Manual to sc-dmri-myelopathy scripts

All implemented code is available at: https://github.com/renelabounek/sc-dmri-myelopathy

Will be published as a part of the scientific journal manuscript:

Labounek R, Valošek J, Horák T, Svátková A, Bednařík P, Vojtíšek L, Horáková M, Nestrašil I, Lenglet C, Cohen-Adad J, Bednařík J and Hluštík P. HARDI-ZOOMit protocol improves specificity to microstructural changes in presymptomatic myelopathy. *Scientific Reports*, **10**, 17529. https://doi.org/10.1038/s41598-020-70297-3

Manual has been written by *Rene Labounek* and *Jan Valosek*.

Content

1 Prerequisites

To successfully execute the in-house implemented codes, several prerequisites have to be fulfilled on your computer machine:

1. Installed bash programming and scripting interface (more details at: https://www.gnu.org/software/bash/)
2. Installed Spinal Cord Toolbox (SCT) software library (available at: https://github.com/neuropoly/spinalcordtoolbox) and set its path in your default bash profile; The SCT version "git-master-d50102be4e00ead24651e5bfe701e645965961cb" was the latest version which was tested for the in-house implemented code compatibility.
3. Installed FSL (FMRIB Software Library, Oxford, UK; available at: https://fsl.fmrib.ox.ac.uk/fsldownloads_registration) software library and set its path in your default bash profile; The FSL version "5.0.11" was the latest version which was tested for the in-house implemented code compatibility.
4. Installed and licensed MATLAB (MathWorks, Natick, USA; available at: https://www.mathworks.com/downloads/) programming environment and set its path in your default bash profile; The MATLAB version "R2017b" was the latest version which was tested for the in-house implemented code compatibility.
5. Installed ANTs (Advanced Normalization Tools, available at: http://stnava.github.io/ANTs/) software library and set its path in your default bash profile. The ANTs version "2.1.0" was the latest version which was tested for in-house implemented code compatibility.
6. We would suggest and recommend to use a computer machine with at minimum 6 CPUs (Central Processing Units) and 32GB as a minimal total RAM memory. GPU (Graphics Processor Unit) is not necessary to have installed in your computer for the successful execution of the in-house implemented codes.
7. Matlab toolboxes freely available at mathworks.com (i.e. mi, nmi): https://www.mathworks.com/matlabcentral/fileexchange/13289-fast-mutual-information-of-two-images-or-signals https://www.mathworks.com/matlabcentral/fileexchange/29047-normalized-mutual-information
8. Statistical Parametric Mapping version 12 (SPM12) available at: https://www.fil.ion.ucl.ac.uk/spm/software/
9. Licensed MATLAB toolboxes: Statistics and Machine Learning Toolbox
10. HTCondor. If you need to make working the bedpostx_condor bash script. It is available at: https://research.cs.wisc.edu/htcondor/

2 Data

Acquired raw data are available here: https://hdl.handle.net/20.500.12618/0000-5c13d342-4798-41d9-8d2a-bf750ab79fdb

If you keep the data folder structure as it is, all bash or matlab scripts should work properly.

3 Bash_scripts

3.1 Anatomical pipeline, dMRI data preprocessing, dtifit and warping dtifit results into anatomical space

Analysis of anatomical (i.e. T_2*-w axial (transversal) - T2TRA and T_2-w sagittal - T2SAG) and diffusion (i.e. HARDI-ZOOMit and RESOLVE) data is done by *run_analysis.sh* bash script.

Anatomical analysis contains:

1) Bias-field correction of original T2TRA and T2SAG images using N4BiasFieldCorrection tool (part of ANTS) and thresholding of low intensity values.

2) Slice-by-slice correction of slicewise zig-zag artifact of T2TRA image caused by interleaved mode (could be switched off inside the script).

3) Resampling and cropping of T2SAG image (for fitting with resolution of T2TRA image)

4) Spinal cord segmentation of T2SAG either by *sct_deepseg_sc* or *sct_propseg* function (could be set inside script); *sct_deepseg_sc* is recommended.

5) Vertebrae labeling of T2SAG spinal cord segmentation using *sct_label_vertebrae* function.

6) Registration of T2SAG image to template (T2SAG spinal cord segmentation is necessary); if T2SAG segmentation is wrong, this step is skipped.

7) Spinal cord segmentation of T2TRA image either by *sct_propseg* (using T2SAG spinal cord segmentation as init-centerline or manually created init-centerline) or by *sct_deepseg_sc* (could be set inside script); *sct_deepseg_sc* is recommended.

8) Segmentation of T2TRA gray matter either by *sct_segment_graymatter* or by *sct_deepseg_gm*; *sct_deepseg_gm* is recommended.

9) Vertebrae labeling of T2TRA spinal cord segmentation and correction of interpolation artifact at borders of adjacent spine segments.

10) Registration of template and atlas (PAM50) to T2TRA image (T2TRA space) using information from previous template to T2SAG registration.

11) Improving of template registration to T2TRA space by accounting for WM/GM shape (requires correct WM and GM segmentation of T2TRA image).

Diffusion analysis contains:

1) Diffusion preprocessing (merging of AP and PA b0 images) and possible motion correction by *sct_dmri_moco* function

2) Correction of geometrical distortions and eddy current artifacts using FSL's *topup* and *eddy* functions (with whole FOV or with manually segmented mask of SC from topup_mean image))

3) Estimation of diffusion tensor (DTI) model using FSL's *dtifit* function

4) Registration between T2TRA and DIFF spaces - 2-step registration:

    a) T2TRA spinal cord segmentation (T2TRA_thr_bias_corr_seg.nii.gz) is registered to DWI space

    b) Spinal cord segmentation of DWI mean image (registered T2TRA spinal cord segmentation is used as init-centerline)

    c) Mean b0 image (b0_mean) is registered to T2TRA image. Segmentations of SC from both images (T2TRA and DWI) are used for improving of final registration. Warping field from previous step is used.

5) Vertebrae labeling of DIFF space

6) Warping of dtifit metrics (i.e. FA, MD, V1, V2, V3, L1, L2, L3) to T2TRA space.

Analysis is run by:

run_analysis.sh <DATA_FOLDER> <SUB_ID> <SEQ_ORDER>

where:

DATA_FOLDER - directory containing subject(s) data

SUB_ID - subject ID

SEQ_ORDER - information which diffusion procol(s) was/were required (are analysed) (e.g. 11001 means ZOOMit_inter, ZOOMit_notiterp and RESOLVE protocols were acquired and are analyzed)

Diffusion is called automatically from *run_analyis.sh* script and is performed by *diff_analysis.sh* script (this script can be also run/rerun manually).

3.2 bedpostx and warping bedpostx results into anatomical space

Estimation of Ball-and-Sticks model using FSL´s *bedpostx* function (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT/UserGuide#BEDPOSTX) is done using *diff_analysis_bedpostx.sh* script which allows to run voxel-wise estimation of this model in parallel mode using *HTCondor* parallelization tool (https://research.cs.wisc.edu/htcondor/). This script also performs warping of Ball-and-Sticks's metrics (f1, f2 and d) to T2TRA space using warping field estimated by previous *diff_analysis.sh* script. Script is called similar as *run_analysis.sh* script:

diff_analysis_bedpostx.sh <DATA_FOLDER> <SUB_ID> <SEQ_ORDER>

where:

DATA_FOLDER - directory containing subject(s) data

SUB_ID - subject ID

SEQ_ORDER - information which diffusion procol(s) was/were required (are analysed) (e.g. 11001 means ZOOMit_inter, ZOOMit_notiterp and RESOLVE protocols were acquired and are analyzed)

NB - it is necessary to run this script from "master condor machine", i.e., machine which is set for distribution of *HTCondor´s* jobs.

USEFUL COMMANDs for HTCondor:

- condor_q
- condor_status

4 Matlab_scripts

When all bash_scripts were successfully executed, matlab scripts evaluating results can be executed. First script which you should execute is the extract_descriptive_statistics.m script after proper setting of a few input variables.

4.1 extract_descriptive_statistics.m

The script extract_descriptive_statistics.m extracts descriptive statistics parameter from diffusion MRI microstructural maps warped into the T2TRA space, stores the results into the predefined dmri_comparison.mat file in the save_path folder, and visualize the categorical scatterplot results for various examined variables.

Before executing the script, you should properly set following variables of the script:

data_folder … path to the folder at the disc where your input data and output results are stored. Default script value is: data_folder='/home/user/data'. The script expects to find inside the data folder the subject specific folders, e.g. S001, S002, S003, etc. Each subject specific folder S0*/ contains three subfolders, i.e. Anat/, Diffusion/ and Results/ respectively.

save_path … path to the folder where results will be stored, i.e. .mat files and .png figures. Default script value is: data_folder='/home/user/results'.

read_data … binary value deciding whether results would be extracted from Result nifti files. When it is equal to 1, the results are extracted and stored into the dmri_comparison.mat file. If it is equal to 0, this step is skipped.

plot_results … binary value deciding whether results would be visualized or not and whether results extended about Wilcoxon rank-sum test p-value tables will be stored in the file dmri_comparison_pvaltable.mat or not. When it is equal to 1, the results are visualized and saved as .png figures into the folder save_path and extended results are stored into the dmri_comparison_pvaltable.mat file. If it is equal to 0, this step is skipped.

If you would like to define your own subject list, you need to redefine the variable **subject** inside the script based on the comments close to the place in the code where the variable is defined.

4.2 extract_heuristic_parameters_wm.m

If you have successfully executed full extract_descriptive_statistics.m script and the extended result file dmri_comparison_pvaltable.mat is stored at your harddrive, you can continue and execute the script extract_heuristic_parameters_wm.m.

Define well the variable **save_path** inside the script at two places. Same value at both places is recommended for a stable run. After that, the script should work and extend/rewrite the file dmri_comparison_pvaltable.mat about heuristic parameter results extracted from the C3-C6 white matter. The script also generates the Figure 3 of the Labounek et al. (2020) Scientific Reports and categorical scatterplots for the heuristic parameters extracted from white matter.

4.3 extract_heuristic_parameters_gm.m

If you have successfully executed full extract_descriptive_statistics.m and extract_heuristic_parameters_wm.m scripts and the extended result file dmri_comparison_pvaltable.mat is stored at your harddrive, you can continue and execute the script extract_heuristic_parameters_gm.m.

Define well the variable **save_path** inside the script at two places. Same value at both places is recommended for a stable run. After that, the script should work and extend/rewrite the file dmri_comparison_pvaltable.mat about heuristic parameter results extracted from the C3-C6

gray matter. The script also generates the FigS. 3 of the Labounek et al. (2020) Scientific Reports and categorical scatterplots for the heuristic parameters extracted from gray matter.

4.4 crosscorrelation_susceptibility_and_ancova.m

If you have successfully executed full extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m scripts and the extended result file dmri_comparison_pvaltable.mat is stored at your harddrive, you can execute the crosscorrelation_susceptibility_and_ancova.m script.

I will estimate cross-correlation over dMRI metrics with significant differences between patient and control groups and measured susceptibility artifact effects, and will estimate the post-hoc ancova of between-group differences where age is used as the confounding variable.

You need to properly set following variables inside the script:

save_path … path to the folder where results are stored. Default script value is: data_folder='/home/user/results'.

demographic_file … full path to the .xlsx table with subjects' demographics (i.e. the table is in the root folder after download and unzipping of the shared dataset at: https://hdl.handle.net/20.500.12618/0000-5c13d342-4798-41d9-8d2a-bf750ab79fdb).

var_indxs … positions of variables with significant differences between patient and control groups or with measured susceptibility artifact effects in the *table_ZOOMit_Int*, *table_ZOOMit_NotInt* or *table_RESOLVE* tables. The precise position of each tested variable in the table you can find in scripts extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m there where the table variables are defined.

table _names … If you have updated positions in the var_indxs variable (compared to original values), you should also update the list of variable names / shortcuts (to get readable and understandable result outputs in visualizations).

4.5 stepwisereg_and_kmeans.m

If you have successfully executed full extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m scripts and the extended result file dmri_comparison_pvaltable.mat is stored at your harddrive, you can execute the stepwisereg_and_kmeans.m script.

You need to properly set following variables inside the script:

save_path … path to the folder where results are stored. Default script value is: data_folder='/home/user/results'.

demographic_file … full path to the .xlsx table with subjects' demographics (i.e. the table is in the root folder after download and unzipping of the shared dataset at: https://hdl.handle.net/20.500.12618/0000-5c13d342-4798-41d9-8d2a-bf750ab79fdb).

var_indxs … positions of variables with significant differences between patient and control groups in the *table_ZOOMit_Int*, *table_ZOOMit_NotInt* or *table_RESOLVE* tables. The precise position of each tested variable in the table you can find in scripts extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m there where the table variables are defined.

table _names … If you have updated positions in the var_indxs variable (compared to original values), you should also update the list of variable names / shortcuts (to get readable and understandable result outputs in visualizations).

For the HARDI-ZOOMit Int protocol, the K-means clustering provided the highest sensitivity+specificity measure for the 2nd most common clustering result, not for the first. The first most common clustering result is neglected with the row number 100 in the code:

cls_count(cls_count==max(cls_count)) = 1;

If you comment it (i.e. add % in the line beginning) you can also check the first most common result. Similar code rows are written and commented also for the DTI-RESOLVE protocol (row number 287) and HARDI-ZOOMit NonInterp. Protocol (row number 490) at appropriate code rows. With their commenting/uncommenting you can change how the results change.

4.6 offresonance_effects.m

If you have successfully executed full extract_descriptive_statistics.m script, you can execute the offresonance_effects.m script.

You need to properly set following variable inside the script:

save_path … path to the folder where results are stored. Default script value is: data_folder='/home/user/results'.

The script estimates group means and standard deviations of off-resonance effects from C3-C6 ROI (i.e. variable offres_c3c6), from C3 ROI (i.e. variable offres_c3) and from C5-C6 ROI (i.e. variable offres_c5c6). The first row in each matrix are mean values, the second row consists of standard deviations. First column corresponds to HARDI-ZOOMit Interp protocol, second column to HARDI-ZOOMit NonInterp protocol and third row to DTI-RESOLVE NonInterp protocol.

4.7 build_table1.m

If you have successfully executed full extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m scripts and the extended result file dmri_comparison_pvaltable.mat is stored at your harddrive, you can execute the build_table1.m script.

You need to properly set following variables inside the script:

save_path … path to the folder where results are stored. Default script value is: data_folder='/home/user/results'.

subnum … number of used scan sessions

The script build Table 1 presented in Labounek et al. (2020) Scientific Reports.

4.8 similarity_measurement_figure2d.m

This script estimates non-normalized mutual information between microstructural maps and white/gray matter structures segmented from medic T_2^*w scans. The script provides the same visual output as shown in the Figure 2d in Labounek et al. (2020) Scientific Reports.

Before executing the script, you should properly set following variables of the script:

data_folder … path to the folder at the disc where your input data and output results are stored. Default script value is: data_folder='/home/user/data'. The script expects to find inside the data folder the subject specific folders, e.g. S001, S002, S003, etc. Each subject specific folder S0*/ contains three subfolders, i.e. Anat/, Diffusion/ and Results/ respectively.

save_path … path to the folder where results will be stored, i.e. .mat files and .png figures. Default script value is: data_folder='/home/user/results'.

read_data … binary value deciding whether results would be extracted from Result nifti files. When it is equal to 1, the results are stored into the dmri_similarity_256_bins_wmgm.mat file. If it is equal to 0, this step is skipped.

plot_results … binary value deciding whether results would be visualized or not. When it is equal to 1, the results are visualized and saved as .png figures into the folder save_path. If it is equal to 0, this step is skipped.

If you would like to define your own subject list, you need to redefine the variable **subject** inside the script based on the comments close to the place in the code where the variable is defined.

4.9 boxplot_visualization_figure5.m

If you have successfully executed full extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m scripts and the extended result file dmri_comparison_pvaltable.mat is stored at your harddrive, you can execute the stepwisereg_and_kmeans.m script.

You need to properly set following variables inside the script:

save_path … path to the folder where results are stored. Default script value is: data_folder='/home/user/results'.

var_indxs … positions of variables with significant differences between patient and control groups in the *table_ZOOMit_Int*, *table_ZOOMit_NotInt* or *table_RESOLVE* tables. The precise position of each tested variable in the table you can find in scripts extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m there where the table variables are defined.

table _names … If you have updated positions in the var_indxs variable (compared to original values), you should also update the list of variable names / shortcuts (to get readable and understandable result outputs in visualizations).

table_pvals_rows … row positions in table of p-values demonstrating differences between patient and control groups in the **table_pvals** variable. The precise position of each tested variable in the table you can find in scripts extract_descriptive_statistics.m, extract_heuristic_parameters_wm.m and extract_heuristic_parameters_gm.m there where the table_pvals variable is called for a specific tested variable (i.e. it is spread over various rows of all three mentioned scripts). In other words, you are looking for code rows appearing as:

```
%% VISUALIZATION OF FA STD VALUE DISTRIBUTIONS FROM GM
pCR_ZOOMint = ranksum(FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==1),FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==1));
pCP_ZOOMint = ranksum(FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==1),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==1));
pRP_ZOOMint = ranksum(FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==1),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==1));
pSC_ZOOMint = ranksum(FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==1),FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==1));
pSR_ZOOMint = ranksum(FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==1),FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==1));
pSP_ZOOMint = ranksum(FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==1),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==1));
pCR_ZOOMnotint = ranksum(FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==2),FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==2));
pCP_ZOOMnotint = ranksum(FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==2),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==2));
pRP_ZOOMnotint = ranksum(FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==2),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==2));
pSC_ZOOMnotint = ranksum(FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==2),FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==2));
pSR_ZOOMnotint = ranksum(FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==2),FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==2));
pSP_ZOOMnotint = ranksum(FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==2),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==2));
pCR_RESOLVE = ranksum(FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==3),FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==3));
pCP_RESOLVE = ranksum(FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==3),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==3));
pRP_RESOLVE = ranksum(FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==3),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==3));
pSC_RESOLVE = ranksum(FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==3),FAgmstd_boxplot(subject_grp==1 & FAgmstd_grp==3));
pSR_RESOLVE = ranksum(FAgmstd_boxplot(subject_grp==3 & FAgmstd_grp==3),FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==3));
pSP_RESOLVE = ranksum(FAgmstd_boxplot(subject_grp==4 & FAgmstd_grp==3),FAgmstd_boxplot(subject_grp==2 & FAgmstd_grp==3));
table_pvals(10,:) = [pCR_ZOOMint pCP_ZOOMint pRP_ZOOMint pSC_ZOOMint pSR_ZOOMint pSP_ZOOMint …
    pCR_ZOOMnotint pCP_ZOOMnotint pRP_ZOOMnotint pSC_ZOOMnotint pSR_ZOOMnotint pSP_ZOOMnotint
…
    pCR_RESOLVE pCP_RESOLVE pRP_RESOLVE pSC_RESOLVE pSR_RESOLVE pSP_RESOLVE];
```

This example code presents that all evaluated Wilcoxon rank-sum test p-values are stored at 10th row of the matrix table_pvals for the FA standard deviation variable extracted from gray matter.

IMPORTANT!!! Unfortunately, values of var_indxs and table_pvals_rows differ for the same examined variable. Please, be careful about it to avoid some mistakes in your own analyses.

If you have set the script properly, it will generate boxplots which were presented in Figure 5 in Labounek et al. (2020) Scientific Reports.

4.10 subgroup_demography_stats.m

You need to properly set following variable inside the script:

demographic_file … full path to the .xlsx table with subjects' demographics (i.e. the table is in the root folder after download and unzipping of the shared dataset at: https://hdl.handle.net/20.500.12618/0000-5c13d342-4798-41d9-8d2a-bf750ab79fdb).

The script evaluate mean age and age standard deviation for healthy controls (ageHCstat), ADCCC patients (ageADCCCstat), mild ADCCC patients (ageADCCCMstat), severe ADCCC patients (ageADCCCSstat) and reproducibility group (ageREPRODUCIBILITYstat). In the p matrix, it also estimates p-values of age-differences between healthy controls and all other mentioned groups. The p-values are evaluated with Wilcoxon rank-sum tests (1st column) or two sample t-tests (2nd column).