

# Total Activation and iCAPs Toolbox

Version 1.0

Daniela Zöller, (daniela.zoller@epfl.ch)

August 13, 2018

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Example data</b>	<b>2</b>
<b>3</b>	<b>Total Activation</b>	<b>2</b>
3.1	Data Input and Preprocessing Options . . . . .	2
3.2	TA-related Options . . . . .	2
<b>4</b>	<b>Thresholding</b>	<b>2</b>
4.1	Thresholding-related Options . . . . .	2
<b>5</b>	<b>Clustering</b>	<b>2</b>
5.1	Aggregating data . . . . .	3
5.2	Clustering . . . . .	3
5.3	Consensus Clustering . . . . .	3
5.4	Cluster Consensus . . . . .	3
5.5	Clustering-related Options . . . . .	3
<b>6</b>	<b>Time Courses</b>	<b>4</b>
6.1	Time courses-related Options . . . . .	4
6.2	Transient-informed regression . . . . .	4
6.3	Unconstrained regression . . . . .	5

## 1 Introduction

This document gives a step by step guide on how to use the total activation (TA) and innovation-driven coactivation patterns (iCAPs) toolbox. Theory and background of the algorithms can be found in [Karahanoglu et al., 2011, Karahanoglu et al., 2013, Karahanoglu and Van De Ville, 2015, Farouj et al., 2017, Zöller et al., 2018].

Main functionalities of the toolbox include:

- Total activation for regularized deconvolution of fMRI data
- Thresholding of innovation frames based on surrogate data
- Clustering of innovation frames to extract spatial brain activity patterns
- Spatio-temporal regression for time course recovery of networks

For each step, there exists a script “main-[stepname]”, which first calls a script “Inputs-[stepname]” to specify the parameters and then the function “Run-[stepname]” to execute the selected step according to the specified parameters. In general, all parameters are saved in a matlab struct variable called “param”.

## 2 Example data

The toolbox includes scripts and results for an example resting-state fMRI dataset. The example data was obtained from the OpenfMRI database (<https://openfmri.org/dataset/ds000030/>) [Poldrack et al., 2016] and pre-processed using Statistical Parametric Mapping (SPM12, Wellcome Trust Centre for Neuroimaging, London, UK: <http://www.fil.ion.ucl.ac.uk/spm/>). Initial preprocessing steps included realignment, spatial smoothing with an isotropic Gaussian kernel of 6 mm full-width half-maximum, co-registration of structural scans to the functional mean and segmentation with the SPM12 *Segmentation* algorithm [Ashburner and Friston, 2005]. Further motion correction steps are included in the total activation pipeline.

Look at the example main and input scripts to see how to run the complete pipeline on the example fMRI data of two subjects. Note that the results are not stable for this few data, the example only exists to demonstrate how input and output data is saved. Resulting iCAPs and time courses are very noisy and unstable because there is not enough input data for the clustering. For stable results, more subjects or longer runs would be necessary!

## 3 Total Activation

This sub-routine of the toolbox reads input fMRI data, conducts preprocessing steps for motion correction according to the options specified by the user and saves the results for every subject in a subdirectory “TA\_results/<title of project>” in every subject’s folder. After execution on the TA pipeline, this folder will contain the following subfolders:

- *inputData*: contains preprocessed input data
- *Total Activation*: contains results of total activation on real data
- *Surrogate*: contains results of total activation on surrogate data

### 3.1 Data Input and Preprocessing Options

Have a look at the file *Inputs\_TA\_Data.m* for the required inputs related to fMRI data reading and preprocessing. If you don’t want to do any motion correction (possibly if you have done it already beforehand), set the fields *doDetrend*, *doScrubbing* to 0.

### 3.2 TA-related Options

Have a look at the script *Inputs\_TA.m* for the list and explanation for all parameters related to total activation.

## 4 Thresholding

This sub-routine applies a two step thresholding procedure to determine, which innovation frames (in Total Activation/Innovations.nii) contain significant transients. After execution of the Thresholding procedure, the following folder will be created in each subject’s “TA\_results/<title of project>” subfolder:

- *Thresholding*: contains results from thresholding. A subfolder is created for each set of thresholding options.

### 4.1 Thresholding-related Options

Have a look at the file *Inputs\_Thresholding\_Data.m* for the required inputs related to reading of TA results files. In *Inputs\_Thresholding.m* all options for thresholding are provided and explained.

## 5 Clustering

**Attention!**

For clustering, all subjects must be normalized to the same space. If you have run TA/thresholding in the subject-specific space, you need to normalize the thresholded innovation frames to MNI before continuing!

Tip: create a separate folder with the normalized files in “TA\_results/<title of project>\_MNI” and save the normalized data in it with the correct structure (i.e., in a subfolder called “TotalActivation”). Like that you can

simply change the *param.title* to read the normalized files. In *example\_normalization.m* there is an example for the normalization of required files and updating of *param* for clustering.

## 5.1 Aggregating data

The clustering routine first reads and concatenates TA and thresholding results of all subjects. Concatenated files will be saved in a folder “iCAPs\_results/<title of project>\_<thresholding specifications>”. The following clustering input files will be saved:

- *final\_mask* (.mat and .nii): the final mask of voxels with data in all subjects (see clustering options for the options to create this mask, by default it will be the intersection between all subjects’ grey matter masks)
- *AI* ( $N_{timepoints} \times N_{voxels}$ ): concatenated activity inducing signals of all subjects
- *AI\_subject\_labels* ( $N_{timepoints} \times 1$ ): subject assignment of each row in *AI*
- *Lsig* ( $N_{significant\_innovations} \times N_{voxels}$ ): Concatenated significant innovations for all subject. Positive and negative innovations are split into separate frames.  $N_{significant\_innovations} = N_{significant\_positive\_innovations} + N_{significant\_negative\_innovations}$ . See [Karahanoğlu and Van De Ville, 2015] for more details.
- *subject\_labels* ( $N_{significant\_innovations} \times 1$ ): subject assignment for each row of *Lsig*.
- *time\_labels* ( $N_{significant\_innovations} \times 1$ ): time point assignment for each row of *Lsig*.

## 5.2 Clustering

After aggregating data from all subjects, clustering will be applied to significant innovation frames *Lsig*. Clustering results will be stored in a subfolder named according to the specific clustering options. The following outputs will be created by this part of the routine:

- *iCAPs* (.mat and .nii): cluster centers of all resulting clusters, these are the spatial iCAPs maps
- *iCAPs\_z* (.mat and .nii): z-scored cluster centers
- *dist\_to\_centroid* ( $N_{significant\_innovations} \times N_{clusters}$ ): distances from every clustered innovation frame to each cluster center
- *IDX* ( $N_{significant\_innovations} \times N1$ ): for each frame the index of the cluster it is assigned to
- *param*: structure containing all input parameters, which lead to these results

## 5.3 Consensus Clustering

If requested according to the options, consensus clustering [Monti et al., 2003] is applied to evaluate cluster stability and find the best number of clusters to consider. After the routine, consensus matrices for each requested cluster number will be saved in a consensus clustering subfolder, as well as CDF and AUC of values in the consensus matrices. These files can then be used according to [Monti et al., 2003] to decide on the best number of clusters.

## 5.4 Cluster Consensus

If clustering and consensus clustering has been done, the consensus of every cluster can be computed as the average consensus index between all pairs of frames belonging to the same cluster. Values range from 1 (all frames are always clustered together across resamples) to 0 (frames are never being clustered together). Higher values indicate higher stability of the cluster [Monti et al., 2003].

## 5.5 Clustering-related Options

Have a look at the file *Inputs\_Clustering\_Data.m* for the required inputs related to reading of TA and Thresholding results files. In *Inputs\_Clustering.m* all options for clustering are provided and explained.

## 6 Time Courses

As the iCAPs' spatial maps are retrieved from significant innovations, time courses have to be extracted by backprojection to the activity inducing signals. In this routine, one can choose to apply transient-informed or unconstrained regression for time course retrieval.

### 6.1 Time courses-related Options

Have a look at the file *Inputs\_TimeCourses\_Data.m* for the required inputs related to reading of Clustering results files. In *Inputs\_TimeCourses.m* all options for time course retrieval are provided and explained.

### 6.2 Transient-informed regression

If time courses are extracted by transient-informed regression [Zöller et al., 2018], a subfolder will be created in the corresponding clustering folder. It will contain the following files:

- *Results for all requested soft assignment factors (see options):*
  - *TC*: cell array with time courses of every subject
  - *TC\_stats*: statistics of data fit (BIC, AIC, ...)
  - *tempChar*: cell array with temporal characteristics (duration, coactivations, ...) for every subject's time courses
- *Files related to selection of best soft assignment factor:*
  - *BICsum*: sum of BIC across all subject, knee point marked in red
  - *BICdist*: distribution of BIC across all subject, knee point marked in red
  - *AICsum*: sum of AIC across all subject, knee point marked in red
  - *AICdist*: distribution of AIC across all subject, knee point marked in red

Results for the best soft assignment factor according to BIC will be saved in main iCAPs directory, it will contain the files: *TC* (time courses), *TC\_stats* (statistics: BIC/AIC etc.) and *tempChar* (temporal characteristics).

Computed temporal characteristics are:

- *nTP\_sub*: total number of time points per subject
- *innov\_counts\_total* / *innov\_counts\_perc*: total number / percentage of significant innovation frames
- *innov\_counts* (*\_percOfInnov*): number of innovations in each iCAP (percentage of total number of significant innovations)
- *TC\_norm\_thes*: z-scored and thresholded time courses
- *TC\_active*: binarized time courses (+1/0/-1 for positive/none/negative activation)
- *coactiveiCAPs\_total*: number of co-active iCAPs per time point
- *occurrences* (*\_pos/\_neg*): number of activation blocks (positive/negative activations)
- *duration\_total\_counts* (*\_pos/\_neg*): total number of active time points (positive/negative activations)
- *duration\_avg\_counts* (*\_pos/\_neg*): average duration of every (=occurrences/duration\_total\_counts)
- *coupling\_counts*: number of frames with co-activation for every pairwise combination of iCAPs
- *coupling\_jacc*: coupling Jaccard index (overlap divided by activation union)
- *coupling\_sameSign(\_diffSign)\_counts*: positive (negative) couplings
- *coupling\_sameSign(\_diffSign)\_jacc*: positive (negative) coupling Jaccard index

For the computation of duration and couplings, only activation blocks of 2 timepoints or more have been considered

### 6.3 Unconstrained regression

We do not recommend this option, but provide the possibility to compare with [Karahanoglu and Van De Ville, 2015, Bolton et al., 2018] If time courses are extracted by unconstrained regression, a subfolder “TCs\_unconstrained” will be created in the corresponding clustering folder. It will contain the following files:

- *TC\_unc*: cell array with time courses of every subject
- *TC\_unc\_stats*: statistics of data fit (BIC, AIC, ...)
- *tempChar\_unc*: cell array with temporal characteristics (duration, couplings, ...) for every subject’s time courses

## References

- [Ashburner and Friston, 2005] Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *Neuroimage*, 26(3):839–851.
- [Bolton et al., 2018] Bolton, T. A. W., Tarun, A., Sterpenich, V., Schwartz, S., and Van De Ville, D. (2018). Interactions Between Large-Scale Functional Brain Networks are Captured by Sparse Coupled HMMs. *IEEE Trans. Med. Imaging*, 37(1):230–240.
- [Farouj et al., 2017] Farouj, Y., Karahanoglu, F. I., and Van De Ville, D. (2017). Regularized Spatiotemporal Deconvolution of fMRI Data Using Gray-Matter Constrained Total Variation. *Proc. 14th IEEE Int. Symp. Biomed. Imaging From Nano to Macro*, pages 472–475.
- [Karanoglu et al., 2011] Karahanoglu, F. I., Bayram, I., and Van De Ville, D. (2011). A Signal Processing Approach to Generalized 1-D Total Variation. *IEEE Trans. Signal Process.*, 59(11):5265–5274.
- [Karanoglu et al., 2013] Karahanoglu, F. I., Caballero-Gaudes, C., Lazeyras, F., and Van De Ville, D. (2013). Total activation: FMRI deconvolution through spatio-temporal regularization. *Neuroimage*, 73:121–134.
- [Karanoglu and Van De Ville, 2015] Karahanoglu, F. I. and Van De Ville, D. (2015). Transient brain activity disentangles fMRI resting-state dynamics in terms of spatially and temporally overlapping networks. *Nat. Commun.*, 6:7751.
- [Monti et al., 2003] Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.*, 52(1):91–118.
- [Poldrack et al., 2016] Poldrack, R. A., Congdon, E., Triplett, W., Gorgolewski, K. J., Karlsgodt, K. H., Mumford, J. A., Sabb, F. W., Freimer, N. B., London, E. D., Cannon, T. D., and Bilder, R. M. (2016). A phenome-wide examination of neural and cognitive function. *Sci. Data*, 3:1–12.
- [Zöllner et al., 2018] Zöllner, D. M., Bolton, T. A. W., Karahanoglu, F. I., Eliez, S., Schaer, M., and Van De Ville, D. V. D. (2018). Robust recovery of temporal overlap between network activity using transient-informed spatio-temporal regression. *under Rev.*