# Statistical testing implemented in the KEY-LORETA software: Non-parametric randomization tests, aka Statistical non-Parametric Mapping (SnPM) methodology

## Introduction

The core idea upon which this methodology is based can be attributed to Fisher’s permutation test (R. A. Fisher, The Design of Experiment, New York: Hafner, 1935.). Some general information and references can be found at:

<http://en.wikipedia.org/wiki/Resampling_(statistics)>

(link alive 2015-06-15).

An additional excellent modern reference to this methodology is:

B.F.J. Manly: Randomization, Bootstrap and Monte Carlo Methods in Biology, Third Edition, Chapman & Hall, Boca Raton, 2007.

An excellent manuscript specifically for non-parametric randomization statistics in neuroimaging (SnPM) is:

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Map 15, 1-25.

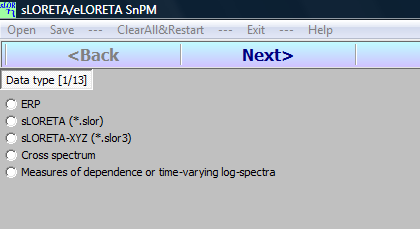
The core idea, which constitutes a cornerstone in statistics, that solved the problem of correction for multiple testing via non-parametric randomization, by making use of the max-statistic in the multivariate case, is in:

Westfall, PH.; Young, SS: Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. Wiley; New York: 1993

## The program

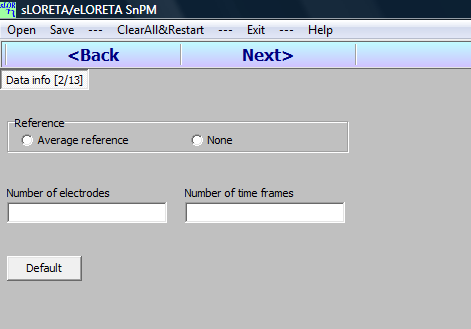
In the following, some pages might not appear during your analysis session, depending on the data type, type of test, etc.

On page 1, Data type [1/13]:



select the basic data type that you want to analyze.

Page 2, Data info [2/13]:

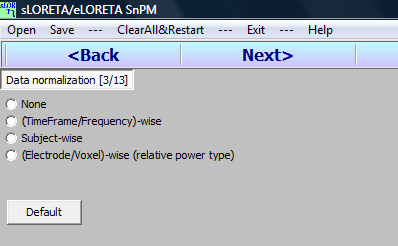


This appears only for ERP data. Specify number of electrodes and time frames. It is recommended to select average reference. Two letters explaining why the average reference should be used are:

Pascual-Marqui RD, Lehmann D. Topographic maps, source localization inference, and the reference electrode: comments on a paper by Desmedt et al. Electroencephalogr Clin Neurophysiol. 1993 Nov-Dec;88(6):532-6.

Pascual-Marqui RD, Lehmann D. Comparison of topographic maps and the reference electrode: comments on two papers by Desmedt and collaborators. Electroencephalogr Clin Neurophysiol. 1993 Nov-Dec;88(6):530-1, 534-6.

On page 3, Data normalization [3/13]:



select the type of normalization or scaling for the data. The option **None** corresponds to no scaling prior to statistics. Scaling is explained in the following references:

The concept of scaling/normalization of neuroimages can be found in:

S.J. Kiebel, A.P. Holmes. The general linear model. (2004) In: K.J. Friston, J. Ashburner, W.D. Penny (editors). Human Brain Function, 2nd edition, Part II – Imaging Neuroscience – Theory and Analysis, Elsevier science.

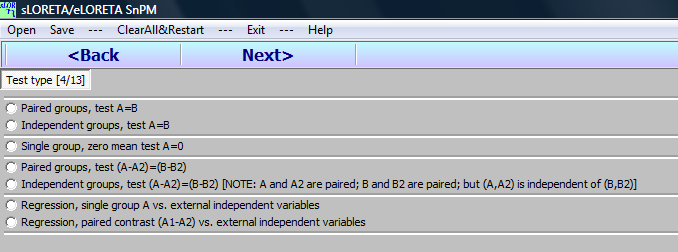
A draft version of this chapter was available at the SPM website:

<http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/pdfs/Ch7.pdf>

(link alive 2015-06-15).

Consider the case of sLORETA/eLORETA images. The option **(TimeFrame/Frequency)-wise** corresponds to image-wise scaling, i.e., each image is normalized (total average power equal to unity) prior to statistics. The option **Subject-wise** corresponds to normalizing the total power (over all images) of each subject. The option **(Electrode/Voxel)-wise (relative power type)** corresponds to normalizing the time signal (if time domain data) or the spectra (if frequency domain data) at each voxel.

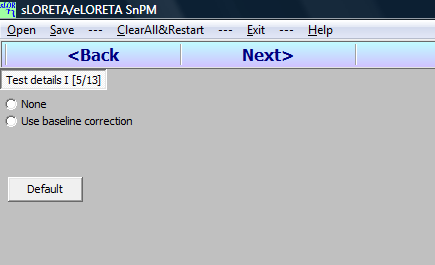
On page 4, Test type [4/13]:



select the type of test to perform.

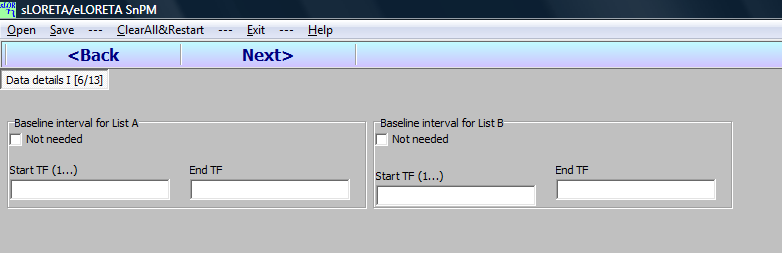
Note that there are a very limited number of possibilities. This package is intended for the simplest of all possible tests.

On page 5, Test details I [5/13]:



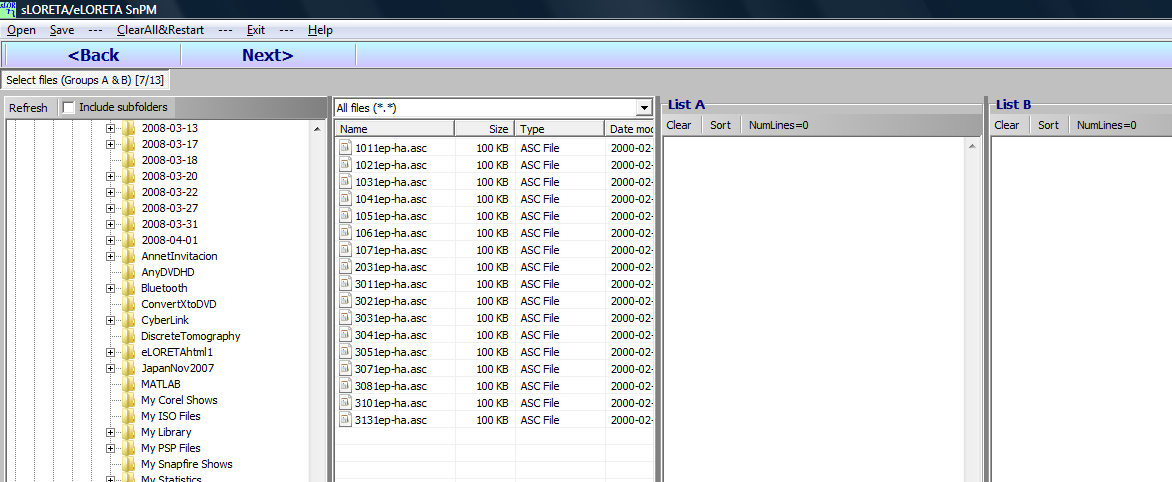
The user can choose to perform baseline correction. There is nothing illegal about performing a baseline correction, since it can be viewed nothing more than a legitimate contrast in the design. Note that the **Default** (preferred) option is **None** (no baseline correction).

On page 6, Data details I [6/13]:



set the baseline (if it was previously selected).

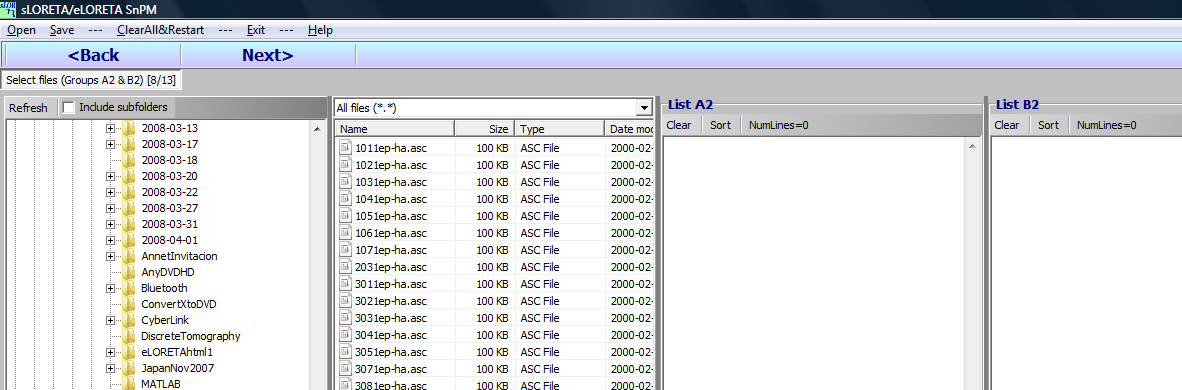
On page 7, Select files (Groups A & B) [7/13]:



select the files for analysis (groups A and B). In a paired design, the number of files must be equal, and their order is essential. The lists must be perfectly paired. This can be achieved by naming the files in such a way that when order alphabetically, they are paired. Otherwise, the test is incorrect.

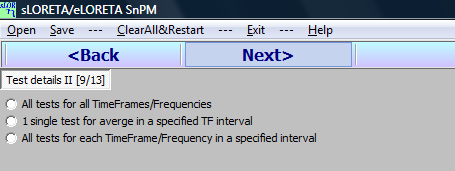
In the special case when a regression analysis has been selected, the list B will accept only one file, which should be a text file prepared by the User, containing the independent variables. This text file with the independent variables must have **NS** rows and **NIV** columns. **NS** is the number of subjects, i.e. the number of files in list A. **NIV** is the number of independent variables. This program will perform regressions against each independent variable, one at a time. It does not perform multiple regressions against all independent variables.

Page 8, Select files (Groups A2 & B2) [8/13]:



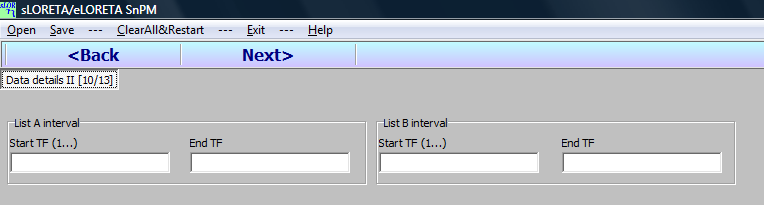
appears only if a contrast of the type (A-A2)=(B-B2) was selected as test. Care must be taken to ensure that files of group A and group A2 be paired; and that files of group B and group B2 be paired. Otherwise, the test is incorrect.

One page 9, Test details II [9/13]:



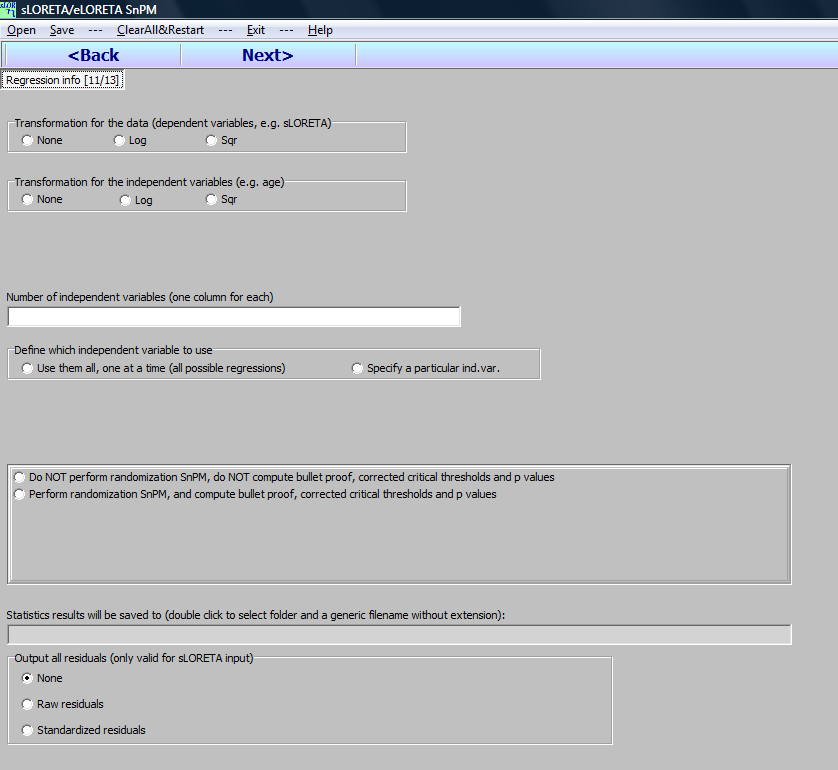
select how the tests should be performed.

Page 10, Data details II [10/13]:



appears only if an interval was selected in the previous page.

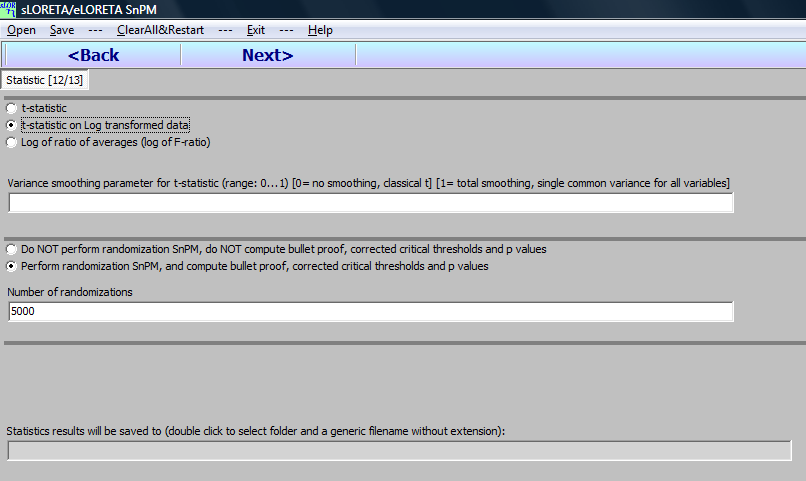
Page 11, Regression info [11/13]:



appears only if a regression analysis was selected.

Recall from **Page 7** above, that when a regression analysis has been selected, the list B of files will accept only one file, which should be a text file prepared by the User, containing the independent variables. This text file with the independent variables must have **NS** rows and **NIV** columns. **NS** is the number of subjects, i.e. the number of files in list A. **NIV** is the number of independent variables. This program will perform regressions against each independent variable, one at a time. It does not perform multiple regressions against all independent variables.

On Page 12, Statistic [12/13]:

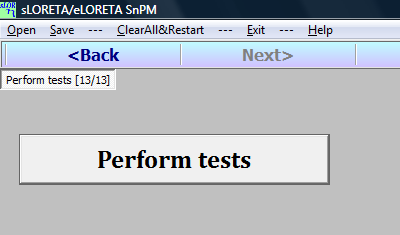


select the actual statistic to be used, the **variance smoothing parameter**, randomizations, and a filename for output.

Variance smoothing is explained in:

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Map 15, 1-25.

On Page 13, Perform tests [13/13]:



click the large **Perform tests** button.

Several files will be produced.

**\*.TestSetup**: this contains information about all the test parameters. It can be read into the program with the **Open** option. In this way one does not have to go though all pages if one wants to repeat the test with other parameters.

**\*-TestSetup.txt**: This is a text file, useful to the User, containing all the test parameters used. One can check exactly which type of test was performed by looking at the contents in any text editor.

**\*.xxx**: The extension xxx can be **slor** or **txt**, depending on the data type analyzed. This file contains the test statistics (either t-values, or log-F-ratios, or correlation coefficients, etc.)

**z-Thresholds&ExtremePs.txt**: This text file is produced only if randomization was carried out. The first part contains the thresholds for one and two tail tests, at p=0.01, 0.05, and 0.1. The last part contains information on the **Exceedence proportion tests**. Information on this test can be found in:

Friston KJ, Frith CD, Liddle PF, Dolan RJ, Lammertsma AA, Frackowiak RS.

The relationship between global and local changes in PET scans.

J Cereb Blood Flow Metab. 1990 Jul;10(4):458–466.

Friston K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. 1991.

Comparing functional (PET) images: The assessment of significant change.

J. Cereb. Blood Flow Metab. 11,690–699.

The exceedence proportion test reports a collection of threshold values and the corresponding probabilities that the number of suprathreshold voxels is significant. See the references quoted above.

In the exceedence proportion test, it may happen that a threshold needs not be high to be significant, e.g. a t-threshold as low as 1.5 might correspond to a significant number of voxels. This test is not always well defined, and the results may be trivial. For instance, consider two conditions such as slow wave sleep and awake, resting, eyes closed. The spectral power of the generators is overwhelmingly larger during sleep over the whole head. In this case the exceedence proportion test will report that the set of all voxels is significant.

The exceedence proportion test might find significant activation for a group of voxels, even when the voxel-by-voxel test finds no single significant voxel. This is because the significant event here corresponds to the simultaneous activation (even if low) of many voxels.

## An example “Thresholds” file

Consider the following thresholds results, obtained when using randomization:

t(0.01) t(0.05) t(0.10) ExtremeP

One-Tailed (A>B): 5.050 4.313 3.994 0.13380

One-Tailed (A<B): -4.913 -4.255 -3.944 0.79000

Two-Tailed (A<>B): 5.228 4.574 4.272 0.24580

-------------------

Exceedence proportion tests:

Thrsh(1Tailed>0) Prob(1Tailed>0) Thrsh(1Tailed<0) Prob(1Tailed<0) Thrsh(2Tailed) Prob(2Tailed)

0.383396 0.501000 -0.272538 0.696600 0.383396 0.732400

0.766792 0.358000 -0.545077 0.660400 0.766792 0.577000

1.150187 0.267000 -0.817615 0.821400 1.150187 0.695800

1.533583 0.154600 -1.090153 0.942600 1.533583 0.507600

1.916979 0.086600 -1.362692 0.938000 1.916979 0.280200

2.300375 0.011000 -1.635230 0.901400 2.300375 0.056000

2.683771 0.002800 -1.907768 0.780000 2.683771 0.026200

3.067167 0.017000 -2.180307 0.739400 3.067167 0.052400

3.450562 0.026200 -2.452845 0.735800 3.450562 0.061600

3.833958 0.133600 -2.725384 0.790000 3.833958 0.245600

In this example, there are two conclusions:

1. No single voxel is significant, since the most extreme voxels have p-values >0.05 (first four lines).

2. The set of all voxels with absolute t-values greater than 3 are jointly significant, according to the exceedance proportion results.

## Some additional notes:

Non-parametric randomization makes use the t-statistic or the “t-statistic on log transformed values” or the log-F-ratio statistic. These are good measures of “statistical distance” between groups or conditions. However, inference does not rely on these measures having an actual t- or F-distribution. The actual distributions are estimated, and thus the tests are in a way “bullet proof”, in that they work regardless of the distribution of the original data (Gaussian or not), and the inference corrects for multiple testing exactly.

The t-test for two independent samples is described at:

https://en.wikipedia.org/wiki/Student's\_t-test

(link alive 2015-06-15).

The log-F-ratio statistic is based on the F-test for equality of variances (which is equivalent to equality of spectral power), described at:

https://en.wikipedia.org/wiki/F-test\_of\_equality\_of\_variances

(link alive 2015-06-15).