

# MIDAS: Regionally linear multivariate discriminative statistical mapping

Erdem Varol  
erdem.varol@uphs.upenn.edu  
software@cbica.upenn.edu

April 3, 2018

## 1 Introduction

This software yield statistical maps for group comparisons or regressions. The statistical maps are based on regionally linear multivariate discriminative analysis. P-value maps are based on an analytic approximation of permutation testing.

## 2 Method

Voxel-based morphometry (VBM) has become an essential tool in summarizing and understanding the brain patterns that exhibit change through either pathological or normal processes. Over the years, many variants and improvements for this technique were proposed that aimed to improve the signal-to-noise ratio of the resulting statistical significance maps.

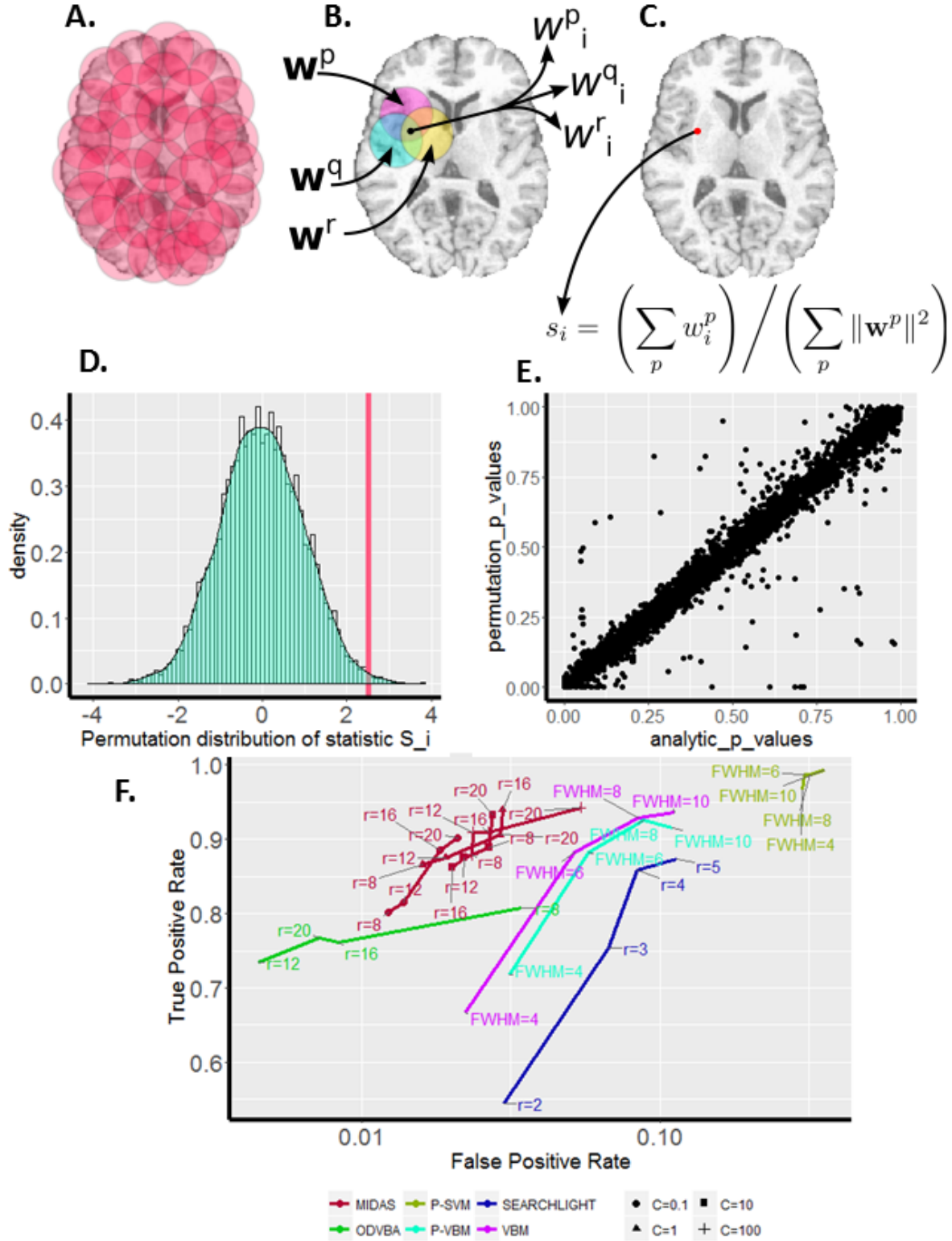
The most common approach is to pass the input images through Gaussian filters to reduce the voxel-wise noise before performing statistical analyses. However, the bandwidths of these filters are fixed and not adaptive to the anatomical structure of the brain which often leads to false positives and negatives.

To enhance voxel-based analysis, a novel statistical method MIDAS (multivariate inference with discriminative adaptive smoothing) is introduced. MIDAS utilizes overlapping local multivariate classifiers (Fig.1A,B) to obtain voxel-wise statistics (Fig.1C) whose distribution under random permutation of groups is approximately Gaussian with parameters that can be estimated from data (Fig.1D). This allows for fast and accurate voxel-wise p-value estimation for group comparisons (Fig.1E). MIDAS is validated in a simulated setting as well as in a clinical setting using a task-based fMRI dataset of truth-telling and lying.

For more information, please refer to:

Varol, Erdem, Aristeidis Sotiras, Christos Davatzikos.

"MIDAS: regionally linear multivariate discriminative statistical mapping."  
NeuroImage (2018)



**Figure 1: Outline of the proposed method MIDAS: A) Local neighborhoods that cover the brain are sampled. B) Multivariate classifier/regressor is applied to each neighborhood C) Statistic using the classifier/regressor weights is obtained D) P-value is computed through permutation testing E) Permutation based distribution is approximated analytically F) True positive/False positive rate of MIDAS (in red) and compared methods for a simulated validation set.**

## 3 Testing & Installation

This software has been primarily implemented in MATLAB for Linux operating systems.

### 3.1 Dependencies

- NIFTI Matlab toolbox (necessary files have been included in package)

### 3.2 Installation

Midas can be run directly in a matlab environment without compilation.

#### OPTIONAL:

If user wants to run midas as a standalone executable, then it must be compiled as following (using the additionally obtained matlab compiler "mcc"):

Run the following command in a MATLAB environment:

```
mcc -m midas.m
```

### 3.3 Testing

We provided a test sample in the test folder.

To test in matlab enviroment, use the command:

```
midas('-i','test.csv','-o','.','-r',15,'-p',200,'-c',0.1)
```

To test in command line using compiled executable, use the command:

```
midas -i test.csv -o . -r 15 -p 200 -c 0.1
```

This runs a MIDAS experiment which may take a few minutes. The test case contains a gray matter RAVENS maps of 46 subjects from the ADNI dataset. The output should yield statistical maps for diagnosis, age, and sex along with p-value maps in `.nii.gz` format. An accompanying MATLAB `.mat` file that stores these results is also output as `MIDAS_results.mat` in the output directory.

### 3.4 Test verification

Pre-computed statistical and p-value maps along with the `.mat` file have been included in directory "Pre\_computed\_test\_results". The user may verify that their test results match the pre-computed results to confirm proper set-up.

## 4 Usage

function returns MIDAS statistical maps and associated p-values computed by analytically estimating permutation testing

### INPUT

#### REQUIRED

[--input, -i] : .csv file containing full paths to input images. (REQUIRED)

We assume that the first column contains subject identifying information; the second column contains the path to the images, while the last column contains label information. First line of the file should contain header information.

[--outputDir, -o] : directory where the output from all folds will be saved (REQUIRED)

#### OPTIONAL

[--c, -c] : regularization parameter (positive scalar) (default C=0.1)

[--radius, -r] : neighborhood radius in voxels (positive scalar) (default R=15)

[--num, -p] : number of neighborhoods (positive scalar) (default P = 200 )

[--usage, -u] Prints basic usage message.

[--help, -h] Prints help information.

[--version, -v] Prints information about software version.

### OUTPUT:

map = structure that stores MIDAS statistics and p-values for every group/covariate provided with the following structure

map.stat = cell that stores MIDAS statistic for groups/covariate in same order (output as .nii.gz file as well)

map.p = cell that stores MIDAS statistic for groups/covariate in same order (output as .nii.gz file as well)

map.N = neighborhood coverage amounts (output as .nii.gz file as well)

NOTE: to compile this function do

mcc -m midas -A [NIFTI toolbox directory]

EXAMPLE USE (in matlab)

midas('-i','test.csv','-o','.','-r',15,'-p',200,'-c',0.1)

EXAMPLE USE (in command line)

midas -i test.csv -o . -r 15 -p 200 -c 0.1

## 4.1 Input images

MIDAS requires input images in nifti gz format (.nii.gz file extension). The images should be registered to a common template space and share the same dimensionality.

## 4.2 Running MIDAS

Here is a brief introduction to running MIDAS. For a complete list of parameters, see `--help` option.

To run this software, you will need an input csv file, with the following mandatory fields in the following column order: (Column 1) ID: ID for subject (Column 2) filepath: path for input image for subject (Column 3 and on) covariate\_1, covariate\_2, ...: covariates to obtain statistical maps for (Needs to be numerical)

NOTE: column header names can be arbitrary, only the order matters.

An example input csv file looks as following:

ID,	filepath,	DIAG,	AGE,	SEX
subject_1,	./data/subject_1.nii.gz,	1,	79.3,	-1
subject_2,	./data/subject_2.nii.gz,	1,	71.4,	1
subject_3,	./data/subject_3.nii.gz,	1,	82.7,	-1

If you install the package successfully, there will be two ways of running MIDAS:

1. Running MIDAS in a matlab enviroment, a simple example:

```
midas('-i','test.csv','-o','.','-r',15,'-p',200,'-c',0.1)
```

2. Running matlab compiled MIDAS executables in the command line, a simple example:

```
midas -i test.csv -o . -r 15 -p 200 -c 0.1
```

The software returns: 1. statistical maps for provided covariates along with p-value maps in .nii.gz format. 2. Same results in a matlab format in MIDAS\_results.mat

## 4.3 Choice of parameters

There are two main parameters in MIDAS. The first parameter is the neighborhood radius,  $r$ , which controls the size of the local discriminative analysis window. The second parameter is the weight  $c$  in the LS-SVM objective. This parameter controls for the amount of slackness in the constraints of the LS-SVM objective, allowing for cases when the data points  $\mathbf{X}$  are not linearly separable with respect to the labels  $\mathbf{y}$ . In other words,  $c$  controls the degree to which  $\mathbf{w}$  fits the data. One particular way by which the  $c$  and  $r$  parameters can be selected is by using the resulting significance maps for feature selection and assessing out of sample predictive performance through nested cross-validation.

One can also set the number of neighborhoods  $P$ , which are sampled such that the full brain volume is covered. The MIDAS statistic is self-normalized to have zero mean and unit variance independent of the selection of  $P$ . In our experiments,  $P$  is selected such that each voxel across the brain is covered at least 20 times for a given neighborhood radius. A practical suggestion for setting  $P$  is to assess to reproducibility of resulting statistical maps over a range of candidate of number of neighborhoods and choose the minimum value that attains stability.

Note that the topology of the regional neighborhoods need not be spherical nor compact for the resulting statistic to be valid. Thus, neighborhoods that are discontinuous or anisotropic may be deployed in implementation. However, for simplicity, spherical neighborhoods were used in the implementation described within.

Lastly, to ensure that the coverage of the brain is relatively uniform, the MIDAS implementation accounts for the number of times each voxel has been covered to adaptively cover undersampled regions at each iteration.

## 5 Citation

If you find this software useful, please cite:

Varol, Erdem, Aristeidis Sotiras, Christos Davatzikos.

"MIDAS: regionally linear multivariate discriminative statistical mapping."  
NeuroImage (2018)

## 6 Licensing

See <https://www.med.upenn.edu/sbia/software-agreement.html> or COPYING.txt file.