Energy Landscape Analysis Toolbox (ELAT) User's Guide (ver. 2.0)

Takahiro Ezaki (https://sites.google.com/site/ezakitakahiro/cv)

Latest update: 3 Sep 2018

Table of Contents

Introduction: What is the energy landscape analysis?

Thank you for considering using the energy landscape analysis toolbox (ELAT). The energy landscape analysis is a computational method that enables intuitive interpretation of multivariate time series. This analysis comprises four steps (Fig. 1.): (1) Binarization of the data, (2) estimation of the maximum entropy model (Boltzmann distribution), (3) construction of a disconnectivity graph and basin of energy local minimums, and (4) computation of dynamics measures on the energy landscape. This method was originally designed for analyzing fMRI data, but it is in principle applicable to other types of data. In our experience, the energy landscape analysis works nicely when the number of variables is roughly 6 to15. For more variables, the computational cost becomes large and interpretation of the results becomes difficult. In such cases, we recommend reducing the number of variables (consider using ICA, merging variables, etc.).

For detailed algorithms of the energy landscape analysis, please refer to Ref [1]. Examples of applications of this analysis are found in Refs. [1-7]. If you have any questions or find an error, please contact ezaki@jamology.rcast.u-tokyo.ac.jp.

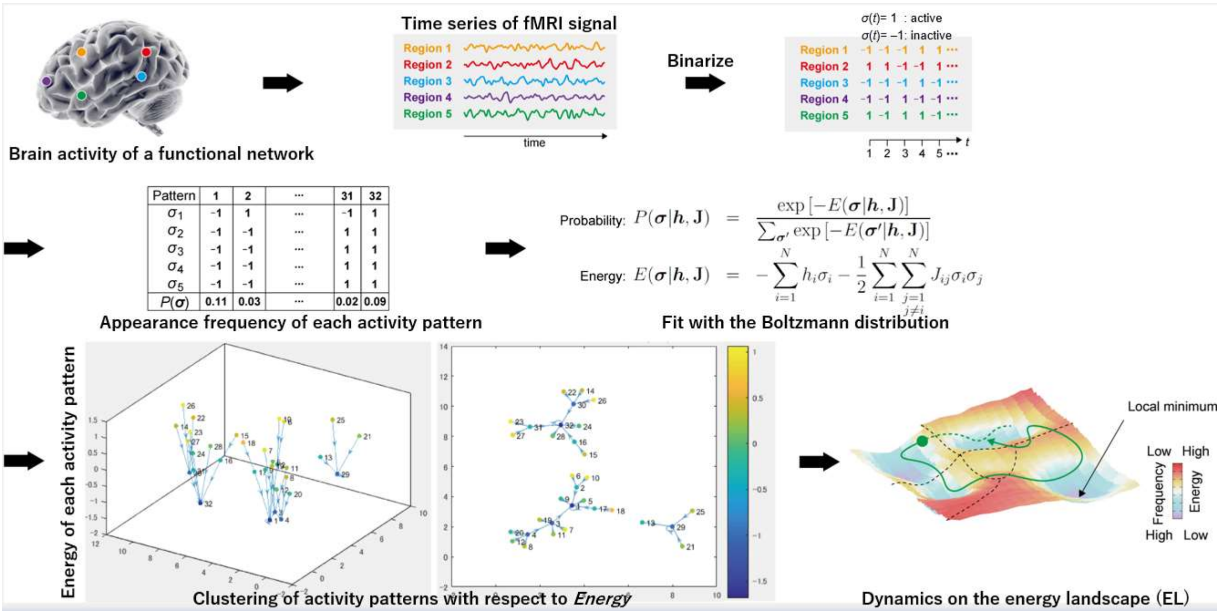

Fig. 1. Overview of the energy landscape analysis.

This toolbox performs the entireall of the computations necessary for the energy landscape analysis (Fig. 2). The Full Analysis option computes everything including the dynamics measures of individuals. The Energy Landscape Construction option computes the energy landscape and does not perform further analysis.

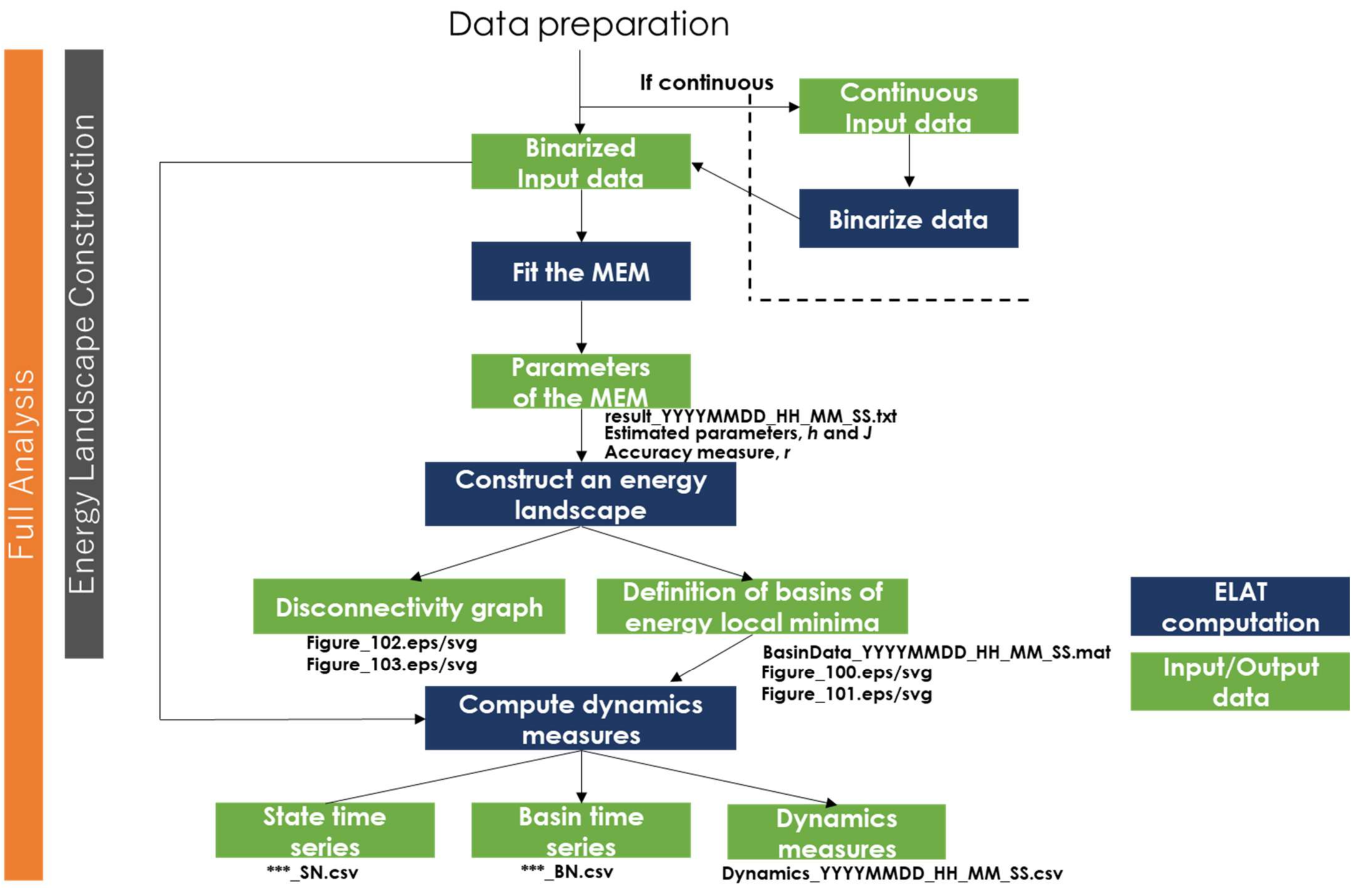

Fig. 2. Overview of ELAT.

Technical details of the energy landscape analysis

We denote the analyzed data by $\mathbf{x_i(t)}$, where $\boldsymbol{i}$ ($\boldsymbol{i=1,\dots, N}$) labels a variable and $\boldsymbol{t}$ ($\boldsymbol{t=1,\dots, t_{max}}$) represents time.

1. Binarization

The input data for the energy landscape analysis must be binary (i.e., +1 / -1). Thus, if your data $\{x_i(t)\}$ take continuous values, you have to binarize them with an appropriate threshold. In our previous studies [1,7], we binarized our fMRI data by setting a threshold to the average of the signal for each variable, i.e., $\boldsymbol{Y_i(t)} = \mathbf{1}$ **(if** $\boldsymbol{x_i(t) > [x_i]}$**);** $\boldsymbol{Y_i(t)} = \boldsymbol{-1}$ **(otherwise)**, where $\boldsymbol{[A]}$ denotes the time average of $\boldsymbol{A}$. Because the baseline of the signal might not be consistent across individuals, we recommend to perform binarization for each individual and each variable. This procedure is implemented in this toolbox and will be explained later.

2. State number

The state of the system at time $\boldsymbol{t}$ is represented by a binary vector $\boldsymbol{(Y_1(t), Y_2(t),\dots, Y_N(t))}$. This state vector takes one of the 2^N states. For convenience, we enumerate the state with a simple conversion from binary to decimal: $\boldsymbol{s} = 1+\sum \boldsymbol{2^i}\ \boldsymbol{(Y_i+1)/2}$. The factor $\boldsymbol{(Y_i+1)/2}$ maps {–1,1} to {0,1}, respectively. This conversion transforms, for example,

(–1, –1, –1, –1) to 1

(1, 1, 1, 1) to 16

(1, –1, 1, –1) to 6 .

Thus, the binarized states are enumerated from 1 to 2^N.

3. Accuracy measure of MEM fitting

The accuracy of the fitting is measured by $\boldsymbol{r}$, which is output in the console. The definition of this measure is found in Ref. [1]. When $\boldsymbol{r}$ is close to 1, it means the MEM fitting is successful. We recommend reporting this value. Note that this value inevitably becomes small when the length of the data is small [1].

(a) Data

Time →

Variable 1	1	1	1	1	1	1	-1	...
Variable 2	-1	-1	-1	-1	-1	-1	-1	...
Variable 3	1	1	1	-1	-1	-1	-1	...
Variable 4	-1	-1	-1	1	1	1	1	...
Variable 5	1	1	1	1	1	-1	-1	...
Variable 6	1	1	1	1	-1	-1	-1	...
Variable 7	1	1	1	1	1	1	-1	...

(b) List of variable names

Variable 1	left aPFC
Variable 2	right aPFC
Variable 3	left aI/fO
Variable 4	right aI/fO
Variable 5	dACC/msFC
Variable 6	left ant thal
Variable 7	right ant thal

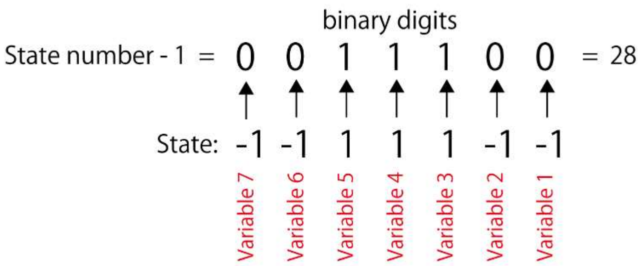

Fig. 3. Input data and the definition of the state number.

How-to 1: A tutorial

Here we show how the toolbox works in the shortest path. (See Fig. 4.)

1. Prepare data

First, prepare the data with ".dat" file(s). The shape of the data must be N (number of variables) rows and t_{max} columns (see also Fig. 3). Each digit must be separated by a tab. Please refer to "testdata.dat" contained in the toolbox folder. If the data is continuous, and you wish to binarize them in your own way, please do so at this stage. If you accept the thresholding with respect to the average of each variable, just prepare the raw data. If you do not have the data and want to see how it works, please use "testdata_1.dat"~"testdata_4.dat" in the folder.

2. Launch the toolbox

Unzip the toolbox and place it somewhere convenient for you. Open "StartProgram.m" with MATLAB, and execute. Older versions of MATLAB might not support some of the functions used, so please update yours to the latest version (in many cases, updating is available for free).

3. Select type of analysis

Here, simply select "Full Analysis" (default).

4. Select input data

Select the data you prepared in 1. Multiple files can be selected. For example, if you are studying the data obtained from participants in two (or more) groups, and just want to see what happens, we recommend that you add all of the participants' data. If you want to use the test data, select "testdata_1.dat"~"testdata_4.dat" in the folder.

5. Select binarization option

Select "Binarized data" or "Continuous data." If you select "Continuous data" and wish to use the average thresholding, please leave the "threshold" as 0 (default). If you change this value to **x**, the threshold will be set to **average + x**. If you selected the test data, please select "Binarized data."

6. Select Basin Data Generation option

Here, just go with "Construct energy landscape from input files." We will explain the "Read basin data" option later (p. 12).

7. Select variable name (ROI name)

If you have a list of the names of the variables, please select it here. The format of the file must be a ".dat" file with N rows and 1 column (same as the input data prepared in 1.). This is optional. For reference, please see "roiname.dat" in the folder.

8. Select output folder

Select the output folder. We recommend that you create a new folder for only this analysis. If you want to see the list of the energy landscape basin, please also check "Save Basin List."

9. Execute

Execute!

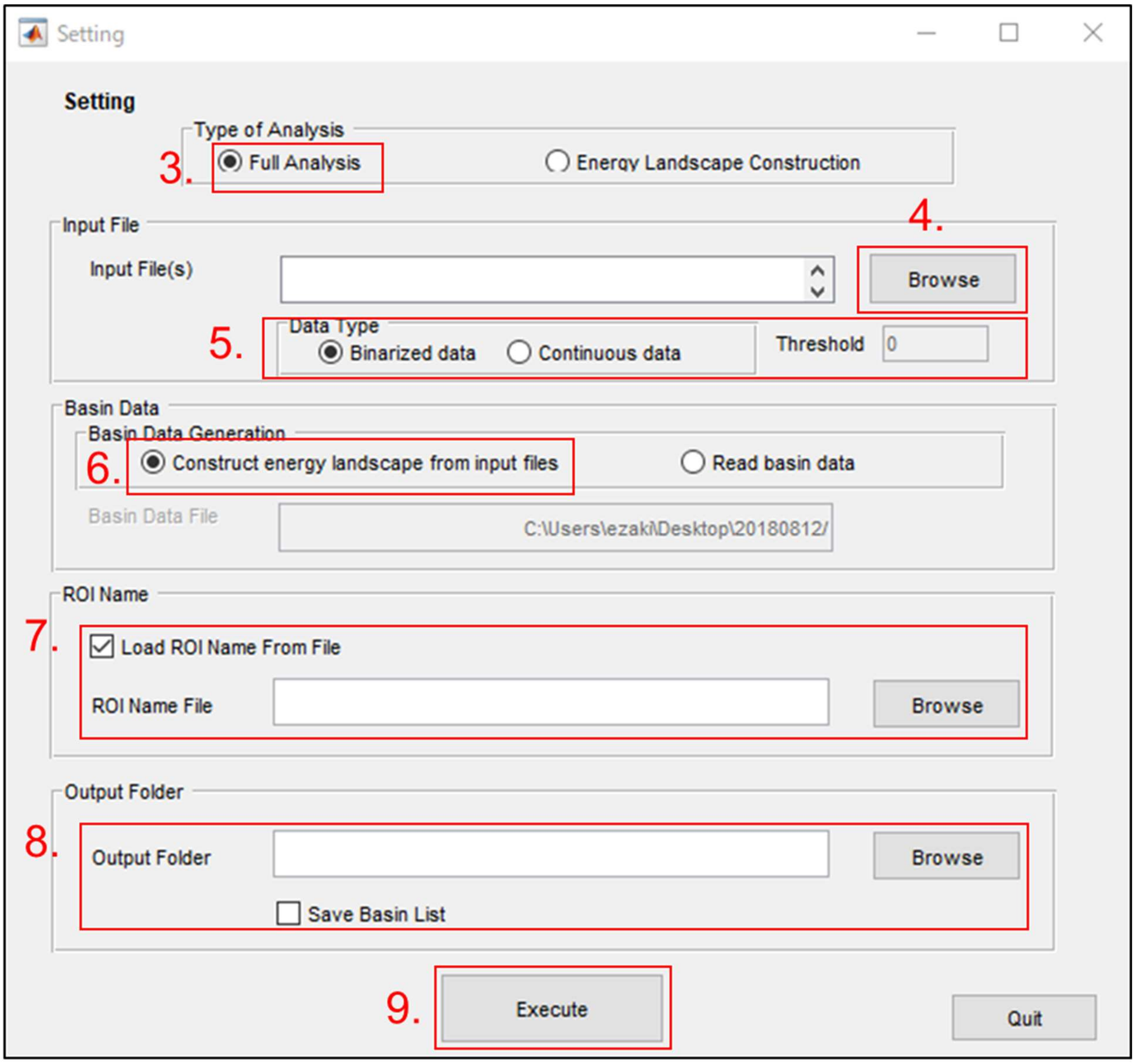

Fig. 4. Configuration window.

How-to 2: Results

1. Energy landscape

As a result, first, you will get four pictures. The first two visualize the energy landscape (Fig. 5). Each node represents a state specified by its state number. Each link visualizes the path between neighboring states which had the largest energy difference (i.e., the steepest path to the energy local minimum). The absolute positions of the nodes are not important. Each isolated cluster is a "**basin**" of each local minimum. The energy value is shown with a color, which is also visualized by the height in the 3D version of the figure (right). The states with a small energy value are considered to appear frequently, and thus they should be important states in many cases.

The two figures are saved in the output folder.

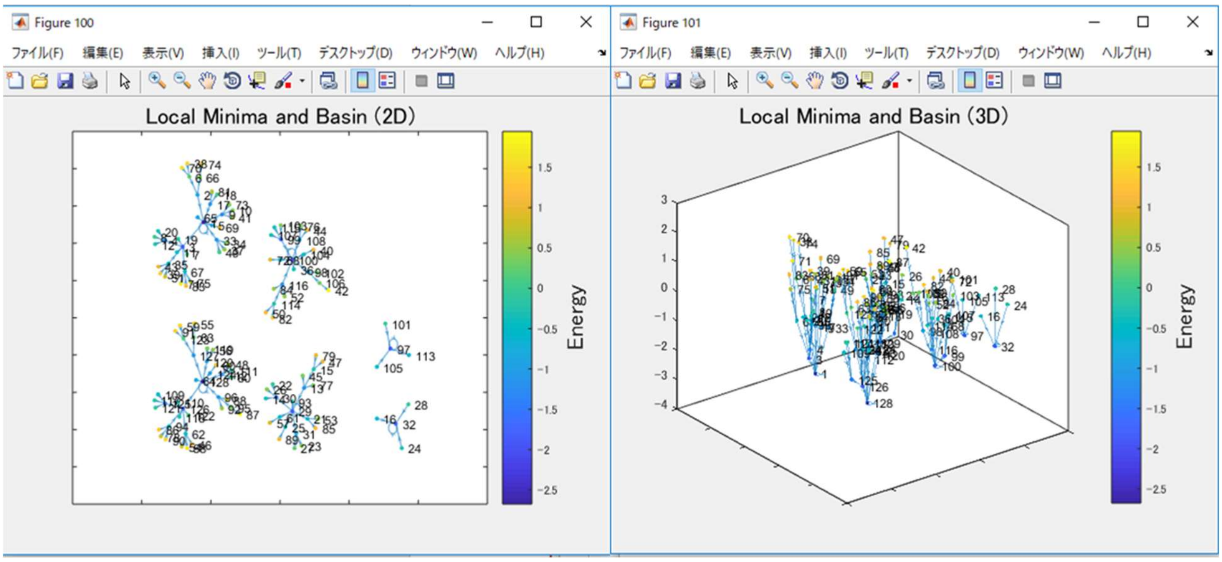

Fig. 5. Output figures 1 and 2.

The other two figures give a more abstract representation of the energy landscape. The disconnectivity graph (Fig. 6 left) shows the positions of the local minimum states (specified in Fig. 6 right) and their relationships. If you perform this analysis for more than one group, this disconnectivity graph is useful for comparisons. Note that the basin information is dropped here.

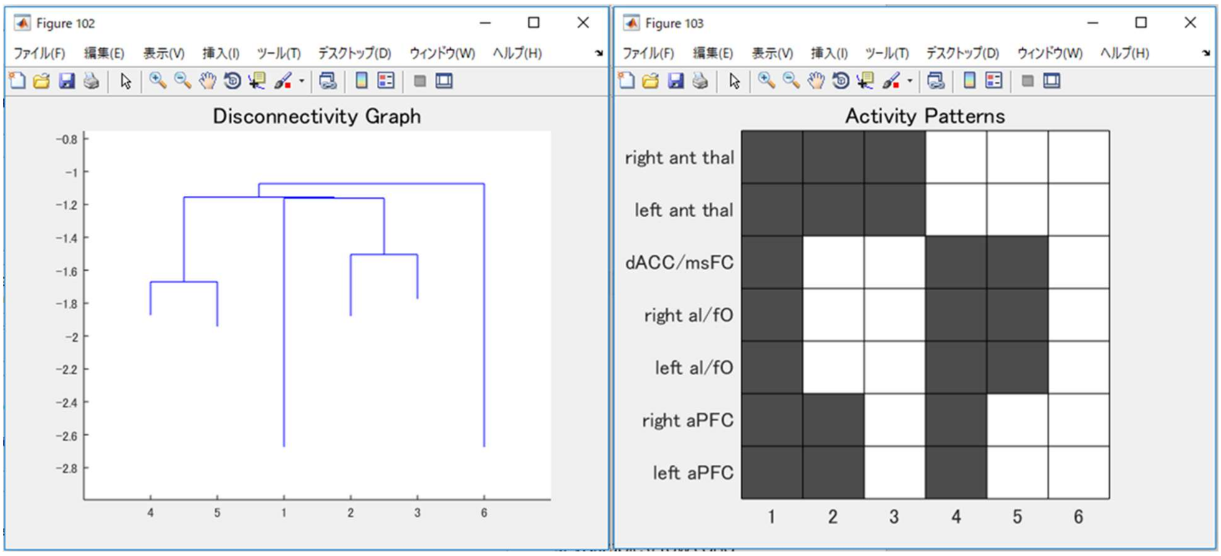

Fig. 6. Disconnectivity graph (left) and activity patterns (states) of each energy local minimum (right). The numbers on the x-axis labeling the energy local minimum states are consistently used in both panels. White and black cells in the right panel represent that variables (ROIs) are active (+1) and inactive (–1), respectively.

2. Time series

In the output folder, csv files are created for each piece of input data.

"*_SN.csv" (***: input file name):**

The time series of the state number, ***s(t)***, is saved.

"*_BN.csv":**

The time series of basin number, ***b(t)***, is saved. The basin number is the label of the basin in which the state is located at time ***t***.

3. Dynamics measures

In the output folder, a csv file, "Dynamics_yyyymmdd_HH_MM_SS.csv" is created. For each piece of input data, dynamics measures are saved.

InputFile:

The name of the data file.

Frequency of *Bj*:

The fraction of time that the basin number is ***j***.

Direct transition from *Bj* to *Bk*:

The number of direct transitions of basin number from ***j*** to ***k***, divided by t_{max}. Indirect transitions, e.g., ***j -> m-> k (m≠j,k)*** are **NOT** included.

Transition from n from *Bj* to *Bk*:

The number of transitions of the basin number from ***j*** to ***k***, divided by t_{max}. Indirect transitions, e.g., ***j -> m-> k (m≠j,k)*** are included.

Other options

1. Compute dynamics measures on other energy landscapes

In the tutorial, we constructed an energy landscape from the input data and computed dynamics measures on it. By choosing the "Read basin data" option, you can compute the dynamics measures based on a different energy landscape. When you execute the Full Analysis, the basin data is saved as "BasinData_YYYYMMDD_HHMMSS.mat" in the output folder. If you want to construct an energy landscape from data without computing dynamics measures, select the "Energy Landscape Construction" option.

If you perform a two-group study and separately execute Full Analysis for each group, the differences might be too exaggerated, because the differences in the basin sizes of the two landscapes, based on which the measures are computed, significantly affect the results. In such a case, we recommend using the same energy landscape or check if the basin sizes are not significantly different between the two groups.

2. Basin data

"BasinData_YYYYMMDD_HHMMSS.mat" contains a variable called "BasinGraph." The first column is the state number, and the third column is the local minimum state that the state belongs to. The second column shows the neighboring state that has the minimum energy value. This path is shown in the visualization of the basin (Fig. 5).

References

[1] T. Ezaki, T. Watanabe, M. Ohzeki, and N. Masuda, "Energy landscape analysis of neuroimaging data," Phil. Trans. R. Soc. A 375, 20160287 (2017).

[2] T. Watanabe, N. Masuda, F. Megumi, R. Kanai, G. Rees "Energy landscape and dynamics of brain activity during human bistable perception," Nat. Commun. 5:4765 (2014).

[3] T. Watanabe, et al. "Energy landscape analysis of resting-state brain networks," Front. Neuroinform. 8:12 (2014).

[4] A. Ashourvan, S. Gu, M.G. Mattar, J.M. Vettel, D.S. Bassett "The energy landscape underpinning module dynamics in the human brain connectome," Neuroimage 157:364–380 (2017).

[5] J. Kang, C. Pae, H.J. Park "Energy landscape analysis of the subcortical brain network unravels system properties beneath resting state dynamics," Neuroimage 149:153–164 (2017).

[6] Watanabe T, Rees G "Brain network dynamics in high-functioning individuals with autism," Nat. Commun. 8:16048 (2017).

[7] T. Ezaki, M. Sakaki, T. Watanabe, N. Masuda "Age-related changes in the ease of dynamical transitions in human brain activity," Hum. Brain Map. 39, 2673–2688 (2018).