforementioned steps according to both the classical and the PN-based methods. In order to do so, first we have carried on a local analysis, by evaluating the accuracy of each single module taken individually in the classical way. Then, we have studied the accuracy of the pipeline as a whole, by evaluating the correlation between connectivity measurements obtained for the 2 time points available. In what follows we briefly overview the analysis approach for each module.

**Segmentation** We have mainly focused our attention on the segmentation of the WM, since it represents the region of interest (ROI) for the subsequent tractography. We have proceeded by applying, with same initial conditions, different state-of-the-art segmentation softwares (SPM8 [8], SPM8-2nd version [9], VBM8 [10], Freesurfer [11], FSL [12]) and averaging their results, in order to obtain a sort of “reference” segmentation. The MatLab-based methods, i.e. SPM8, SPM8-2nd version and VBM8, could bias the definition of a white matter reference mask from a STAPLE-like approach [13]. Therefore, in order to correctly weight the information provided by all the different methods, we have opted for the voting policy [14].

**Registration** The registration step, performed with FSL, is devoted to the alignment of T1-weighted images with the DSI images. To do so, a 2-step process is followed, in order to reduce the geometric distortions arising from the acquisition method: (1) T1- to T2-weighted registration (2) T2-weighted to DSI (b0 image) registration. While using for the first step a linear (affine) registration, we have processed the second



**Fig. 1.** PN focusing on the critical aspects that can affect the overall accuracy of the DSI processing pipeline

registration step in both a linear and a non-linear fashion, to highlight the differences between the two approaches.

**ODF Reconstruction/Tractography** The ODF reconstruction/sharpening issue is being more and more investigated during the last few years, since it has been highlighted the sensitivity of the tractography to the choice of reconstruction parameters. Following the method described in [4] we have reconstructed the data by two different approaches, namely using 1 or 2 ODFs respectively. Details are out of the scope of this paper, and they can be found in the above-mentioned reference.

### 2.3. Petri Net paradigm

A Petri Net is a tuple  $(P, T, F)$  where  $P$  is a finite set of places,  $T$  is a finite set of transitions ( $P \cap T = \emptyset$ ) and  $F \subseteq (P \times T) \cup (T \times P)$  is a set of arcs (flow relation). From the graphical point of view, a classical PN is composed of three primitive concepts: places, transitions and tokens (Fig. 1).

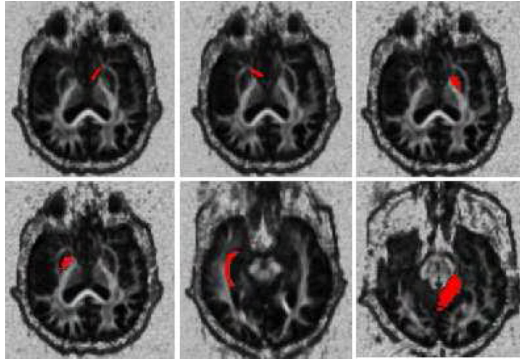
*Processing units* within the imaging chain correspond to *places*, *image features* and *processing methods* correspond to *transitions*. In consequence, every path in the net represents a possible combination of parameters that are responsible for the inaccuracy of the results.

In this phase, all the critical aspects in the processing workflow that potentially lead to inaccuracies were considered and the overall process was formalized using the PN paradigm (Fig. 1). The inaccuracy level is encoded in the number of tokens collected in a specific state/counter that the process reaches at the end of the complete workflow (red-place). As mentioned above, the critical points that could affect the final results have been assigned to each processing step (blue-places).

## 3. RESULTS

### 3.1. Classical approach: local analysis

**Segmentation** We have compared each segmentation to the reference by measuring commonly used overlap agree-

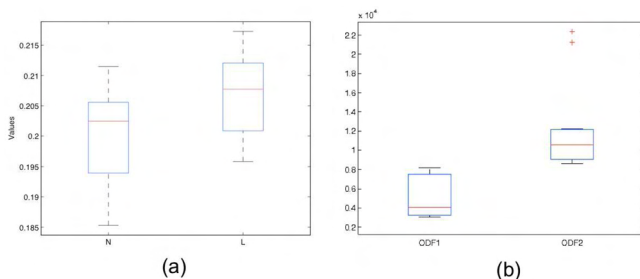


**Fig. 2.** Critical regions according to segmentation variability in proximity of (from above-left) accumbens nucleus sx, accumbens nucleus dx, pallidum sx, pallidum dx, right-inferior-lateral-ventricle and cerebellum cortex

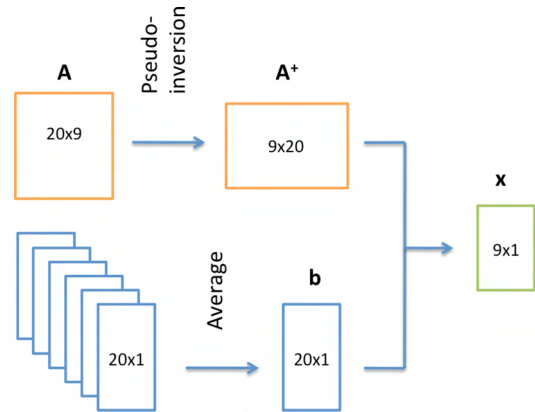
ment, i.e. Dice and Jaccard indexes, and overlap error measures, i.e. false negative and false positive errors. Moreover in order to localize the most critical anatomical regions in terms of variability of the segmentation, we have applied the labeling obtained in Freesurfer to both the segmentations and the reference and we have detected the regions that were more prone to results variability. The main segmentation issues come in the central part of the brain and in Fig. 2 we present the 6 regions that shows the higher variability.

**Registration** We have measured the Edge Overlap Error between the T2-weighted registered image and the DSI (b0 image). In Fig. 3 (a) we can highlight an improvement of the registration outcome going from the linear to the deformable registration.

**ODF Reconstruction/Tractography** We have evaluated the relative difference by measuring the number of fibers reconstructed in the whole brain and also comparing some specific tracts that have been reconstructed defining several ROIs along different slices, as described in [15]. In Fig. 3 (b) we summarize in the box-plot the results obtained in terms of fibers reconstructed, by comparing their distributions in the case of 1 and 2 ODFs.



**Fig. 3.** Box-plots showing a) the improvement in terms of registration accuracy while using the non-linear (N) registration instead of the linear (L) one ( $p < 0,05$ ), b) the difference in terms of fibers reconstructed using 1 or 2 ODFs respectively ( $p < 0,01$ ).



**Fig. 4.** 2-step process for obtaining matrix  $\mathbf{x}$  from the pseudo-inversion of  $\mathbf{A}$  and the measured errors in SA.

### 3.2. PN-based approach: global analysis

Six out of the twelve datasets (SA) were used for the calibration of the PN, and the remaining ones (SB) for validation. Images in SA were processed using different combinations of the parameters listed in Section 2. Then, the correlation between the prediction provided by the PN and the measures obtained *a-posteriori* was assessed. We have based our study on the analysis of the correlation between the connectivity matrices calculated for the 2 scans available per subject in terms of normalized fiber density (number of fibers/ROIs volume). In fact, in healthy subjects of all ages we do not expect any changes in fiber density during a 1-month period. Thus, the discrepancies between connectivity matrices are attributed to variations arising from the processing pipeline itself or to the noise in the acquisition.

In order to calibrate the transitions of the net, i.e. assign a defined weight to each arc that goes from a specific transition to the inaccuracy counter, a binary  $20 \times 9$  matrix  $\mathbf{A}$  (where 9 stands for the pipeline parameters analyzed and 20 is the number of complete processing for each dataset, i.e. the number of possible paths along the PN) has been built. For each dataset in SA, 240 different processing have been carried on and we have collected the respective error measures by comparing the connectivity matrices built from the 2 scans, obtaining a unique representative 20-elements error vector  $\mathbf{b}$  for SA. Matrix  $\mathbf{A}$  has been then pseudo-inverted and multiplied by  $\mathbf{b}$ , in order to obtain  $\mathbf{x}$ , that is the matrix of the weights to be applied on the PN. The process for obtaining  $\mathbf{b}$  from the errors measured is summarized in Fig. 4 To evaluate the robustness of the proposed approach the PN weights have also been estimated from SB. The Pearson's correlation has been then calculated between the measured and predicted errors for both SA and SB in order to validate the whole process: the correlation coefficient measured for the 12 datasets is over 0,95. From this results we can conclude that we can accurately predict the outcome of DSI analysis pipeline ap-

**Table 1.** Weights obtained from SA and SB.

		SA	SB
Segmentation	SPM	0,016850	0,016348
	SPM2	0,013045	0,012763
	VBM8	0,002373	0,002783
	Freesurfer	0,007164	0,007732
	FSL	0,010375	0,012795
Registration	linear	0,033262	0,034579
	non-linear	0,098432	0,108509
Reconstruction	1ODF	0,078562	0,080493
	2ODF	0,056279	0,057342

plied on different data and settings, which is one of the main advantages of the framework developed.

In Table 1 the comparison between the weights obtained from SA and SB is presented. There is quite a wide variability between segmentation methods in terms of how they affect the reproducibility of the connectivity analysis. A surprising result comes from the registration, where we can see that even if the non-linear registration appears more accurate from the local analysis, while investigating its contribution in terms of correlation between matrices, we can see a higher impact in terms of inaccuracy compared to the linear registration, probably due to the registration issues in proximity of the WM-GM interface. For what concerns the reconstruction of the ODF, we can conclude from this preliminary study that this is a field that has to be investigated more, in order to define the optimal reconstruction strategy that could lead to more robust tractography.

#### 4. CONCLUSIONS

In this paper we have applied a new an alternative concept of accuracy assessment in medical image processing to the complex scenario of a complete DSI processing pipeline. Based on the analysis of the critical aspects in the workflow, it was possible to estimate the overall system (in)accuracy and predict the outcome of the processing chain. It has been demonstrated that is possible to correlate the prediction with the quantitative measures. This could potentially have a great impact on the clinical practice as it would enable to predict the inaccuracy of the process *a-priori* or, viceversa, to tune *ad-hoc* the different steps of the considered process.

#### 5. REFERENCES

- [1] S. Warfield P. Jannin, E. Krupinski, "Validation in medical image processing," *IEEE Transactions on Medical Imaging*, 2006.
- [2] V.J. Wedeen, T.G. Reese, D.S. Tuch, and et al., "Mapping fiber orientation spectra in cerebral white matter with Fourier-transform diffusion MRI," *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 8, pp. 82, 2000.
- [3] J. Dauguet, S. Peled, V. Berezovskii, and et al., "Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain," *Neuroimage*, vol. 530-538, no. 37, 2007.
- [4] A. Daducci, X. Gigandet, G. Krueger, and et al., "On the reproducibility and anatomical correspondence of dsi tractography," in *HBM*, 2010.
- [5] F. Pizzorni Ferrarese, G. Menegaz, A. Le Troter, and B. Fertil, "A new paradigm for geometric accuracy prediction in medical image segmentation," in *MIC-CAI 2009 Workshop on Geometric Accuracy in Image Guided Interventions*, 2009.
- [6] F. Pizzorni Ferrarese, F. Simonetti, R. Foroni, and G. Menegaz, "Registration accuracy assessment on noisy neuroimages," in *Proc. of IEEE International Conference on ITAB 2010*, 2010.
- [7] "Lausanne neuroimaging tools for connectomics," <http://www.cmtk.org>.
- [8] J. Ashburner and K.J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, pp. 839–851, 2005.
- [9] J. Ashburner, "Computational anatomy with the SPM software," *Magnetic Resonance Imaging*, 2009.
- [10] M. Wilke C. Gaser, M. Altaye and S.K. Holland, "Unified segmentation without tissue priors," *NeuroImage*, vol. 36, pp. S68, 2007.
- [11] X. Han and B. Fischl, "Atlas renormalization for improved brain MR image segmentation across scanner platforms," *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 479–486, 2007.
- [12] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 1, pp. 45–57, 2001.
- [13] S.K. Warfield, K.H. Zou, and W.M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 7, pp. 903–921, 2004.
- [14] Xiang Li, Ben Aldridge, Jonathan Rees, and Robert Fisher, "Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation," in *Proc. Medical Image Understanding and Analysis*, 2010, vol. 1.
- [15] M. Catani and M. Thiebaut de Schotten, "A diffusion tensor imaging tractography atlas for virtual in vivo dissections," *Cortex*, vol. 44, no. 1105-1132, 2008.