**Manual for LICA-based Denoising Methods**

**1. Introduction**

LICA Denoising combined Multi-Study MRI Data for Removing Scanner and Site Variability from Multimodal MRI Data.LICA denoising method based on data fusion of multiple MRI measurements to identify multimodal spatial patterns related to scanner and/or study variations that can be used to denoise these effects from each individual modality. We capitalize on the strengths of LICA, namely that by using all measurements for every subject together in the LICA, we can better identify scanner/site effects as patterns that are distinct from meaningful patterns of brain structure/function than doing ICA on each modality separately. Once scanner/site effects have been identified, they can be removed from the original measurements to provide a “clean” set of measurements for each modality that are free of study/site effects that can be used for further statistical analyses. To that end it can produce combined data across studies to improve modality-specific statistical processing.

Four manuscripts provides a detailed description and evaluation of LICA and LICA-based denoising method:

1. Li, H.J., Smith, S.M., Gruber, S., Lukas, S.E., Silveri, M.M., Hill, K.P. Killgore, W.D.S., Nickerson, L.D. Combining Multi-Study MRI Data Linked-ICA Denoising for Removing Scanner and Site Variability from Multimodal MRI Data. NeuroImage, under review.

2. Groves, A.R., Beckmann, C.F., Smith, S.M., Woolrich, M.W., 2011. Linked independent component analysis for multimodal data fusion. NeuroImage. 54, 2198-2217.

3. Groves, A.R., Smith, S.M., Fjell, A.M., Tamnes, C.K., Walhovd, K.B., Douaud, G., et al., 2012. Benefits of multi-modal fusion analysis on a large-scale dataset: life-span patterns of inter-subject variability in cortical morphometry and white matter microstructure. NeuroImage. 63, 365-380.

4. Nickerson, L.D., Smith, S.M., Öngür, D., Beckmann, C.F., 2017. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. Front Neurosci. 11, 115.

**2. Requirement**

Install Matlab, FSL, Freesurfer, ImageMagic, epstopdf

Download the latest flica release from FSL webside (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLICA>), and unpack with tar xvzf flica\_2013-01-15.tar.gz. Put it under Matlab folder.

Or use the FLICA in our GitHub, this is more suggested as this has handle the missing data analysis.

**3. General info**

**Data processing**

Optimized modality-specific preprocessing pipelines, including quality assurance to identify data with excessive motion or other artifacts, were used to produce standard-space outcome images for each subject for a given modality. For example, modulated grey matter images generated by FSL-VBM; vertex-wise cortical thickness and pial surface area maps estimated using FreeSurfer by means of automated surface reconstruction scheme; FA, MD and tensor mode (MO) images calculated using FSL FDT; fMRI brain activation maps estimated by FSL FEAT (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT). For each modality, the images should be registered to MNI standard space, the spatial smoothness level and spatial resolution are not required identify among modalities. For each outcome that was derived from the structural MRI, DTI and fMRI data, a “subject series” was created by concatenating the resulting spatial maps across all participants into a single 4D (volume x subjects) data file using the same subject order. These outcomes are hereafter also referred to as modalities even though they are derived quantities.

**Missing data**

To account for the fact that participants from different studies had different measurements; In applying LICA, participants with high noise variance (1/precision) will influence a significant fraction of the LICA components (Groves et al., 2012). These participants should be removed from the LICA to achieve robust results.

A volume of zeros was used to represent a subject’s “missing” data for a given modality.

**3. Run LICA and identify noise components**

The subject-series for all modalities were analyzed simultaneously using LICA implemented in Matlab (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLICA).

The subject loadings for each component were assessed for relationships with scanner/study and participant variables (i.e., demographic, drug use, and task performance measures) using a single multivariate generalized linear regression. Those components whose subject loadings related *only* to SSWV/study variability were identified as noise components (evaluated by multiple comparison corrected p-value < 0.05).

All these can be done based on flica\_template\_script\_example.m

**LICA-R1 and LICA\_R2 Denoising methods**

Once the LICA components associated with scanner/study effects are identified, they can be used to denoise the original data with multivariate regression for each modality.

**LICA-R1 Denoising: Hard regression using LICA subject loadings**

A single multivariate regression of the subject loadings for only the noise components against the original data was implemented to remove noise components from each modality as follows:

1. extract subject loadings of all noise components from LICA, named as R1\_NOISE.txt

2. run fsl\_glm to remove noise effect by regressing R1\_NOISE.txt from original data for each modality. --out\_res output the denoised data.

*fsl\_glm -i Original\_4D\_subject\_Data.nii.gz -d R1\_NOISE.txt -o LICA\_R1\_Denoised --out\_z=LICA\_R1\_Denoised\_Z --out\_res=LICA\_R1\_Denoised\_res --demean -m mask.nii.gz --des\_norm*

**LICA-R2 denoising: Hard regression via dual regression using LICA spatial maps**

A multivariate dual regression procedure (Nickerson et al., 2017) using the LICA component spatial maps was implemented to remove noise components as follows.

1. First run dual\_regression with spatial maps of all the componnets from LICA for each modality (based on dual\_regression.sh usage):

*dual\_regression.sh Spatial\_maps\_from\_LICA.nii.gz 1 -1 0 output\_directiry Original\_4D\_subject\_Data.nii.gz*

2. Then run fsl\_glm to regress the noise components subject loadings generated from dual\_regression for each modality.

Extract subject loadings of all noise components from dual\_regression, named as R2\_NOISE.txt.

Run fsl\_glm to remove noise effect by regressing LD\_NOISE.txt for each modality. --out\_res output the denoised data

*fsl\_glm -i Original\_4D\_subject\_Data.nii.gz -d R2\_NOISE.txt -o LICA\_R2\_Denoised --out\_z=LICA\_R2\_Denoised\_Z --out\_res=LICA\_R2\_Denoised\_res --demean -m mask.nii.gz --des\_norm*