

Documentation for the mICA toolbox

Author: Florian Beissner (beissner.florian@mh-hannover.de)

1. General functionality

The toolbox is capable of four basic analyses, **Single-session mICA**, **Multi-session (group) mICA**, **Back-Reconstruction** and **Reproducibility analysis**. They can be selected in the drop-down menu at the top of the graphical user interface (GUI).

Single-session mICA: Runs a single-subject masked ICA using FSL MELODIC for one subject or independent analyses for a number of subjects in batch mode.

Multi-session (group) mICA: Runs a masked group ICA in FSL MELODIC in temporal concatenation or tensor mode.

Back-Reconstruction: Estimates functional connectivity of your masked ICs from a previous analysis with the same (i.e. masked) region or with any other region of the brain. Two different methods are available: Dual regression and direct back-reconstruction. For the former, the toolbox utilizes a modified version of FSL's dual-regression script that applies the first regression step on the masked data and the following steps on the whole-brain data. For the latter, results are calculated in a single step. The components' time courses serve as a design matrix in a GLM applied on the temporally concatenated whole-brain data of all subjects. As such calculations using FSL GLM require huge computational resources, because the whole data from all subjects are loaded to the memory, the toolbox automatically splits the concatenated 4D data along the Z axis and performs GLM fitting slice by slice.

Reproducibility analysis: Performs random split-half reproducibility estimation. For every one repetition, ICA analysis is performed separately on both split-half samples. A cross-correlation matrix between the ICs' spatial maps is then calculated using Pearson's coefficient. Inter-group matching of ICs is accomplished using a Hungarian sorting algorithm. Mean correlation of all matched IC pairs averaged over N repetitions is then used as the reproducibility measure. Based on the estimated reproducibility curve, the user can then select an optimal dimensionality for the analysis. We recommend the global maximum of the curve. However, there may be cases when this value may be unsuitable, for example, when it leads to nonconvergence in a significant proportion of mICAs, or if the user has a priori knowledge on the anatomical region under study. In these cases one of the other local maxima may be chosen instead. 95% CI can help to assess the stability of such local maxima.

2. Modules

The GUI itself is separated into four modules, **Input**, **Mask**, **Options** and **Output**. Every analysis requires input in these four sections before it can be executed by clicking **Go** at the bottom of the GUI.

2.1 Input

The function of this module is to select the input data for the ensuing analysis. In general, data preprocessing should comprise the usual steps, like distortion/motion correction, high-pass filtering, and spatial normalization to standard space. While the input fMRI data should ideally be preprocessed before being used with mICA toolbox, a number of preprocessing steps, such as smoothing, resampling and cropping can be applied using the toolbox itself. The motivation to apply these operations using the

toolbox is that they will be applied in an optimal order to achieve good mICA results. It is important to mask input fMRI data before applying any spatial smoothing, as this approach prevents signal from voxels outside of the mask influencing the signal of those inside. Resampling is another step to think about as it may make sense to use higher than normal up-sampling for analyses of small regions. Therefore, the toolbox can automatically up-sample input data to the desired resolution.

There are two choices for data entry: **4D nifti file(s)** or **FEAT folder(s)**.

4D nifti file(s) should be used if your input data is already preprocessed (except for smoothing) and in 4D-nifti format. Simply choose the number of files you want to analyze and put in the location of these file(s) by clicking **Select**.

FEAT folder(s) should be used if you used FSL FEAT for preprocessing and want to use the filtered_func_data.nii.gz generated by FEAT for your analysis.

While Single-session ICA mode does not require any spatial normalization, all files need to be coregistered to the same standard space to perform a group ICA analysis. Therefore, the **FEAT folder(s)** option allows for an automated registration to standard space if this has been calculated during preprocessing with FEAT. Enabling **Apply reg2standard** will automatically apply the transformation from the reg folder in the FEAT directory to your filtered_func_data.nii.gz.

Two further options in the Input section control spatial smoothing and resampling.

Resampling resolution (mm) This sets the final spatial resolution in which to run the analysis. For most brain regions, 2mm will be a sensitive choice here. However, if you want to look into small regions, like the brainstem, hypothalamus or the amygdala, you may want to try 1mm.

Spatial smoothing FWHM (mm) This sets the smoothing kernel to be applied after the data have been masked and resampled.

Note: In Back-Reconstruction mode the input module refers to the data containing the region to which you are establishing connectivity from your masked ICs and not to the masked ICA input.

2.2 Mask

The function of this module is to select the mask to be used in the analysis. This mask can have different functions. In ICA mode (single-session or group), this will be the region where you want to perform your masked ICA in. It is also possible to select several regions that will be combined into a single mask. These regions do not need to be adjacent or overlap. In Back-Reconstruction mode, the mask will define the region to which you want to establish connectivity of your masked ICs. This can be the same region (e.g. brainstem to brainstem), another region (e.g. amygdala to prefrontal cortex) or the entire brain (e.g. thalamus to the rest of the brain). In the latter case, you can simply choose a whole-brain mask.

Start by entering the desired total number of (sub-)masks and press **Select**. Then choose those masks from the menu, one for each **Mask #**. There are multiple sources to choose your mask from, most of which are anatomical atlases included in the FSL package. You first choose the atlas and then one of its regions. In case the atlases are probabilistic, you also have to select an **Atlas probability threshold**, which will be applied before the mask is binarized for the analysis. You can also use your own mask by selecting **Other**. All selected masks will be merged for the analysis in a file called **final_mask.nii.gz**.

2.3 Options

This module changes depending on the selected analysis type.

Number of components. You can either enter a single value, comma separated values, or a range here to set the number of components for your ICA. If more than one value is entered, the analysis is repeated for each parameter value.

For example: 0 will perform a mICA using MELODIC automated dimensionality estimation.
20 will perform a mICA with a decomposition dimensionality of 20.
20-29 will perform 10 mICAs, one for each dimensionality between 20 and 29.
20,30,40 will perform 3 mICAs, one for each dimensionality of 20, 30 and 40.
0,20-29,40 is a possible input, too.

Each analysis will be saved in a sub-folder called dim#, where # is the selected number of dimensions.

Advanced options will contain the option **Generate Parcellation** for some types of analysis, which - when checked - will output a file **parcellation.nii.gz** in which each voxel is assigned the component number with the highest z-value in that voxel. This represents a binary parcellation of the masked region based on its intrinsic connectivity. **Crop image** is a further useful tool when running analyses with large number of subjects or with many repetitions. It removes all parts of the image that do not contain a piece of the mask and, thus, drastically reduce data size. **Background image** will let you define a background image for the MELODIC web report and **Additional MELODIC command flags** can be used to pass any additional arguments directly to MELODIC.

For example: "--approach=tica" will tell MELODIC to run a Tensor-ICA instead of the standard temporal concatenation approach.

Timeseries/Session model is only available for single-session and group ICAs. If a time-series "design.mat" model and "design.con" contrast file is selected (like the ones used in first-level FEAT analyses), they will be used in ordering the ICs, and in providing additional information. For multi-session/multi-subjects ICAs, you can achieve the same by selecting a "design.mat" and "design.con" like the ones used in a higher-level FEAT analysis.

Back-reconstruction

The back-reconstruction tool will estimate connectivity of your masked ICs from a previous analysis with the same (i.e. masked) region or with other regions of the brain. In the options module you first have to choose whether to run a **Dual Regression** and / or **Inversion** (a.k.a. direct back-reconstruction). At the moment it is hard to say which one is the better option. If your main interest lies in the connectivity of your masked ICs, you may be better of using inversion, whereas dual regression may be the more appropriate option for running group comparisons. For the details, please consult our toolbox paper.

In both cases you have to provide the toolbox with the path to your original **ICA directory** (i.e. the one containing the melodic_IC.nii.gz and melodic_mix files) as well as the list of files that have been used in this ICA (**ICA input**). If the ICA has been calculated using mICA toolbox, you can simply use the input_preproc.txt file for this.

Mixture model threshold is the last choice you have to make. It only applies to inversions, where the toolbox adapts a mixture model approach for significance testing provided by the MELODIC command-line tool. A threshold level of 0.5 is standard, while higher values of e.g. 0.75 will lead to

fewer false-positive results.

Reproducibility analysis

The only unique input for reproducibility analyses is **Repetitions of Split-Half sampling**, which sets the number of random samplings. You should be warned that entering high numbers here will drastically increase computational time, especially, if you have not set up a parallel computing environment on your computer.

Entering 0 here will omit random sampling and instead compare the first half of subjects against the second half. This can be used for test-retest reproducibility analyses, where the test data need to be entered as the first half and re-test data as the second half of subjects.

2.4 Output

You can select your desired output folder here.

3. Output files and folders

Depending on the type of analysis, different files and folders will be created.

3.1 Output generated by all analyses

Independently of the analysis type, the top folder will contain the following files.

mica_log.txt	General information about the analysis that has been performed.
input_orig.txt	A list of functional images that were selected as in the input module.
input_preproc.txt	The same images as in input_orig.txt, however, after masking and applying the selected preprocessing steps.
final_mask.nii.gz	The mask used in the masked ICA. It may have been merged from several masks selected in the Mask module.
bgimage.nii.gz	The background image (if one has been selected).

If preprocessing (i.e. resampling, smoothing or cropping) has been used, the top folder will further contain a **masked_input** folder with all preprocessed (i.e. masked, smoothed, resampled and/or cropped) input files. These can be re-used for later analyses to save processing time. Furthermore, **tmp_log** folders are created by each analysis. They can be safely deleted. However, the log files in these folders can help to trace errors.

3.2 Output generated by Single-session or Multi-session (group) mICAs

In **Multi-session (group) mICA** mode, the toolbox will create a folder called **dim#** for each dimensionality #. In **Single-session mICA** mode, a **subj_#** folder will be created for each subject (with # starting from 0001). This folder will then contain the **dim#** sub-folder(s) for each dimensionality. These sub-folders contain the standard ICA output that MELODIC generates and in addition a **parcellation.nii.gz** file (if “generate parcellation” was checked).

3.3 Output generated by Back-Reconstruction analyses

Depending on the selection made in the options module, the top folder will contain sub-folders called **inversion** and / or **dR** (for dual regression). Furthermore, it will contain the **final_mask.nii.gz** defining the target region to which connectivity for the masked ICs is to be established.

The **inversion** folder will contain three files:

all_inversion.nii.gz	The concatenated raw inversion results (parameter estimates). The order of components is the same as in melodic_IC.nii.gz
all_zstat.nii.gz	The concatenated unthresholded z-transformed inversion results.

all_thresh_zstat.nii.gz The concatenated z-transformed inversion results thresholded by a mixture-model approach.

The **dR** folder will contain the standard output of a dual regression calculated with FSL's dual regression script. In addition it will contain the concatenated results:

all_dr_stage3_tstat#.nii.gz The concatenated unthresholded t-transformed dual regression results. The order of components is the same as in melodic_IC.

all_dr_stage3_tfce_corr_p_tstat#.nii.gz The concatenated t-transformed dual regression results thresholded using the TFCE approach and FSL RANDOMISE.

all_dr_stage3_vox_corr_p_tstat#.nii.gz The concatenated t-transformed dual regression results thresholded using voxel-level correction and FSL RANDOMISE.

3.4 Output generated by Reproducibility analyses

For each random split-half procedure there will be one **sample_#** folder (with # starting from 0001) containing sub-folders **group1** and **group2** which reflect the two (half) samples to be compared. Each group folder will contain one **dim#** folder for each tested dimensionality #. These folders contain the usual MELODIC ICA output.

The results for each split-half sample can be found in the respective **sample_#** folders in the file(s) **corr_dim#.txt** that contain the spatial cross-correlation matrix for this sample and all tested dimensionalities.

The following files will be created by a successful reproducibility analysis.

all_melodic_commands.txt A list of all melodic commands used for the analysis.

corr.csv Mean correlation coefficients of all matched IC pairs for each sample (columns) and dimensionality (rows).

corr_mean_sd.csv Mean value (column 1) and 95% confidence interval (column 2) of all mean correlation coefficients of all matched IC pairs from all samples. This is the main measure of reproducibility used by the toolbox.

corr_mean.png A figure showing the mean of all mean correlation coefficients with 95% confidence interval for all tested dimensionalities (“dimensionality curve”).

no_convergence_error.png A figure showing for each tested dimensionality the number of mICAs that have not converged. These information can help to identify problematic dimensionalities.