

Restoring statistical validity in group analyses of motion-corrupted MRI data – an extension to relaxometry data (QUIQI II)

Analysis code

Authors: Rita Oliveira, Quentin Raynaud, Nadège Corbin, Giulia Di Domenicantonio, Martina F. Callaghan, Antoine Lutti

Laboratory for Neuroimaging Research

Lausanne University Hospital & University of Lausanne, Lausanne, Switzerland

Copyright (C) 2022 Laboratory for Neuroimaging Research

REFERENCE

'QUIQI using a QQuality Index for the analysis of Quantitative Imaging data'. Di Domenicantonio et al., Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine, 2020.

'Restoring statistical validity in group analyses of motion-corrupted MRI data', Lutti et al, Human Brain Mapping 2022.

INTRODUCTION

This package includes supporting material for the scientific article by Lutti et al. entitled 'Restoring statistical validity in group analyses of motion-corrupted MRI data – an extension to relaxometry data'. This article introduces QUIQI II, an extension of our previously proposed method called QUIQI. QUIQI accounts for the degradation of data quality due to motion in the analysis of MRI data by assigning weights to each dataset, based on a value of image quality, the Motion Degradation Index (MDI). In the context of quantitative relaxometry data, some maps can be computed from multiple raw image volumes, such as $R1$, $R2^*$, and $MTsat$. In QUIQI II, we account for the motion degradation when images are obtained from multiple acquisitions.

The purpose of this package is to allow the scientific community to cross-examine the analysis script written to implement the method and to replicate the main results obtained. This package also constitutes a template to help interested users to implement this method for their own neuroimaging studies. This code has been integrated into a customized version of the hMRI toolbox (<https://hmri-group.github.io/hMRI-toolbox/>).

The complete support package for the QUIQI II method includes:

1. A copy of the original analysis code used to compile the results presented in the original scientific publication.
2. A subset of the data used in the original publication for computation of the results. This data also includes a set of analysis results obtained by running the code described in 1. on the provided data.

The combination of 1. and 2. allows users to replicate the computation of the provided analysis results.

The material provided here only concerns part 1. of the QUIQI II support package described above – original analysis code.

The QUIQI method requires the values of a MDI for each of the datasets to be analysed. The proposed method can accommodate all types of MDI, which need to be computed separately prior to the analysis of the data. In its current form, the analysis script expects the MDI values as a separate field in a Matlab structure that contains relevant metadata (Subject_Details.mat). This can be modified according to users' preferences.

CONTENT

The provided material contains the Matlab scripts required for data analysis. The main function is **RunQUIQI.m**. The provided material also includes, in the SPM folder, two SPM12 files (version 7771) that were edited to provide the Free Energy estimates from the ReML computation as an output (spm_spm.m and spm_est_non_sphericity.m). These files are not a requirement for the analyses but may be used when change in Free Energy need to be considered. Also, note that the provided spm_spm.m script also contains an optional call to spm_reml_sc to enforce positive hyper-parameter estimates.

REQUIREMENTS

A running version of Matlab (<https://ch.mathworks.com/products/matlab.html>) and SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>) are pre-requisites for running these analyses.

Matlab Econometrics toolbox (in particular the *garch* function) is necessary to compute heteroscedasticity maps. This function is only used in the scripts if the user has the necessary toolbox, otherwise, the step is skipped.

This package may be run on the publicly available data or data provided by the user.

INSTALLATION

- Download the provided package.
- In Matlab, add to the paths your SPM installation folder and the folder that contains the provided package.
- Edit **GetParams.m** according to your working directories.
- Run the main function **RunQUIQI.m** with the input variables corresponding to the analysis of interest.

MAIN FUNCTION RunQUIQI.m

Inputs:

- a. 'Subregions': array with brain regions of interest for the analysis. Note that multiple regions can be entered into 'Subregions'. In this case, analyses will be conducted in succession for all entered regions.
 - 'p1': grey matter
 - 'p2': white matter
 - value of the regional labels defined in the neuromorphometrics atlas

- b. 'lambda': array with the desired powers of the motion degradation index used for the computation of the ReML basis functions. If 'lambda' contains multiple entries, analyses will be conducted in succession for each entry. Note that in the main scientific article variable 'lambda' is referred to as 'alpha'. Example options are:
- 0: ordinary least-square analysis
 - [0 1 2 3]: separate basis functions will be computed from the MDI, one for each value.
 - [0 1 2 3 4]: separate basis functions will be computed from the MDI, one for each value.
- c. 'DataType': array with list of type of data to be analyzed. The options are:
- 'PDT1_R2s': R2* maps computed from the PDw and T1w raw image contrasts. In the main scientific article, this variable is named 'R2* (2)'.
 - 'R2s_OLS': R2* maps computed from the PDw, T1w and MTw raw image contrasts. In the main scientific article, this variable is named 'R2* (3)'.
 - 'R1': R1 maps computed from the PDw, T1w and MTw raw image contrasts.
 - 'MT': Magnetization transfer maps computed from the PDw, T1w, and MTw raw image contrasts.
- d. 'AnalType': string with the type of analysis to be conducted. This variable can take the following values:
- 'Full': analysis of all the datasets available, identified from the Subject_details.mat object. Statistical analysis: one-sample t-test of the age-dependence of the data on age (a 2nd-order model of ageing is assumed).
 - 'Residuals': identical to 'Full' analysis but residual images are saved for subsequent analysis vs the MDI values ('MDIvsResAnalysis' function). Note that saving individual residual images increases computation time and requires extensive disk space.
 - 'Exclusion': conducts analysis after exclusion of the fraction of the cohort with the highest value of the MDI. The main purpose of this analysis is to replicate the common use of an MDI in the community and only really makes sense by setting lambda to 0 (Ordinary Least-Square analyses). The value of the fraction of the cohort excluded from analysis is given by AnalParams.Exclusion, in the GetAnalParams function
 - 'Specificity': as for a 'Full' analysis, the statistical analysis involves one-sample t-tests of the age-dependence of the data on age. However, the age assigned to each dataset is scrambled randomly across the datasets to allow for the monitoring of false positives. The number of repetitions of this analysis is set by AnalParams.NRepeats in the GetAnalParams function. Note that in order to limit the size of the provided data, the number of repetitions was set to 10 to compute the provided results (1000 in the scientific article).

Another two types of analysis are enabled in the provided script but were not run to generate analysis results as they require datasets within a narrow age-range, of which not enough are available here:

- 'GroupComparison': as 'Specificity', this analysis aims to monitor the occurrence of false positives in the analyses. Here, this is done in the context of a group comparison using two-sample t-statistics. The number of datasets in the 1st group is set by AnalParams.Ngroup1 in the GetAnalParams function. As for 'Specificity' analyses, the number of repetitions of this analysis is set by AnalParams.NRepeats in the GetAnalParams function.

- 'MotionBias': this analysis examines the dependence of the $R2^*$ data on the MDI (bias). This is achieved using F tests, inserting the MDI values into the design matrix. The powers of the MDI values inserted into the design matrix is set by Params.MotionRegPowers field in the GetParams function.

Note that the variable settings underlying each type of analysis are determined in the function **GetAnalParams.m**

ANALYSIS FUNCTIONS

The analysis functions used in the main function **RunQUIQI.m** are:

PrepAnalysis.m

Reads in the details of the data to be analyzed contained in Subject_Details.mat and completes all preparatory steps required for analysis (e.g. computation of the basis functions for ReML estimation, computation of the explicit masks for image analysis, editing of the analysis cohort if required,...).

Inputs:

- Subject_Details_All: Subject demographic information structure. Matlab structure with columns:
 - ID: string with subject ID
 - Age: double with subject age
 - Confound: Matlab structure containing the confounds of the analysis. Confound.Gender=1 (male) Confound.Gender=0 (female). Confound.BrainVol=double (brain volume)
 - QA: Matlab structure containing the MDI values for the weighted-data: QA.SDR2s.PDw=double, QA.SDR2s.T1w=double and QA.SDR2s.MTw=double
- DataTypes: type of input data for the analysis (see above).
- Subregions: region of interest of the analysis (see above).
- AnalType: analysis type (see above).

Output:

A QUIQI structure which, for each analysis to be conducted, contains information such as the folder location of the data to be analyzed, the output folder for the analysis, the type of data to be analyzed, the basis functions for the ReML computation, etc...

RunAnalysis.m

Runs the analyses and creates the results into specific folders. The inputs are a QUIQI structure and 'AnalType'.

MDIvsResAnalysis.m

In the case of a Residuals analysis type, for a given tissue type, compare the MDI values with the spatial variance of the residuals obtained from the model to be tested. Two measures are computed to assess the heteroscedasticity on the residuals obtained: a global (tissue type level) and a local measure (voxel level). The Econometrics toolbox is necessary to compute the maps of heteroscedasticity. The input is a QUIQI structure.

FOLDER STRUCTURE

Following analysis, the folders that contain the analysis results are organized as follows:

1. Top-level folder – in the home directory defined in GetParams. The name of this folder is given by the Params.AnalysisFolder field of GetParams (default: 'Analysis'), followed by the analysis type (see above).
2. Sub-folder: repetition number (repetitions are relevant for e.g. the 'GroupComparison' and 'Specificity' analyses).
3. Sub-sub-folder: type of input data (MTw, T1w, or PDw)
4. Sub-sub-sub-folder: region of interest of the analysis and powers of the MDI used

ADDITIONAL NOTES

- Regional grey matter analyses can be conducted using the provided script, for example using getROIpairs.m in the definition of the variable SubRegions in RunQUIQI.m (see commented example in this script). However, we do not provide the results of such analyses due to the large size of the data (>100GB).
- The FalsePositiveCount.m function that monitors the rate of false positives in statistical maps relies on the cp_cluster_Pthresh.m function, included in this package, that was written by Christophe Phillips, Cyclotron Research Centre, University of Liege, Belgium.
- These scripts support parallel computing. To use this feature read the header and adapt the scripts **PrepAnalysis.m**, **RunAnalysis.m**, and **MDIvsResAnalysis.m** accordingly.